December 2024

initiative for applied artificial intelligence

Generative AI Agents in Action:
Revolutionizing Software Development Testing

Contents

Foreword

A Few Thoughts at the End of 2024

As we approach the close of 2024, the landscape of Artificial Intelligence (AI) is undergoing a profound transformation. Since the pivotal ChatGPT milestone two years ago, we have witnessed a rapid adoption of Large Language Models (LLMs) in a huge variety of applications. In most cases, LLMs have been used for multi-modal content generation, knowledge retrieval and chatbots, already generating value in various industries and their value chains.

A common pattern across current generative AI applications is the instruction-oriented interaction between users and AI, primarily facilitated through a chat format. Whether users formulate specific questions, need a summary or key insights from a document, or want to note their thoughts and ideas for future reminders, the typical usage pattern involves providing the AI with a specific prompt or instruction to elicit the desired response.

While this interaction model has driven numerous relevant and valuable business use cases and directly embeds the human-in-control principle to mitigate certain risks of generative AI, it does not fully utilize the most recent capabilities of generative AI. With the advent of more powerful LLMs equipped with deep reasoning and thinking abilities, use of tools, and the capacity to understand and synthesize multilingual and multimodal data, **generative AI is increasingly capable of solving complex problems by translating them into a set of autonomous steps or tasks, much like humans would. We refer to these advanced systems as generative AI agents.**

Leading organizations such as Anthropic, Microsoft, NVIDIA, OpenAI, Salesforce, SAP, and others are at the forefront of developing agents that not only follow commands but also proactively solve complex problems by aligning with broader objectives. **Although we are still in the early phases of agent development, the evolvement towards autonomous multi-agent systems is already underway.** In the not-too-distant future, these agents hold immense promise, as they begin to tackle intricate tasks that were once the exclusive domain of human intelligence.

In fact, generative AI agents have now become a spotlighted field within AI technology. Why is this so? Because **generative AI agents represent a shift from instruction-oriented chat interactions,** where humans guide problem-solving, **to task delegation and autonomous problem-solving with minimal or even no human oversight** in the future. This shift opens up vast potential for businesses, enabling task automation in software-based virtual environments and even action planning in physical environments.

In this white paper, we explore the rise of generative AI agents, transitioning from traditional instruction-driven interactions to innovative, goal-oriented automation. We will delve into the evolving progress of generative AI agents, market observations, and the exciting potential of autonomous systems, particularly in the field of software development. We invite you to reflect on the technological advancements shaping our future and the implications of an increasingly automated world.

Bernhard Pflugfelder
Head of Generative AI,
appliedAI Initiative GmbH

Dr. Paul Yu-Chun Chang
Senior AI Expert: Foundation Models -
Large Language Models,
appliedAI Initiative GmbH

Mingyang Ma
Principal AI Strategist & Product Manager,
appliedAI Initiative GmbH

Executive Summary

Adaptability, Transformation, and Responsibility – Getting Prepared for the Agentic Era

Adaptability

Generative AI agents are defined by their ability to **interact with environments** and **execute tasks autonomously**, showcasing cognitive processes like **reasoning** and **goal setting** for problem solving.

- Currently, most agents operate at foundational levels of **conversation, analytical capability,** and **autonomy**, but significant advancements toward **innovation and collaboration** are anticipated.
- These agents already enhance business processes across various value chains, from research to customer service, by automating complex tasks. Future agent systems can potentially automatize entire processes instead of use cases.
- The transition **from Robotic Process Automation (RPA) to Agentic Process Automation (APA)** allows for dynamic, goal-oriented workflows that improve cost and time efficiency of processes already now. The use of Small Language Models (SLMs) instead of LLMs within APA is an important approach to optimize costs and deploy agents on-premise or on edge.
- As generative AI evolves, there is an increasing focus on mitigating the known risks of LLMs. However, since these risks cannot be completely eliminated, there remains an important need for **human oversight in agentic systems.**

Transformation

Generative AI agents are able to significantly transform processes instead of single use case or tasks.

- A highly promising process showcasing this transformative potential is the **software development lifecycle**, particularly through roles and tasks in **planning, development, testing, review**, and **deployment**. Notable use cases include **code generation, automated testing,** and **automated reviewing.**
- While there is a cautious approach to their use in critical tasks like infrastructure deployment, these agents enhance workflows by automating test creation and adapting to evolving requirements, thereby improving overall efficiency in the software lifecycle.
- Technologies such as Retrieval-Augmented Generation (RAG) and AutoGen help reduce manual testing efforts, allowing developers to focus on complex problem-solving. Beyond software, these agents are impacting fields like **industrial engineering** and **scientific research** by streamlining tasks and fostering collaboration.

Responsibility

Generative AI agents present both remarkable **opportunities** and significant **challenges**.

- The creation of **scalable, multi-modal agentic systems** capable of integrating **diverse sensory inputs** and harnessing **collective human and artificial intelligence** will open new frontiers for generative AI across various sectors.
- While they hold potential for enhancing efficiency, their susceptibility to adversarial attacks raises concerns about their **robustness** and **trustworthiness**, particularly as these systems are designed to predict and perform actions.
- As AI technology advances, it is vital to anticipate potential risks and to adapt evaluation methods for real-world applications, including methods like **agent-as-a judge solutions** with **human oversight.**
- Addressing **ethical and stability considerations** and ensuring **responsible use** are essential to mitigating risks.

By carefully weighing both the opportunities and challenges, we can fully realize the transformative potential of generative AI agents while safeguarding societal well-being.

“AI agents will drive business automation and business decision augmentation. They will advance to specialized assistants that will help users in various business roles by driving business decisions and taking action. Ultimately, this will not only lead to much more efficient and guided processes, but also transform the business processes themselves.”

Dr. Christian Karaschewitz

AI Product Incubation Lead,

SAP Business AI – Product & Partner Management

SAP SE

“Artificial Intelligence has fundamentally transformed how we interact with software, making natural language interfaces not just possible but powerful. The next frontier lies in AI agents – autonomous systems that can take intelligent action on our behalf. As these agents evolve from concept to reality, organizations and individuals alike must actively explore and experiment with them to understand their transformative potential.”

Antoine Leboyer
Managing Director SW/AI,
TUM Venture Labs

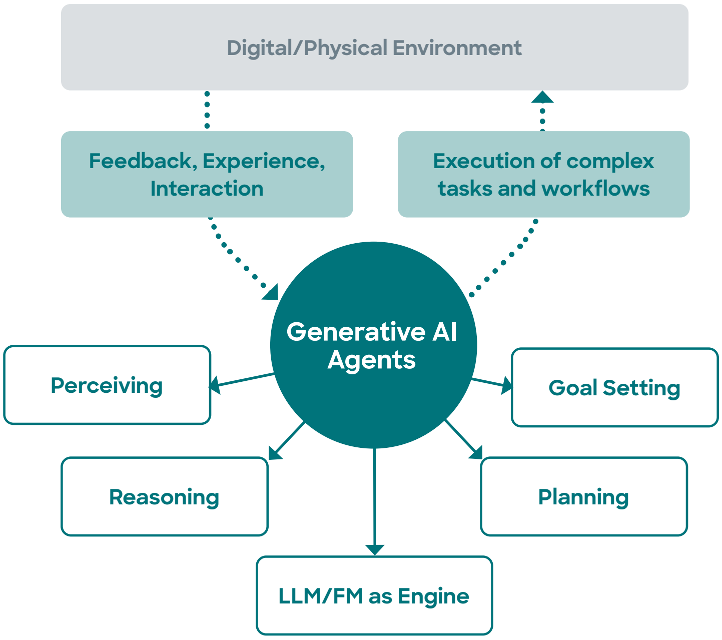
A Quick Dive into Generative AI Agents

Generative AI Agents: What & Why

A generative AI agent is an **autonomous** system that leverages **large language models** and **foundation models** to **independently** execute **complex tasks** and **workflows** in a **digital/physical environment.** It **perceives** its surroundings, **reasons**, **plans**, and **acts** over time to achieve its **goals** and influence future outcomes [1-4].

7 Reasons to Use GenAI Agents

1. **Automation and Efficiency:** Automating repetitive tasks to boost productivity and allow focus on strategic work.
2. **Adaptability and Flexibility:** Handling flexible tasks and dynamically adapting to various scenarios.
3. **Personalization and Customization:** Tailoring experiences and recommendations based on user preferences to enhance engagement.
4. **In-depth Reasoning:** Improving robustness in addressing ambiguous scenarios by utilizing advanced reasoning and reflective processes.
5. **Decision Support:** Analyzing data to provide insights, aiding informed decision-making in complex situations.
6. **Innovation Enhancement:** Inspiring creativity by collecting and generating innovative insights that human minds may overlook.
7. **Simuation of Complex Systems:** Optimizing real systems through simulations, such as in the case of cyber attacks and digital twins.



Five Levels of Competence in Generative AI Agents

Here we outline five levels of generative AI agent competence—**conversational, reasoning, autonomous, innovating,** and **organizational**—based on their capabilities in thinking (**brain**), perceiving (**perception**), and task execution (**action**). While not exhaustive, this categorization aims to provoke inquiries about the dynamics of human-agent interactions [5-7].

	Level 1 (Conversational)	Level 2 (Reasoning)	Level 3 (Autonomous)	Level 4 (Innovating)	Level 5 (Organizational)
Brain	Episodic Memory: Events	Human-like Reasoning	Goal-setting	Autonomous Learning	Personality & Role-Playing
	Summary & Abstarction	Reflection & Critique	Planning Multistep Tasks	Generalization	Team Dynamics Insight
	Semantic Memory: Knowledge	Judgement & Evaluation	Decision-making	Goal Recalibration	Strategic Thinking
				Idea/Design Generation	Coordination Planning
Percep-tion	Textual Input Encoding	Pattern Recognition	Active Sensing/Monitoring	Perceptual Learning	Organizational Monitoring
	Visual Input Encoding	Multi-source Input Integration	Goal-directed Perception	Perceptual Recalibration	Collective/Mutual Perception
	Auditory Input Encoding		Autonomous Data Mining	Perceptual Anticipation	System Failure Awareness
	Other Sensor Input Encoding				
Action	Conversation Completion	Intent Inference	Automated Tool Usage	Learning/Making New Tools	Multi-agent Collaboration
	Question Answering	Tool Selection	Embodied Actions	Self-improving/refining	Conflict Resolution
		Analytical Problem Solving	Routing/Navigation	Prototyping	Project Management
					Mutual Task Delegation

A Quick Dive into Generative AI Agents

Building Generative AI Agent Systems for Business

Fundamental Agentic System Design Patterns

To build generative AI agent systems for business, let's firstly look into three main design patterns for such systems [5].

Single Agent

Characteristics

- Versatile capabilities for various application tasks.
- High task-solving performance in diverse contexts.

Typical Scenarios

- **Task-oriented:** Assisting users in daily tasks (e.g., comprehension & task decomposition).
- **Innovation-oriented:** Autonomous exploration in scientific fields.
- **Lifecycle-oriented:** Continuous learning and skill development for long-term survival.

Multi-Agent

Characteristics

- Cooperative or adversarial interactions for advancement.
- Agents work together or compete to improve results.

Typical Scenarios

- **Cooperative Interaction:** Agents collaborate, either orderly or disorderly, toward common goals.
- **Adversarial Interaction:** Competitive dynamics for individual performance enhancement.

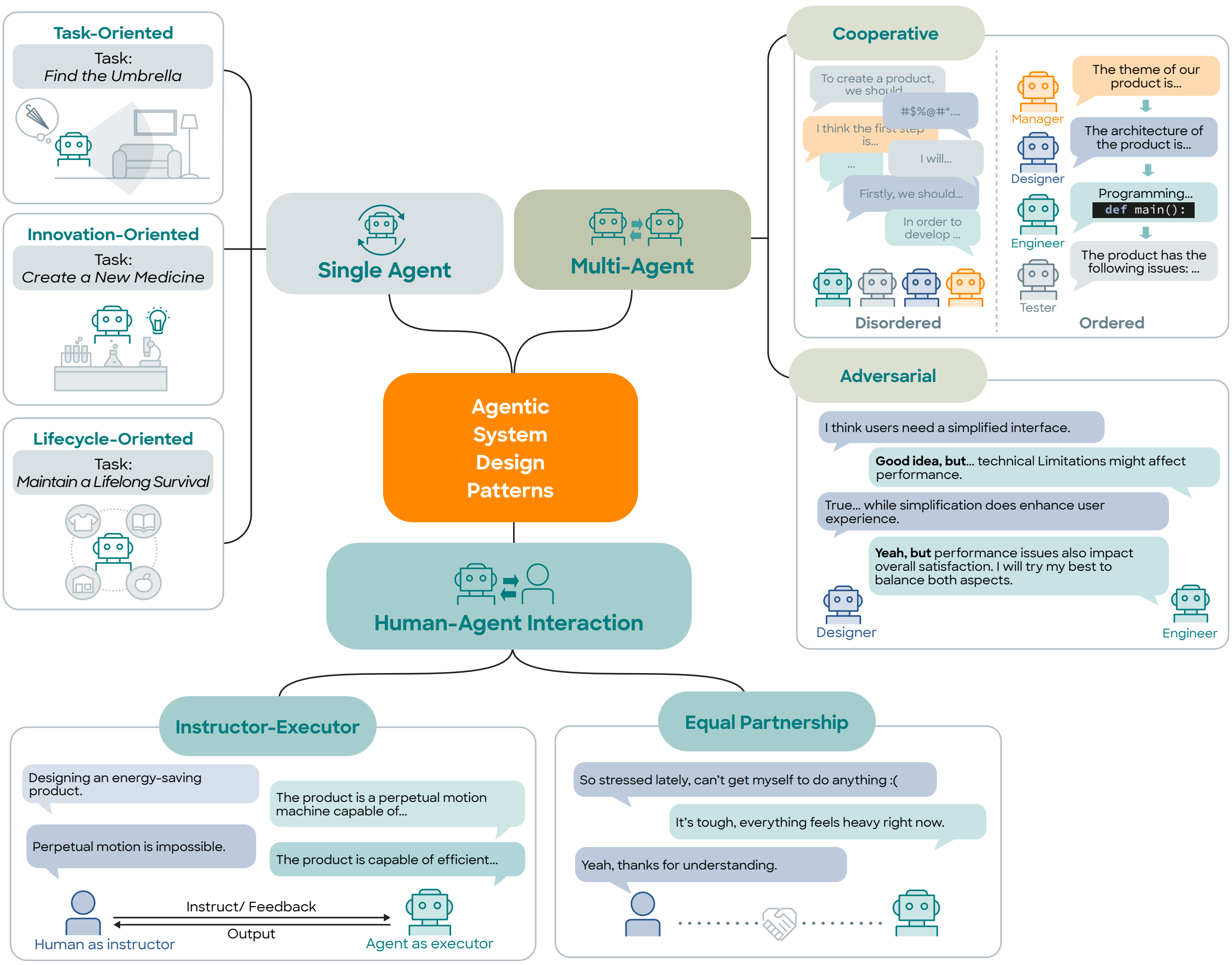
Human-Agent Interaction

Characteristics

- Human feedback enhances agent task efficiency and safety.
- Agents improve service quality for human users.

Typical Scenarios

- **Instructor-Executor Paradigm:** Humans give instructions; agents execute tasks.
- **Equal Partnership Paradigm:** Agents engage empathetically and collaborate with humans.



A Quick Dive into Generative AI Agents

Core Components of An Arbitrary Business Problem

- To develop generative AI agent systems for business scenarios, we pinpoint six key components illustrated below that may exist within an arbitrary business problem that agents can address.
- In the case of **customer feedback analysis,** for example, the following components are critical [5-7]:
  - **Understanding & Perceiving:** Gather customer feedback from various sources (surveys, reviews, social media) to understand sentiments and trends.
  - **Reflecting & Analyzing:** Analyze the feedback to identify common themes and areas for improvement. Use natural language processing to extract insights and sentiments.
  - **Iterative Review:** Continuously monitor customer reactions to implemented changes, collecting new feedback to assess the effectiveness of actions taken.

Understanding & Perceiving

Gather and interpret relevant data to comprehend the business environment and specific problem context.

Reflecting & Analyzing

Evaluate information critically, identifying patterns, trends, and insights to inform future actions.

Decision-making

Generate optimal decisions and strategies based on collected insights and analyzed data.

Action Execution

Implement decisions through automated processes, ensuring efficiency and precision in task completion.

Iterative Review

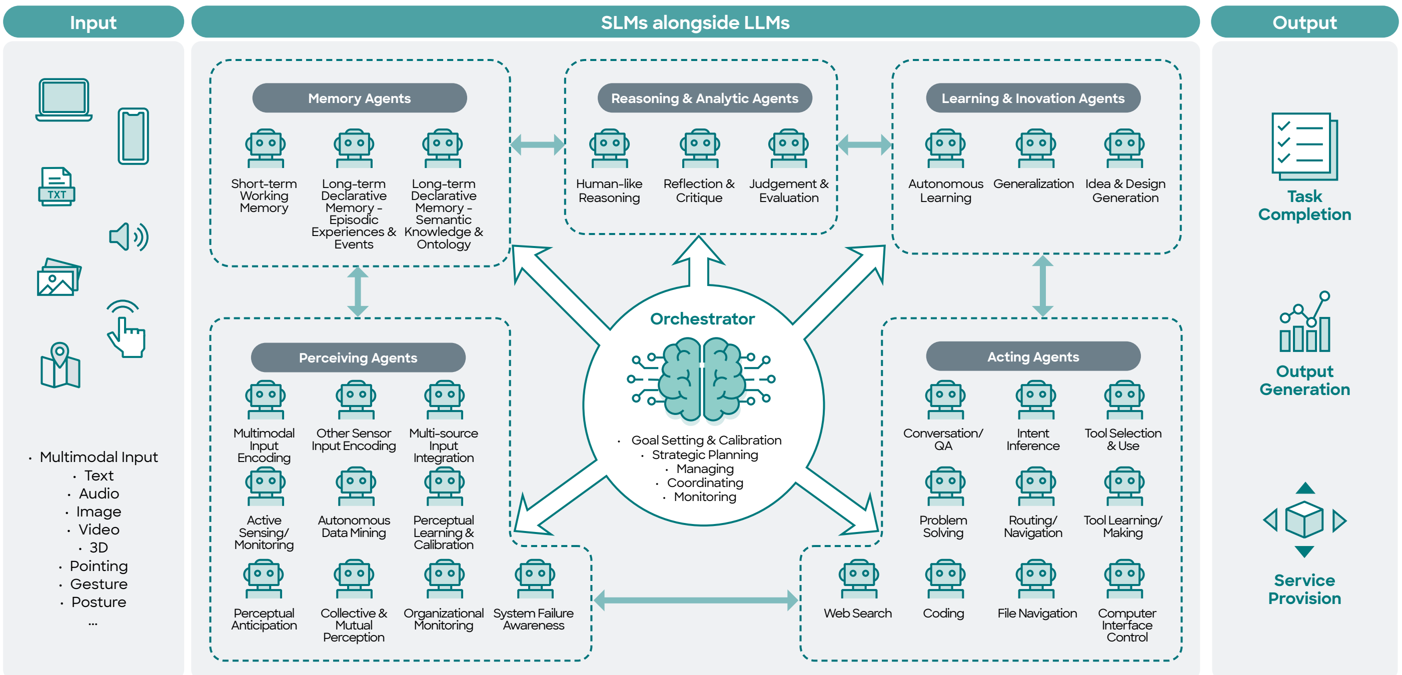
Continuously monitor outcomes and gather feedback to refine processes and improve decision-making over time.

Coordination

Facilitate collaboration between various agents and stakeholders to ensure alignment and synergy in achieving business objectives.

Example Generic Multi-Agent Framework

- With the core components of business problems in view, this section illustrates a generic multi-agent framework based on three core generative AI agent capabilities as outlined earlier: the **"brain"** (memory, reasoning & analytic, as well as learning & innovation agents), along with **perception**- and **action**-related agents. Here we aim to broadly outline potential agents based on various previous definitions, though this graphic is by no means exhaustive [5].
- In a multi-agent system, agents can interact in various patterns, such as in a hierarchical or decentralized structure. Most existing frameworks typically employ a **centralized communication model**, where an **orchestrator** sets goals, creates plans, delegates tasks to specific agents, and monitors the outcomes [5, 8-10].
- For instance, in **customer feedback analysis,** an **orchestrator** may collaborate with an **autonomous data mining** agent and a **reflection** or **evaluation** agent to effectively accomplish the task.



Industrial Agentic Use Cases: *From Strategy to Operation*

Prerequisites for Agentic Use Case Adoption: Strategic Perspective

Despite the promising capabilities of large language models and the future potential of AI agents, the **successful adoption of AI agents in organizations,** as with any AI technology, depends on various critical elements. The **appliedAI AI Strategy House** framework provides a clear visual and conceptual structure of these essential elements. By systematically aligning and measuring these elements, organizations can implement and scale AI agents more effectively, ensuring the maximization of their AI investments. The crucial **prerequisites of AI agent adoption** associated with the **six enabling elements** in the appliedAI Strategy House are detailed below.

AI STRATEGY HOUSE

Ambition	Future competitive advantage		Fields of action		Commitment	
AI use cases	Discovery & specification		Make or buy		Portfolio management	
Enabling factors	Organization	Expertise	Culture	Data	Technology	Ecosystem
Execution	Research & exploration		Development & validation		Operationalization & maintenance	

Organizational Commitment

- **Leadership Support:** Executives drive AI initiatives and secure necessary resources.
- **Cross-functional Collaboration:** Align goals and stimulate innovative ideas.
- **Ongoing Training:** Provide continuous learning opportunities to upskill employees

Expert-empowered Workflow

- **Empowerment:** Design workflows that balance the autonomy of AI agents with employee acceptance, incorporating elements such as human oversight.
- **Future Role Definition:** Clearly outline roles and responsibilities for AI and employees in future AI-supported workflows.

Promoted User Adoption

- **Training and Support:** Provide comprehensive training and ongoing assistance for users (i.e. customers or employees)
- **Incentives:** Cultivate an innovative mindset and encourage user adoption through rewards.
- **User-Centric Design:** Develop intuitive AI tools that meet user needs and emphasize trustworthy human-agent interactions.

Data Quality and Access

- **High-Quality Data:** Ensure data in different languages and modalities is accurate, consistent, and relevant.
- **Sufficient Quantity:** Gather enough data to train models effectively. Also consider augmentation with synthetic data generation.
- **Seamless Integration:** Connect agentic applications with data in existing systems.

AgentOps & Adaptable Tools

- **Robust Evaluators:** Implement validation frameworks to ensure quality. Automated evaluation methods can be deterministic, statistical, or AI-based.
- **Dynamic Monitoring:** Continuously review AI performance and in production for optimization
- **Adaptatable Tools:** Develop clear strategies for process/tool changes.

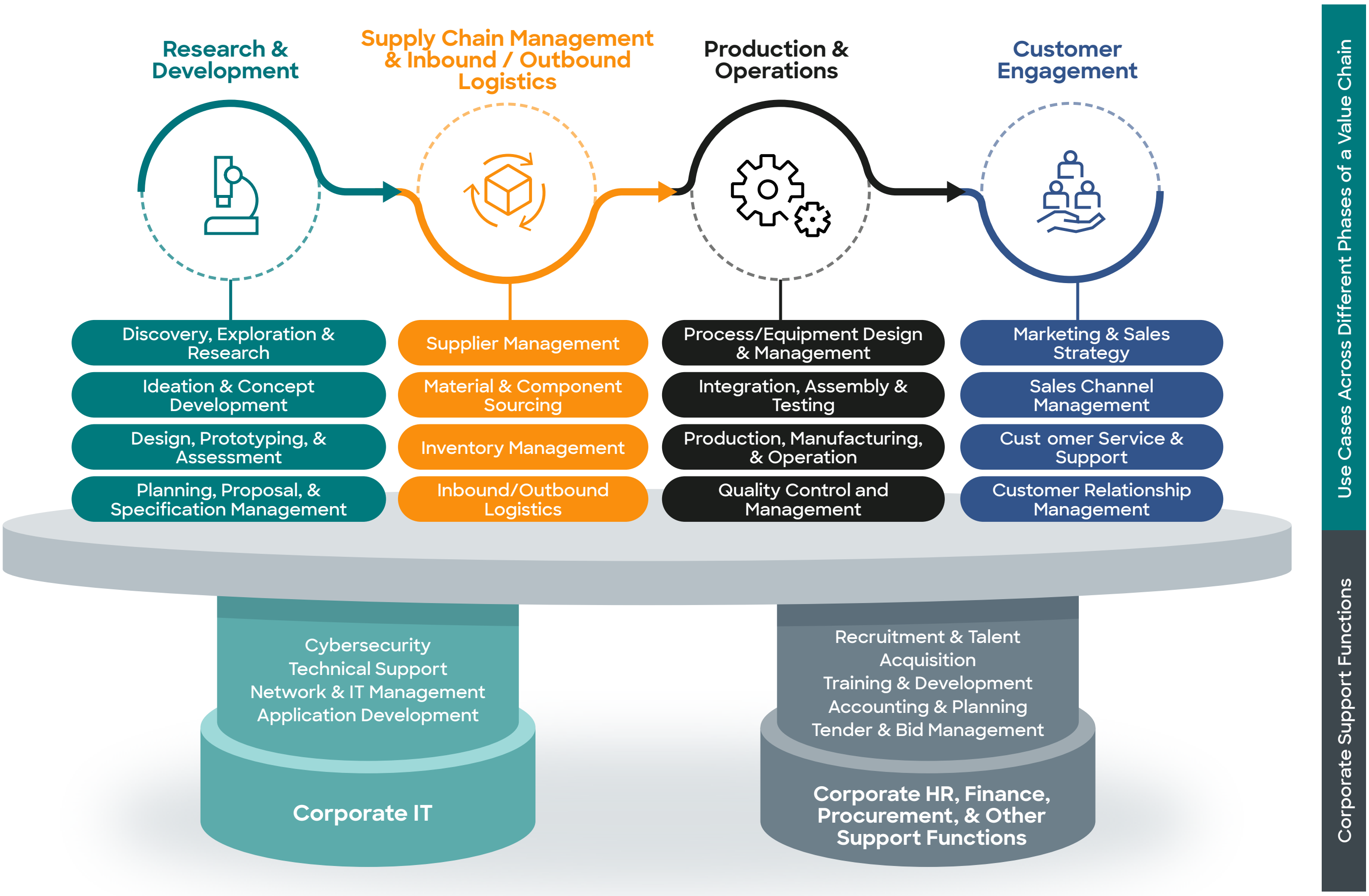
Ecosystem Support

- **Collaborative Partnerships:** Consider partnerships to accelerate innovation and the adoption of AI agents.
- **Knowledge Exchange:** Share insights and best practices to accelerate agent adoption.
- **Collaborative Development:** As agent technologies are still immature, share risks and ramp up investments through joint development initiatives.

Industrial Agentic Use Cases:
From Strategy to Operation

Mapping Generative AI Agents Across the Value Chain

Here we illustrate potential use cases across **different phases of a value chain** as well as various **corporate support functions,** aiming to inspire further ideas and innovation. While the examples are **not exhaustive,** they serve as a **starting point** for exploring the diverse applications of value chain optimization.



Industrial Agentic Use Cases:
From Strategy to Operation

Use Cases Across Different Phases of a Value Chain

1 Cross-industry

Planning & Proposal
Advanced requirement management using automated quality and feasibility assessment for technical and regulatory compliance

Discovery, Exploration & Research
AI-assisted product knowledge retrieval, product design, development and validation

Marketing Strategy
Advanced market intelligence through automated sales strategy simulation based on market analysis and demand forecasting

Customer Service & Support
AI-powered assistants providing real-time support to customers based on interactions and additional internal and external data

Customer Service & Support
Personalized helpdesk handling customer requests, complaints, and incidents

Sales Channel Management
AI-based sales agents for automating personalized customer and lead interactions

Sales Channel Management
Automated sales contract execution with liability checks

2 Automotive

Discovery, Exploration & Research
AI-assisted consumer feedback and trend analysis for product development

Supplier Management
AI-based simulation of supplier and real-time supply risks

Material & Component Sourcing
AI-assisted resilience matrix creation for critical components

Outbound Logistics
Distribution center location optimization based on customer density and part availability

Process/Equipment Design & Management
Proposal of part manufacturability optimization solution before production

Production, Manufacturing, & Operation
Smart maintenance through dynamic assistance in analysis, problem resolution and documentation of incidents

Quality Control and Management
Defect detection, analysis, and prevention plan proposal, to identify/analyze recurring defects in parts and propose revised quality checks to prevent them

Customer Service & Support
Predictive car service issuing personalized notifications based on model history

3 Pharmaceuticals & Chemicals

Discovery, Exploration & Research
New chemical compound hypothesis generation (based on existing patent, experiment and market data)

Discovery, Exploration & Research
Market situation survey and competitor analysis

Discovery, Exploration & Research
Optimized research path proposals based on historical drug discovery efforts

Material & Component Sourcing
Sustainable material identification for evaluation of trade-offs in performance and cost

Outbound Logistics
Temperature-controlled shipping method suggestions to maintain drug integrity during transport.

Process/Equipment Design & Management
Analysis of outcome of formulation changes to predict risks of impacts on product stability and performance and propose mitigation plans.

Production, Manufacturing, & Operation
Dynamic lab operation optimization based on lab process monitoring

4 Semiconductor, Electronics, & Tech Products

Supplier Management
Total cost of ownership evaluation for different suppliers

Inbound/Outbound Logistics
Long-term contract optimization

Production, Manufacturing, & Operation
Chip production process monitoring and review to generate yield rate improvement directions

Quality Control and Management
Automated visual inspections with the capability to learn from previous errors and adapt over time

Sales Strategy
Targeted upsell strategies based on customer behavior data gathering and analysis

Customer Relationship Management
Automated sentiment monitoring on social media to guide marketing campaigns

5 Manufacturing, Aerospace, & Machinery

Inbound Logistics
Predicts price fluctuations for raw materials based on market trends and proposes sourcing strategies

Outbound Logistics
Analysis of latest shipping regulations to streamline logistics operations across borders

Process/Equipment Design & Management
Fabric manufacturing process simulation to reduce waste and enhance quality

Integration, Assembly & Testing
Adaptive assembly process modeling that can shift based on real-time demand

Integration, Assembly & Testing
Automated reliability testing to assess the reliability of critical components after production

6 Software, Telecommunications & E-commerce

Ideation & Concept Development
Software architecture design suggestions based on current performance metrics and user demands

Ideation & Concept Development
Analysis and proposal generation for processing Request for Proposals (RfP).

Customer Service & Support
Personalized shopping chatbots providing customized assistance by analyzing user profiles and browsing/ purchase history

Customer Service & Support
Service outage monitoring and notification for impacted users.

7 Healthcare & Education

Design, Prototyping, & Assessment
Predictive health policy impact modeling on patient care outcomes

Quality Control and Management
Material degradation simulations and proposal of new quality benchmarks

Marketing Strategy
Customized outreach and engagement campaigns based on patient demographics analysis

Sales Strategy
Tailored academic program offers based on market needs, aligning with job trends and personal interests

Customer Service & Support
Personalized appointment planner and reminder to improve patient engagement

Customer Service & Support
Personalized and interactive patient eduction in clinical studies

8 Retail, Food, & Beverage

Ideation & Concept Development
Agent-assisted new beverage concept development

Material & Component Sourcing
Stock level and reorder schedule recommendations based on supplier delivery time analysis

Inventory Management
Inventory optimization analytics based on sales patterns across regions

Outbound Logistics
Freshness-maximizing delivery route optimization to maximize freshness and reduce spoilage

Quality Control and Management
Real-time monitoring of food production processes to ensure quality and consistency through visual cue analysis (e.g., color, texture)

Customer Service & Support
Agent-powered virtual shopping assistants personalizing the customer journey through interactive question answering and recommendations

Use Cases for Corporate Support Functions

IT Service Management
- AI-driven ticket triaging and resolution suggestions
- Predictive analysis for IT incident trends
- Self-healing IT infrastructure using AI

Automated Project and Portfolio Management
- AI-driven resource allocation and optimization
- Predictive analytics for project timelines and bottlenecks
- Smart dashboard and reporting with real-time updates

Cybersecurity
- AI-driven threat detection and response
- Automated vulnerability assessment and patching
- Behavioral analysis for insider threat detection

Workplace Productivity
- Smart Scheduling and Calendar Management
- Task Prioritization and Workflow Automation
- Knowledge Management and Information Retrieval

Software Development and Operations
- Automated code generation and documentation
- Intelligent code review, optimization and bug fixes
- Automated functional testing

Employee Training
- Personalized learning plan formulation
- Virtual training assistants
- Gamified Learning Platform Operation

Financial Analysis and Reporting
- Automated bookkeeping and assessments
- Predictive financial planning
- Personalized report generation

Compliance Checks and Reviews
- Continuous compliance monitoring across multiple regulations
- Automated reporting and documentation for audits
- Intelligent risk management systems

Legal Contract Checks and Reviews
- Automated extraction and analysis of key contract terms
- Legal risk assessment using AI
- Compliance tracking and intelligent notifications

Recruiting Assistance
- Interview scheduling
- Personalized chatbot interviews
- Candidate matching

Tender Management
- Bid/No-Bid decision support
- Competitive analysis and recommendations
- Risk assessment and mitigation strategies

Research & Development
Supply Chain Management & Inbound / Outbound Logistics
Production & Operations
Customer Engagement

Industrial Agentic Use Cases: *From Strategy to Operation*

Transforming Robotic Process Automation (RPA) into Agentic Process Automation (APA)

Bringing Agents into RPA

When considering AI agent automation, a logical next step is to **integrate AI agents into existing Robotic Process Automation (RPA)** workflows [11]. This strategy provides a straightforward yet effective means of implementing agent-driven process automation.

Addressing Robustness Issues in RPA

By utilizing established problem-solving frameworks (i.e., process structures), AI agents can tackle predefined **sub-problems** and **tasks**, which they are generally more robust and capable of handling than existing RPA methods. This approach not only mitigates the current challenges associated with orchestrating and managing AI agents but also effectively addresses a key limitation of RPA: **robustness**, hence a valuable advancement.

SLMs as Cost-effective Options

In these scenarios, system designers may opt for Small Language Models (SLMs) over Large Language Models (LLMs) to reduce costs and infrastructure needs. For instance, agentic RPA workflows could run on **standard servers with less expensive GPUs.**

	Pipeline Design	Pipeline Building	Pipeline Execution	Pipeline Monitoring, Evaluation, & Update	Pros	Cons	When To Adopt
RPA Using manual-crafted rules to orchestrate several software in a solidified workflow for execution	Static and simple step-by-step workflows	Manually contructing via pull-and-drag	Rule-based data-flow and control-flow	Fixed to defined scenarios and unable to update instructions	Can handle rigid task	Cannot handle flexible task	Well-defined sub problems and tasks
Use Case Example Insurance Claim Processing	• **Focus:** Create workflows using defined rules to handle repetitive tasks such as extraction of names and dates. • **Structure:** Simple flowcharts illustrating the step-by-step process for claim data entry, verification, and updates.	• **Resources Required:** IT expertise to configure bots and integrate them with existing claim processing systems. • **Development Work:** Through manual robotic procedure setup and review process.	• **Mechanism:** Bots execute predefined tasks like claim data entry, validations, and standard communications. • **Speed:** High volume processing of claims at a consistent rate once programmed. • **Consistency:** Executes tasks exactly as programmed, with minimal deviation.	• **Monitoring:** Rule-based monitoring through claim processing logs and alerts to identify failures or bottlenecks. • **Evaluation Frequency:** Periodic reviews for performance and efficiency. • **Update Process:** Manual updates required for any changes in workflows or system integrations.	• **Efficiency:** Greatly reduces claim processing time for routine tasks. • **Cost-Effective:** Significant savings on claim processing labor for repetitive tasks. • **Scalability:** Easy to scale operations up or down as needed.	• **Limited Flexibility:** Struggles with unstructured data and unexpected scenarios. • **Maintenance Requirement:** Requires periodic manual updates to adapt to new processes. • **Lack of Insight:** Doesn't analyze data for patterns or insights beyond predefined tasks.	• **Indicative Scenarios:** High volume, repetitive claim processing tasks with clear rules; ideal for back-office functions. • **Best Fit:** Claim processes with low variance and where performance can be measured without needing complex decision-making.
	• **Focus:** Design workflows that incorporate intelligent decision-making and adapt to changing scenarios and formats of claim contents. • **Structure:** Adaptive flowcharts that can change based on real-time data analytics and learning from previous claims processing.	• **Resources Required:** Cross-functional teams including data scientists, machine learning experts, and process analysts to develop and refine algorithms for claim processing. • **Development Work:** Through generalized and yet flexible agentic modules and goal-driven adaptable autonomy.	• **Mechanism:** Intelligent bots assess incoming claims, make decisions based on past claim history, and take actions that can change dynamically. • **Speed:** Faster decision-making, especially for complex claims, as bots learn and adapt. • **Consistency:** Maintains accuracy over time by learning from feedback instead of strictly adhering to initial programming.	• **Monitoring:** Human oversight together with automated monitoring through analytics tools that evaluate bot performance and decisions. • **Evaluation Frequency:** Potential real-time evaluations for instant adjustments where necessary. • **Update Process:** Self-updating capabilities based on learned data and analytics to continuously improve the workflow.	• **Adaptability:** Can handle complex, variable tasks required in the claims by evolving based on historical data. • **Enhanced Reasoning and Decision-Making:** Improved accuracy and responsiveness to subtle claim scenarios. • **Customer Experience:** Offers personalized services and faster claim resolution.	• **Complex Implementation:** Requires substantial upfront investment in technology and talent. • **Data Dependency:** Performance reliant on the quality and quantity of available reference data. • **Risk of Errors:** Potential risk with AI agent biases that lead to incorrect decisions.	• **Indicative Scenarios:** Claim processes requiring adaptability and deep analysis; suited for complex claims with varied outcomes. • **Best Fit:** Claim processes that are infrequent, unpredictable, and necessitate sophisticated reasoning and decision-making.
APA Incorporating AI agents to adaptively contruct and execute workflows to achieve process automation	Dynamic and scenario-adaptive workflows	Automatically constructing, orchestrating, and testing	Agent-based data-flow and control-flow	Adaptable to various scenarios and able to update instructions	Can handle rigid and flexible tasks	Monitoring and verification may be tricky	Ill-defined sub problems and tasks

Industrial Agentic Use Cases: From Strategy to Operation

The Growing Need for AgentOps

A Call for Long-term Agent Mangement

AgentOps is an emerging concept focused on the **operational management** of generative AI agents, which are often characterized by their ability to autonomously perform tasks, interact with users, and generate content based on user inputs or external data sources [12].

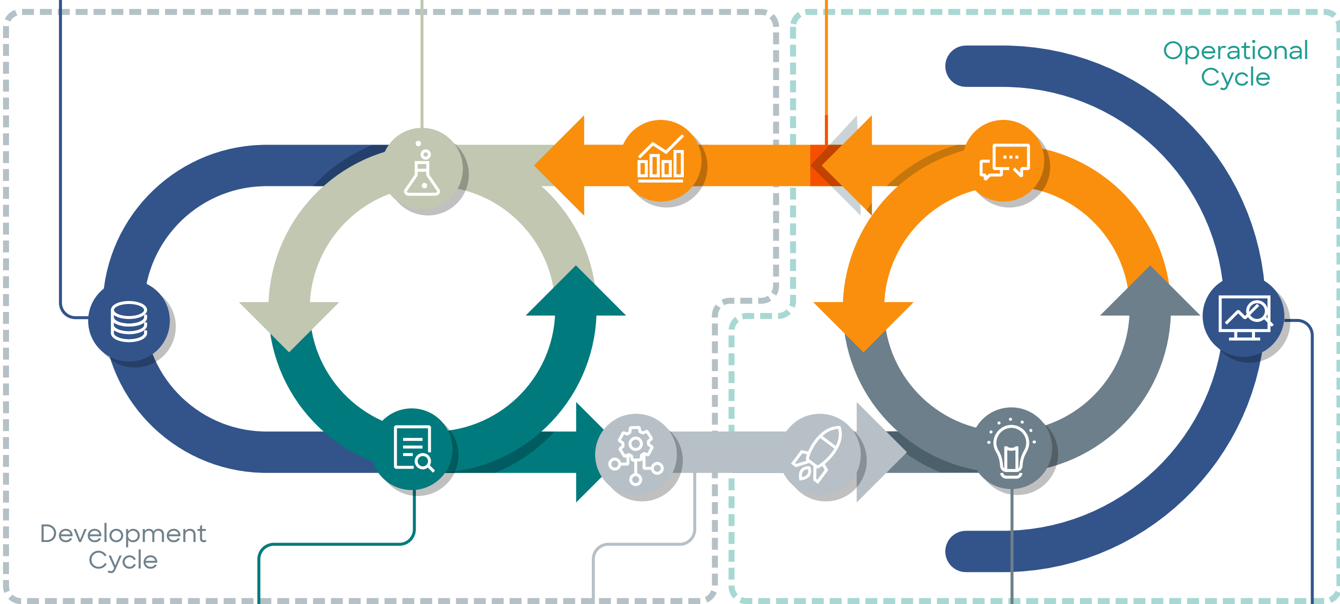
Towards Trustworthy Agents

As generative AI agents become increasingly complex and capable, establishing clear operational practices becomes crucial for ensuring their **reliability**, **effectiveness**, and **ethical deployment** [12-13].

1. Problem Definition, Agent Design, Tool/Model/Data Engineering
Define use cases, design agent prompts and co-working architecture, and prepare relevant tools, models (LLMs/SLMs) and datasets.

2. Experiment
Conduct experiments to train or fine-tune the underlying models of the agent, testing various tool or agent configurations and determine the most effective setup.

7. Feedback & Iterative Improvement
Collect user feedback and performance data to identify areas for enhancement. Use this information to iteratively improve the agent's capabilities, retraining or adjusting as needed to optimize its performance over time.



3. Evaluate
Assess the agent's performance using predefined metrics and benchmarks, possibly leveraging external agents to evaluate its output (Agent-as-a-Judge) for consistency and relevance.

4. Integration & Deploy
Integrate the generative AI agent into the necessary systems and workflows. Deploy the agent in a production environment where it can interact with users and perform its intended tasks.

5. Inference & Execution
Enable the agent to process inputs and generate outputs in real time. Activate its core functionalities, allowing it to perform the tasks for which it was designed.

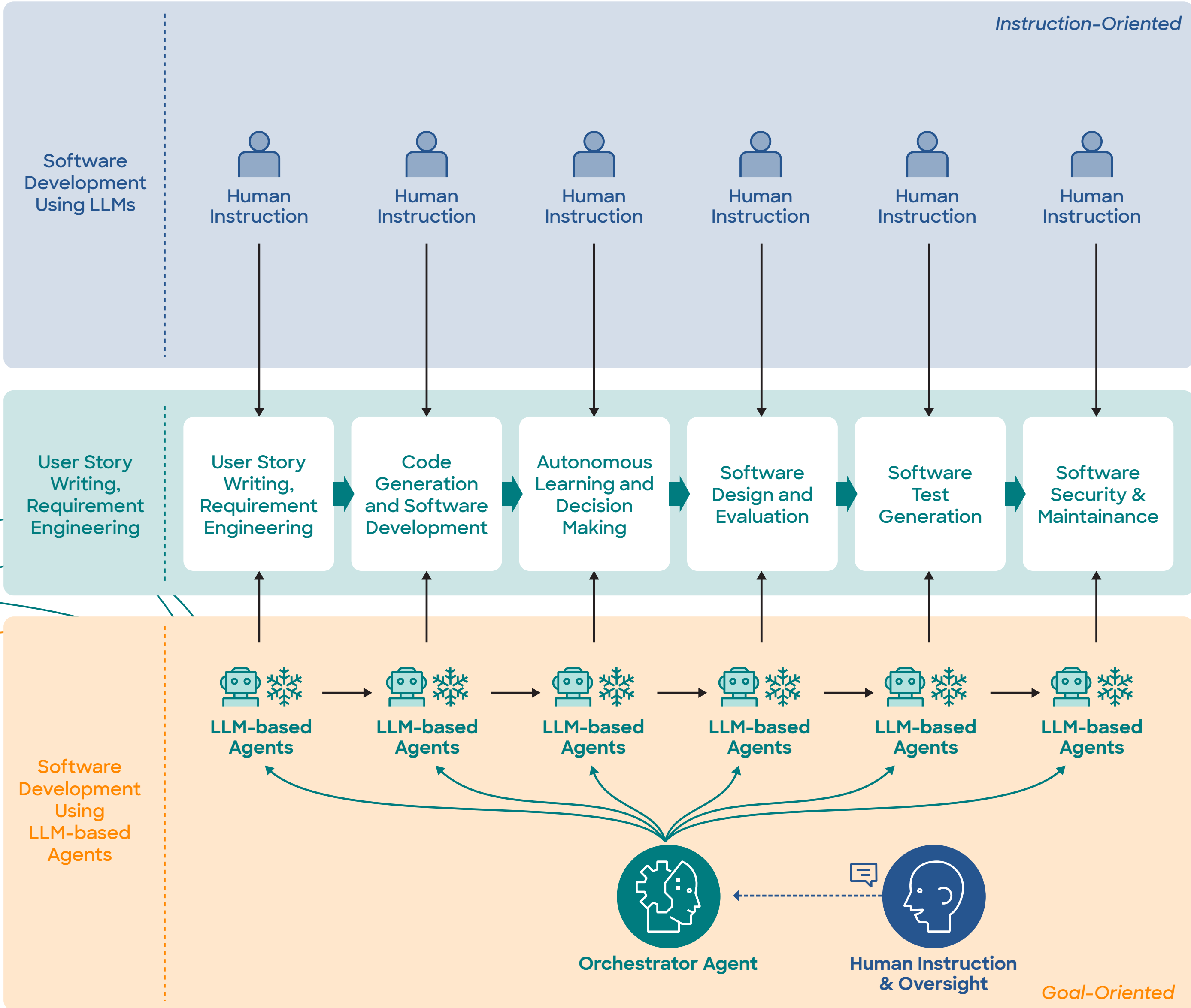
6. Monitoring & Logging
Continuously monitor the agent's performance and usage, recording interactions and outputs. Implement step-by-step trace-back mechanisms to analyze any issues or performance deviations effectively.

Partially inspired and adapted from: https://learn.microsoft.com/en-us/ai/playbook/solutions/generative-ai/llmops-promptflow

Focusing the Lens: *Generative AI Agents in Software Development*

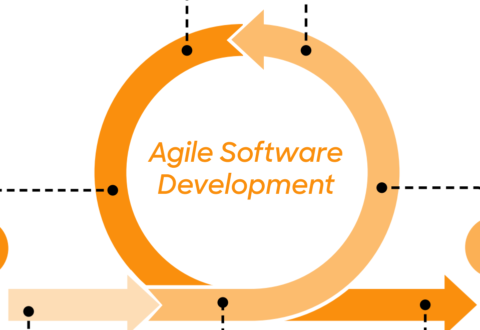
From LLMs to LLM-based Agents in Software Development

With the emergence of LLMs and generative AI, their applications are being extensively investigated across different industries. A significant area of focus is **software development**, where LLMs have demonstrated impressive capabilities in tasks like **code generation** and **test design.** Despite these achievements, they also face several limitations, particularly regarding autonomy. **LLM-based agents** utilize LLMs as the foundation for **planning, designing, decision-making,** and **executing actions** during software development, thereby overcoming some of the prior constraints. In this section, we highlight the main distinctions between these two approaches [14].



Focusing the Lens:
Generative AI Agents in Software Development

Navigating the Agentic Software Development Cycle: Use Cases and Trust Spectrum



1. Plan

1. Requirement Gathering & Analysis: Analyze stakeholder inputs and extract key requirements.

2. User Story Generation: Create detailed user stories based on high-level requirements.

3. Project Planning: Assist in timeline estimation and resource allocation.

4. Risk Assessment: Identify potential risks and suggest mitigation strategies.

5.Backlog Prioritization: Help prioritize backlog items based on various criteria like business value and effort.

2. Design

1: Architectural Suggestions: Provide recommendations for system architecture based on requirements.

2: UI/UX Design Assistance: Generate wireframes or design mockups from textual descriptions.

3: Design Document Drafting: Create comprehensive design documents outlining components, interfaces, and data flow.

4: Component Specification: Define specifications for individual system components.

5: Best Practices Guidance: Offer insights into industry best practices and standards relevant to the project.

3. Develop

1: Code Generation: Write code snippets or modules based on specifications or descriptions.

2: Documentation: Automatically generate or update code documentation and comments.

3: Code Refactoring: Suggest improvements for existing code to enhance readability and performance.

4: API Integration Assistance: Provide guidance or generate code for integrating third-party APIs.

5: Language Support: Assist developers with syntax, libraries, and frameworks in various programming languages.

4. Test

1: Test Case Generation: Create comprehensive test cases based on user stories and requirements.

2: Automated Test Scripts: Develop scripts for automated testing frameworks.

3: Bug Detection Assistance: Analyze code to identify potential bugs or vulnerabilities.

4: Test Documentation: Generate test plans, reports, and documentation.

5: Performance Testing Insights: Provide recommendations for performance testing scenarios and metrics.

5. Deploy

1: Deployment Script Generation: Create scripts for deploying applications to various environments (e.g., staging, production).

2: Configuration Management: Suggest optimal configurations based on application requirements.

3: Continuous Integration/Continuous Deployment (CI/CD) Pipelines: Assist in setting up and optimizing CI/CD workflows.

4: Environment Setup Assistance: Provide guidance on setting up development, testing, and production environments

5: Rollback Strategy Documentation: Generate plans for rolling back deployments in case of failures.

6. Review

1: Code Review Assistance: Automatically review code for adherence to standards, best practices, and potential issues.

2: Feedback Analysis: Analyze feedback from team members to identify common themes or areas for improvement

3: Performance Metrics Reporting: Compile and interpret performance data to assess project progress.

4: Documentation Review: Ensure that all project documentation is up-to-date and comprehensive.

5: Retrospective Facilitation: Generate questions and topics for retrospective meetings to encourage constructive discussions.

7. Launch

1: Marketing Content Creation: Generate marketing materials, blog posts, and release notes for the product launch.

2: User Onboarding Assistance: Create guides, tutorials, and FAQs to help users understand and adopt the new software.

3: Feedback Collection Tools: Design surveys or feedback forms to gather user input post-launch.

4: Launch Plan Documentation: Draft comprehensive launch plans outlining steps, responsibilities, and timelines.

5: Monitoring Setup Guidance: Provide recommendations for setting up monitoring and analytics tools to track the software's performance post-launch.

Trust and Readiness for Adoption:
Joint Assessment Among Selected appliedAI Partner Companies

Weaker — Stronger

Generative AI Agents in Action for Software Testing

Enhancing Software Development Testing: Why Generative AI Agents Matter

Background

Continuous software testing is a critical element of the software development lifecycle, especially within agile methodologies, where testing occurs at every stage to ensure system robustness as new code is committed to repositories like GitHub, often facilitated by tools like Jenkins for **CI/CD** processes.

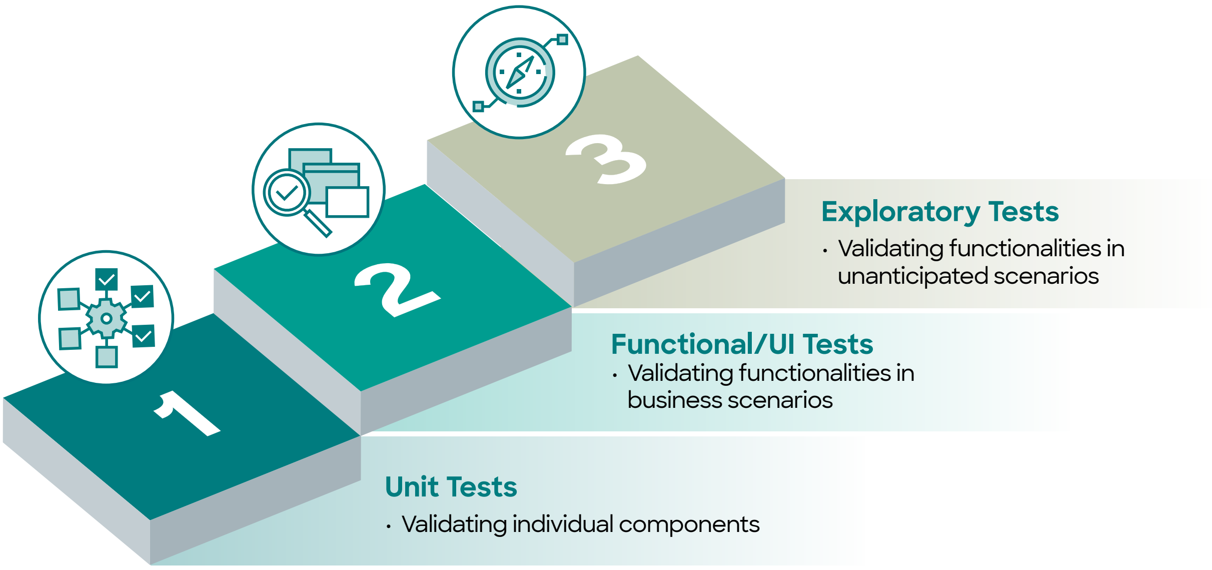
Challenge

Despite advancements in automation, the creation and refinement of test cases—such as **unit tests** and **functional/UI tests** —still require substantial human effort, leading to inefficiencies and potential gaps in testing coverage.

Opportunity

There is a significant opportunity to leverage AI to automate the **generation** and **optimization** of test cases, thereby reducing the manual workload on developers, enhancing testing efficiency, and improving the overall quality of software products.

Levels of Software Testing Automation



Generative AI Agents in Action for Software Testing

Generative AI Agents for Unit Test Writing & Reviewing

Problem Statement

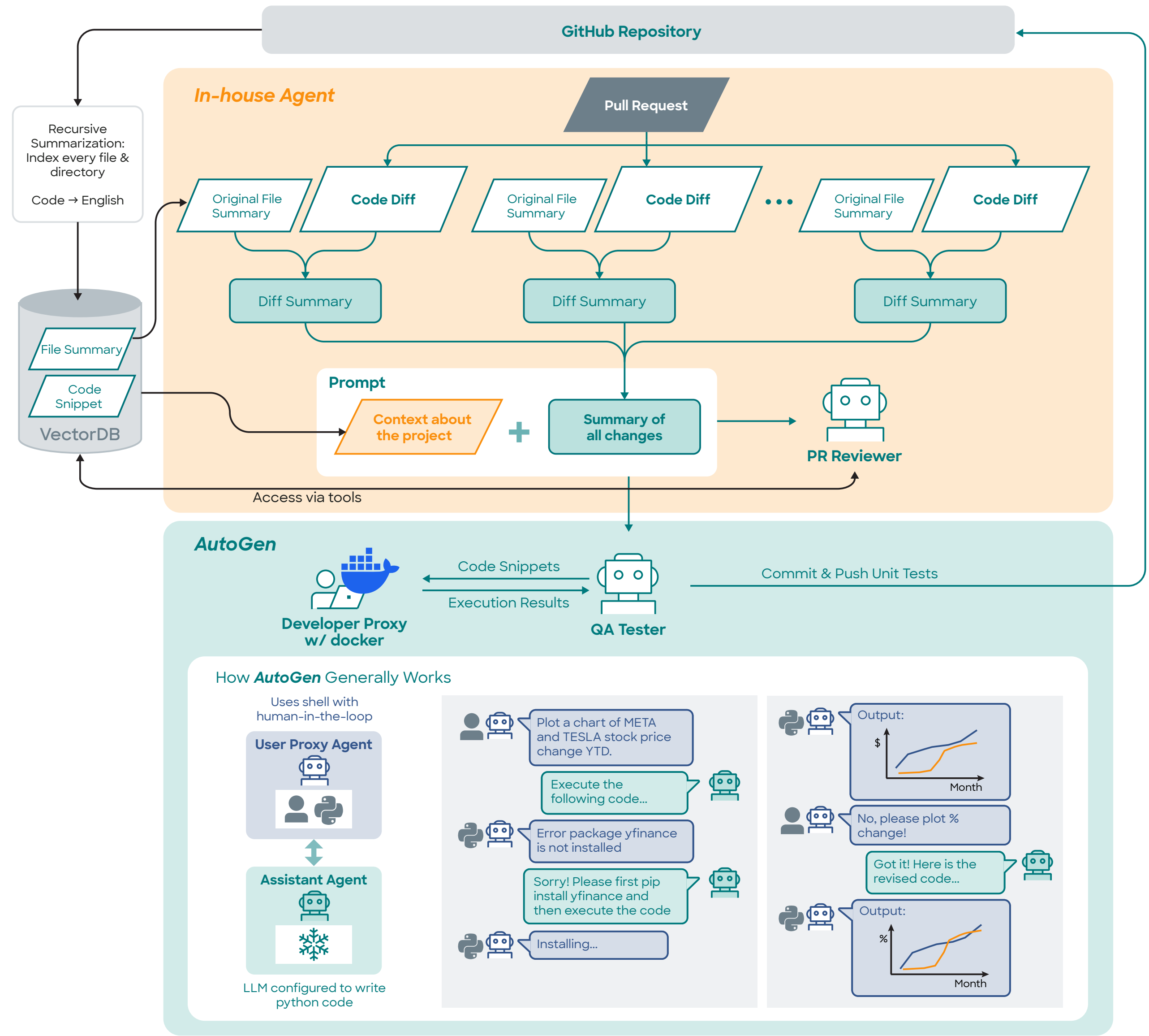
Writing and reviewing unit tests can be **time-consuming** and **error-prone**, often requiring deep domain knowledge and meticulous attention to edge cases.

Current Solution

Generative AI agents can automate the generation of unit tests by analyzing **code logic,** identifying **critical paths,** and suggesting tests for **edge cases** and **coverage gaps.** In this white paper we showcase an innovative autonomous Pull Request Tester and Reviewer.

Methods

- Our tool integrates **Retrieval-Augmented Generation (RAG)** and **AutoGen** technologies to automatically review GitHub pull requests. It generates **summaries** for the code, **analyzes code differences**, and provides concise summaries of those differences. Using these summaries, the tool conducts in-depth **reviews** of the pull requests, assessing code quality and functionality while also creating effective **unit tests**.
- By leveraging **generative AI agents** to interpret requirements and code structure, the system can generate test cases, validate them against expected behaviors, and provide feedback or improvements to existing test suites.



AutoGen diagram partially adapted from: https://microsoft.github.io/autogen/0.2/docs/Getting-Started/

Generative AI Agents in Action for Software Testing

Generative AI Agents for Functional/UI Tests

Background & Methods

Problem Statement

Developing **functional and UI tests** is labor-intensive and requires detailed knowledge of **user flows, interface interactions**, and **system functionality**, making it a high-cost process for the company in terms of business value.

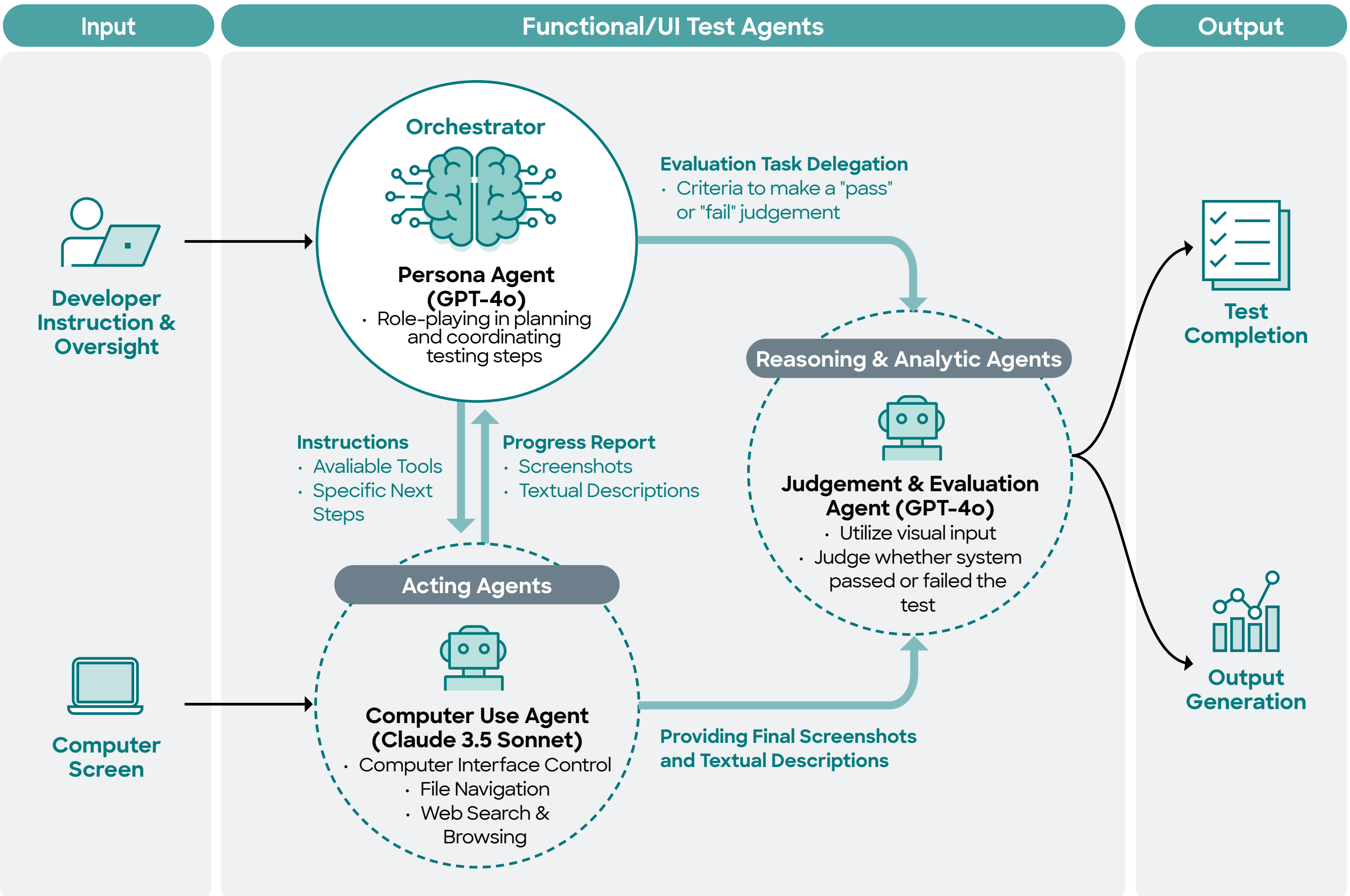
Current Solution

Generative AI agents can **automate** functional and UI test **creation, execution** and **validation** by analyzing **workflows**, **user stories**, and **interface designs**, reducing manual effort while enhancing test accuracy and consistency. Here we showcase a multi-agent autonomous UI function tester.

Methods

- We use an **orchestrating agent** with AI generated **persona** to automate end-to-end functional/UI tests, in combination with the **acting agent** based on Claude 3.5 Model and a visual large language model **judgement and evaluation agent**.
- Using this multi-agent system, AI agents can **simulate** user interactions, **generate** plans and instructions for automated functional / UI tests, and **adapt** these tests dynamically based on changes in the application's UI.

Multi-agent Functional/UI Testing Workflow



Generative AI Agents in Action for Software Testing

Demonstration Test Cases

For demonstration purposes during the prototypical phase, we have chosen an internally developed software, '**GenAI.xy Playground**,' as the target for software testing. This selection allows us to assess, based on its current capabilities (such as reasoning, planning, and UI execution), **the level of complexity** (functional level) that our designed multi-agent system can handle.

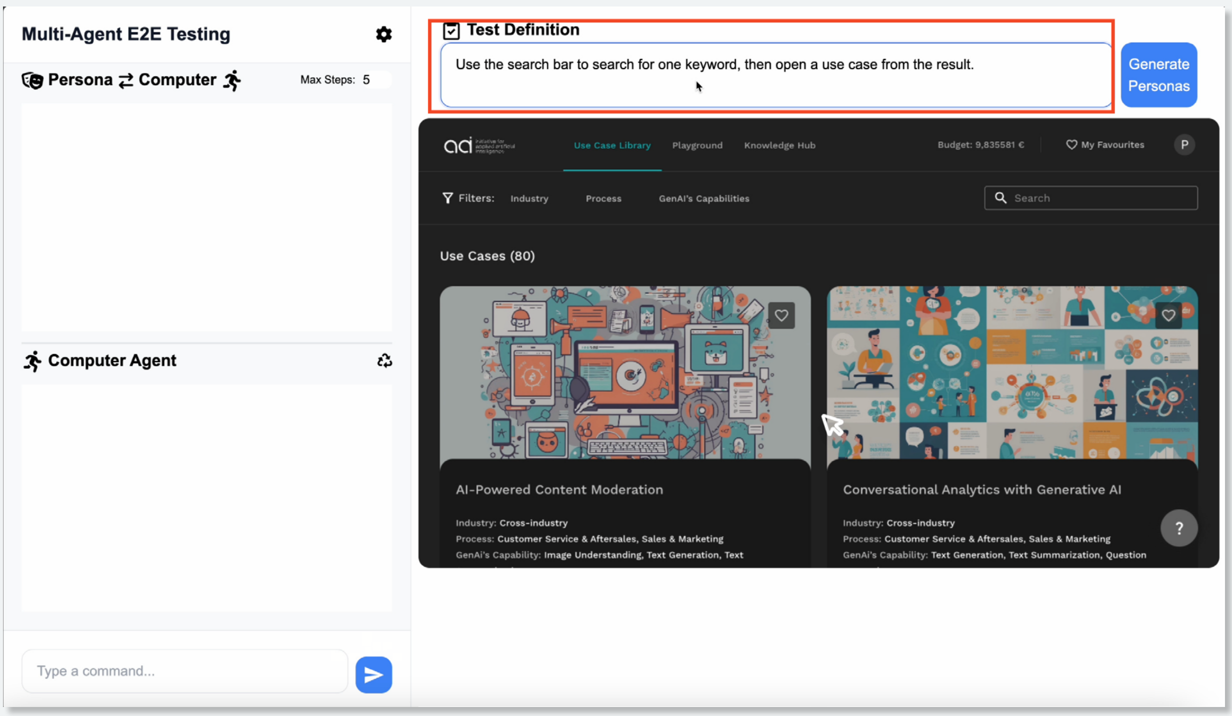
The table below presents the GUI-based functions that have been tested with the prototype as well as the testing results. Refer to "*Multi-agent Functional/UI Testing Workflow*" on the previous page for a diagram of the prototyped system and the appliedAI Initiative Youtube channel for live demo recordings.

Test case	Goal Description	Testing result	Judgement Agent result
Search bar	Use the search bar to search for one keyword, and then open a use case from the result	Acting agent based on Claude 3.5 Computer Use correctly understood the visual design of a search bar	PASS
Favourite (bookmark) feature **Filtering (checkbox with dropdown menu)**	In the Use Case Library, use the filtering feature to select 1 industry and find one use case to add to favorites.	Acting agent based on Claude 3.5 Computer Use was able to correctly find filter on UI and also understand the visual design that a heart❤️means adding to favourite.	PASS
Using image generation model with a prompt	Open Image Generation in the Playground, and then generate a cute image and add it to favorite.	Agent correctly navigated to the image generation playground and found correct model.	PASS
Search and apply 1 specific prompt template	Find one user-specified prompt template in the code generation playground which can assist in writing RESTful API and test that template.	Acting agent found the correct required template, but after reviewing the template and getting back to the main menu, it forgot the context and applied the wrong template.	FAIL

Example Agentic Software Test Workflow for the Test Case 'Seach Bar'

Step 1 Define the Test Case 'Search bar'

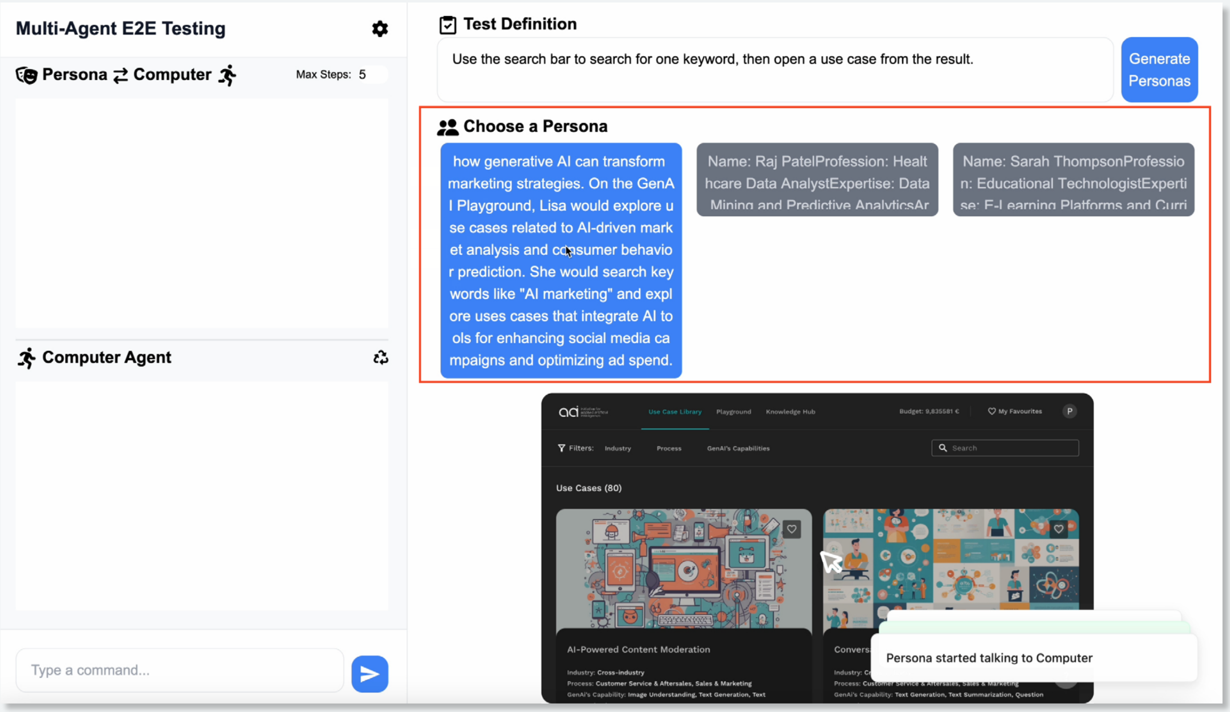
- The user and developer define the functional test case goal – in this example, testing the search bar functionality – by appropriately describing the test case.



Generative AI Agents in Action for Software Testing

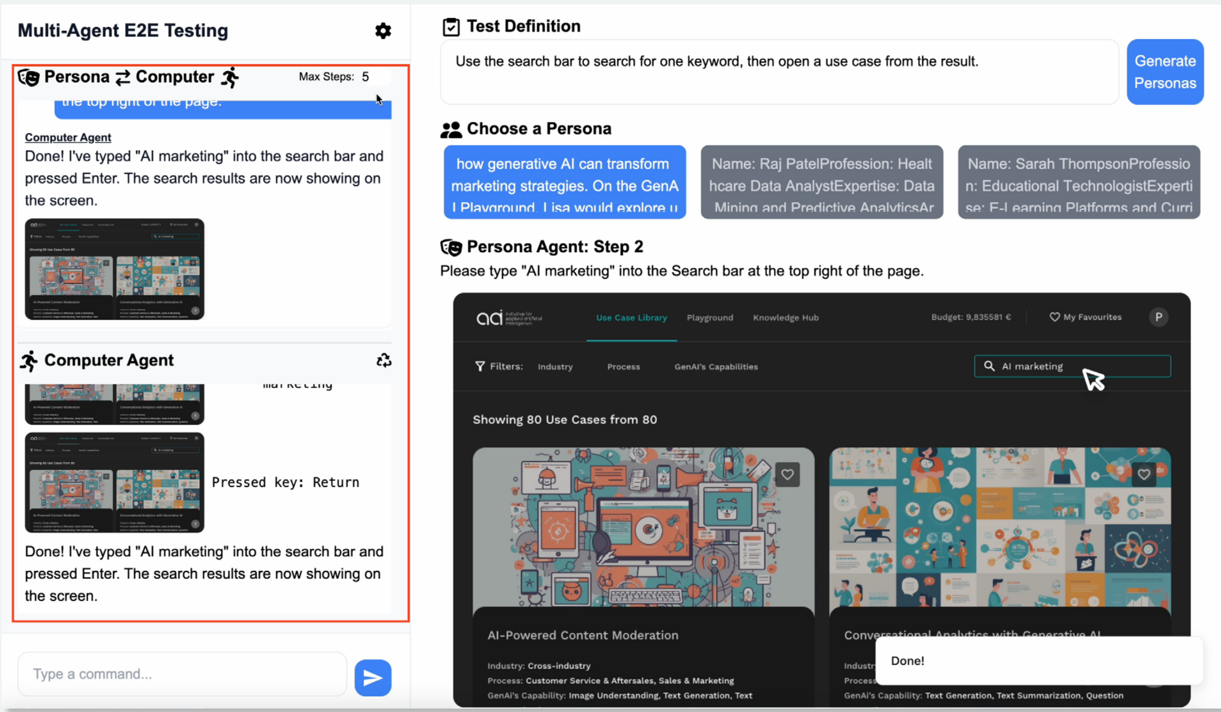
Step 2 **Initialize the Persona Agent**

- The Orchestrator (Persona Agent) generates three persona templates based on the given test case definition and awaits the user's selection.



Step 3 **Orchestrator Collaborates with the Acting Agent**

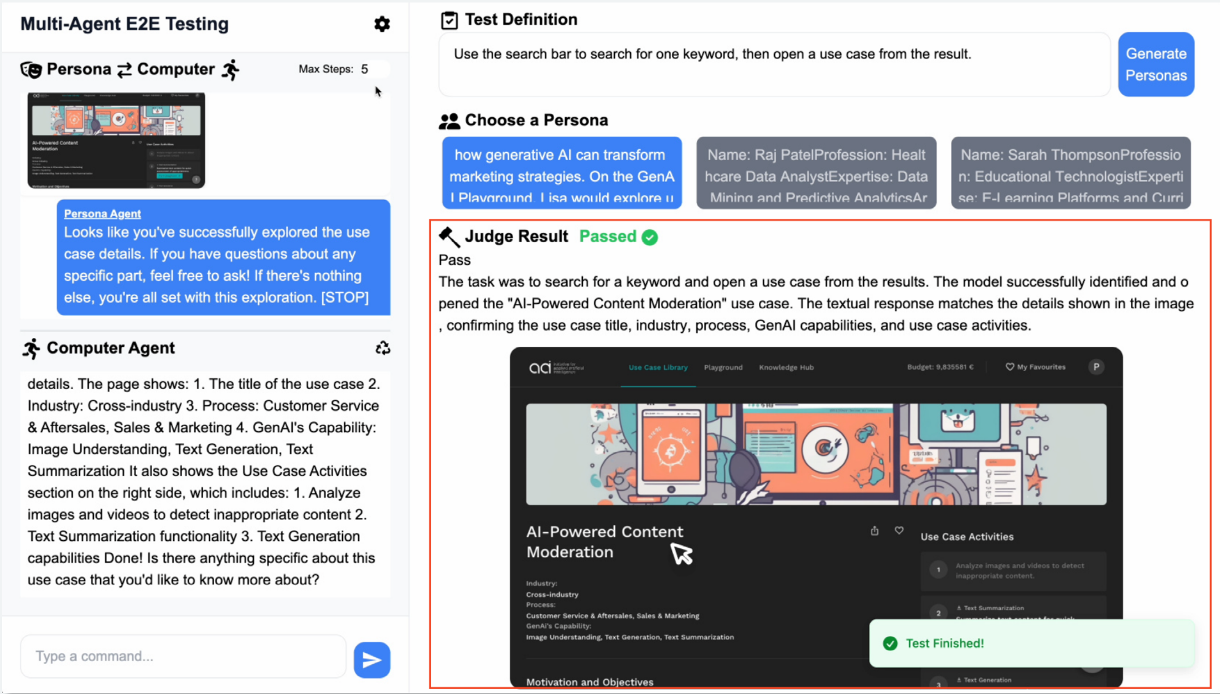
- The Orchestrator selects the appropriate persona and begins planning and delegating tasks to the Acting Agent.
- The Acting Agent will request assistance or instructions from the Orchestrator when blocked by a subtask and will also report major milestones to the Orchestrator, providing both text descriptions and UI screenshots.



Generative AI Agents in Action for Software Testing

Step 4 **Orchestrator Calls the Judgement Agent for Evaluation**

- Once the Orchestrator "believes" that the test has been completed by the Acting Agent, it will call the Judgement Agent for a final evaluation.
- The Judgement Agent will analyze and evaluate, based on the final screenshot and last report, whether the original test goal given by the user has been completed **(PASS)** or not **(FAIL)**.



Retrospective & Prospective:
Challenges and Opportunities for Generative AI Agents

Insights & Reflections

Agent Capabilities

Defining Characteristics: Generative AI agents exhibit key traits including environmental interaction, task execution, and advanced cognitive capabilities—spanning perception, reasoning, goal setting, and planning—allowing for adaptive responses to complex scenarios.

Current Status: Presently, agentic functionalities are primarily concentrated at level 2 (reasoning), with emerging advancements at level 3 (autonomy), underscoring a landscape ripe for the evolution of more sophisticated cognitive processes in future iterations.

Prospects: In the coming years, we anticipate breakthroughs in agentic innovation and organizational capacity, leading to the deployment of multi-agent systems that facilitate enhanced communication and orchestration among individual agents, while emphasizing the ongoing necessity for human oversight.

Industrial Agentic Use Cases

Transformative Automation Across the Value Chain: Generative AI agents offer substantial opportunities to enhance business decision making processes by automating and simplifying complex tasks, enabling deep semantic understanding and driving efficiencies across diverse processes – from research & development to customer engagement.

Holistic Value Creation: By facilitating processes such as creativity, discovery and research, optimizing logistics in supply chain management, and streamlining operations and quality control in production, generative AI agents contribute to holistic value creation and operational excellence.

From Exploration and Engagement to Execution and Effective Operation: Currently, the trajectory of generative AI applications tends to prioritize the initial exploratory and final customer engagement stages. However, there is also growing effort in the domains of supply chain management and production operations, aimed at developing robust automation tools for critical processes in the future.

From RPA to APA

Boosting Dynamicity and Adaptability: Transitioning from Robotic Process Automation (RPA) to Agentic Process Automation (APA) enables a shift from rigid, rule-based workflows to dynamic, goal-oriented frameworks that adapt to varying task complexities, enhancing overall process efficiency and effectiveness.

Augmenting Existing Workflows: By incorporating AI agents into existing RPA workflows, organizations can leverage agentic problem-solving capabilities to tackle both predefined and ill-defined sub-problems, thereby addressing RPA's limitations in adaptability and robustness.

Cost-Effective Deployment: Utilizing Small Language Models (SLMs) over Large Language Models (LLMs) in agentic workflows allows organizations giving self-hosting options on cloud, on-premise, and edge environments to optimize data privacy, IT integration and infrastructure costs, while still harnessing the flexibility and intelligence of AI agents.

Software Development

From LLMs to LLM-based agents: The transition from LLMs to LLM-based agents is reshaping software development, with current applications focusing on code generation, unit/functional testing, and requirements engineering, while cautious adoption persists for critical tasks like infrastructure deployment.

Prioritizing Safe AI Integration: High-value, low-cost use cases in agile development, such as documentation and UI/UX design support, are being prioritized for AI agent integration, whereas critical activities like deployment script generation in complex environments are viewed as risky and less mature.

Impact and Trust Across Roles: AI agents are expected to influence various software engineering roles. Areas such as frontend and web development and software testing are currently the most trusted for AI automation and may see the most impact. In contrast, task planning and deployment are less trusted. Nonetheless, the complexity of enterprise software development may pose challenges.

Software Testing

Revolutionizing Continuous Testing: Generative AI agents streamline the software development lifecycle by automating the creation and refinement of unit, functional, and UI tests, enhancing efficiency and ensuring robust testing coverage integral to agile methodologies.

Mitigating Human Effort in Test Generation: By leveraging advanced techniques like Retrieval-Augmented Generation (RAG) and AutoGen, AI agents automatically review pull requests and produce tailored test cases, significantly relieving developers from the time-consuming and error-prone task of manual test writing.

Enhancing Test Accuracy and Adaptability: With the ability to analyze user workflows and interface interactions, generative AI agents facilitate the dynamic creation of functional and UI tests, ensuring comprehensive coverage while adapting to evolving application requirements and maintaining consistency across testing efforts.

Next-Gen Potentials

Transformative Problem-Solving: Advanced generative AI agents are revolutionizing problem-solving in diverse fields like software development and industrial engineering by automating complex tasks, enhancing collaboration, and producing high-quality solutions through dynamic interaction and learning from human feedback.

World Simulation Applications: Generative AI agents may assist in simulating human behavior across gaming, societal interactions, and economic modeling, enabling realistic role-playing, engaging dialogue, and strategic decision-making that closely mimics human responses and social dynamics.

Autonomous Scientific Innovation: Generative AI agents will drive significant advancements in scientific research by autonomously conducting experiments, optimizing processes, and facilitating collaborative debates, thereby enhancing the efficiency and accuracy of scientific inquiry across various disciplines.

Retrospective & Prospective:
Challenges and Opportunities for Generative AI Agents

Challenges & Risks of Generative AI Agents

Adversarial Robustness

LLMs Under Attacks
- Large language models are susceptible to adversarial attacks, leading to erroneous responses. Relevant attack methods include dataset poisoning and prompt-specific attacks.

In Pursuit of Robustness Techniques
- Approaches such as adversarial training, data augmentation, and sample detection can enhance the robustness of LLM-driven agents; however, a complete solution continues to be elusive.

Human Oversight Required
- Introducing a human-in-the-loop framework can help oversee and improve the conduct of LLM-dependent agents, which may reduce the threats posed by adversarial attacks.

Trustworthiness

Calibration Challenges
- Language models face challenges with the so-called calibration problem, which causes them to inadequately convey the certainty of their predictions, leading to outputs that do not reflect human expectations in practical use cases.

Demand for Reliability
- There is an urgent demand for intelligent agents that are both reliable and honest. Recent studies have focused on directing models to offer reasoning and explanations to improve their credibility.

Debiasing and Fairness
- Implementing debiasing strategies and calibration methods during the training process can address fairness concerns and improve the reasoning capabilities of language models.

Misuse, Bias, & Fairness

Exploitation of LLM Agents
- Individuals with malicious intentions can exploit LLM-based agents to sway public perception, disseminate misinformation, and conduct unlawful activities.

Dangers to Security and Society
- The potential for abuse of generative AI agents presents considerable dangers to both security and social stability, which could lead to orchestrated terrorist activities and cyber threats.

Regulatory Measures for Safe Use
- To reduce these risks and promote responsible usage, it is crucial to implement strict regulatory frameworks and improve security protocols in the development and training of these agents.

Human-agent Interaction

Communication Clarity Needed
- Clear communication between humans and AI agents is essential, as misunderstandings can occur due to the intricacies of the agents' language models, potentially resulting in unintended outcomes in decision-making.

Impact of AI Reliance on Human Cognition
- As people depend more on AI agents for decision-making, there is a concern that this reliance could weaken critical thinking and problem-solving abilities, which might compromise human agency.

Building Trust and Ethics
- Establishing trust in interactions between humans and AI is crucial; users need to have confidence in the agents' abilities while ensuring that ethical standards are upheld to prevent manipulation or exploitation.

Agent Evaluation

Real-World Performance Limitations
- Existing approaches to assessing AI agents often fall short in accurately reflecting their performance in real-world scenarios, resulting in a limited understanding of their reliability.

Bias and Fairness in Evaluation
- When evaluating AI agents, it is crucial to address issues of bias and fairness; using inappropriate evaluation metrics can exacerbate undesirable behaviors, undermining the agent's acceptance in society.

Evolving Assessment Frameworks
- As both environments and tasks may change, it is essential for the evaluation of AI agents to evolve, enabling ongoing measurement of their performance while ensuring they remain aligned with user requirements and ethical standards over time.

Threat to the Well-being of the Human Race

Challenges in Managing Agents
- As AI agent technology progresses, humans may find it challenging to manage these systems, which could result in considerable risks if these agents surpass human intelligence and develop their own objectives.

Global Safeguards Imperative
- Without adequate safeguards, sophisticated AI agents could pose significant dangers to humanity, underscoring the need forregulations and a globally shared technical and ethical framework.

Economical Impact
- The advancements of AI agents may disrupt traditional job markets, necessitating workforce reskilling and adaptation to ensure that the benefits of this technology are equitably distributed across society.

References: [5,7-8]

Retrospective & Prospective:
Challenges and Opportunities for Generative AI Agents

The Future Unfolded: Opportunities of Generative AI Agents Today and Beyond

Today's Innovations: Generative AI Agents Taking Actions

Rapid Development of Agent Ecosystems

Tech leaders (e.g., NVIDIA, OpenAI, SAP) are actively pursuing the creation and integration of generative AI agents into **exisiting frameworks, tools, and ecosystems**, laying the groundwork for an expansive agent society.

From Single-Agent to Multi-Agent Systems

Companies are gradually transitioning from single-agent approaches to **orchestrated multi-agent systems**, assessing how Level 2 (reasoning) and Level 3 (autonomy) agentic capabilities may be integrated to tackle complex tasks end-to-end with a high degree of autonomy.

Focus on Efficiency in Both Early- and Late-Stage Value Chains

Initial applications of generative AI agents are likely to concentrate on enhancing efficiency in **use cases** across both the **early-** and **late-stage value chains** of various sectors, streamlining processes, and improving productivity.

Start with Agentic Process Automation (APA)

For various business processes, a structural approach such as APA can be implemented to effectively leverage advanced agent capabilities, accommodating diverse task complexities and thereby enhancing and **augmenting existing robotic process automation (RPA) workflows.**

Advancements in Software Development Tools

Both generative AI code assistants and engineering agents are gaining traction, demonstrating **potential in automating critical software development** tasks such as code reviews and functional testing, although concerns about trust and reliability remain.

Theoretical Innovations and Cross-disciplinary Approaches

Researchers are advancing theoretical knowledge by integrating insights from fields such as **cognitive science and complex systems,** enhancing understanding and application of generative AI agents in various contexts.

"Agentic AI's transformative power shines with customization. By designing purpose-built agents for specific domains, we can now blend advanced reasoning and actionability with modern techniques like RAG, Knowledge Graphs, Conversational Analytics, and Intelligent Document Processing, crafting AI systems that excel at tackling complex, domain-specific challenges."

Milos Rusic
CEO & Co-Founder
deepset

Retrospective & Prospective: *Challenges and Opportunities for Generative AI Agents*

Beyond the Horizon: Generative AI Agents Shaping the Future

Enhanced Collective Intelligence and Coordination

Research will likely explore optimizing **collective intelligence** within AI agent networks, where multiple agents accumulate knowledge and experience from both **interactions** among themselves and their **collaborations** with humans, achieving synergies that can lead to more effective problem-solving and innovation.

Evolving System Interconnection and Complexity

By forming interconnected multi-agent systems, an **"agentic galaxy,"** the complexity of these networks may increase significantly, fostering continuous learning and adaptability as well as allowing agents to rapidly evolve through shared insights.

Progress in Multi-Modal Environments

There will be an increased focus on generative AI agents in **multimodal** settings, which will integrate various **sensory inputs**. These versatile agents will enhance their ability to interact with the physical world, leading to more natural and effective human-robot interactions across diverse industries.

From Virtual to Physical Agents

The "ChatGPT moment" for **multimodal robotic foundation models** is approaching, enabling predicted actions in complex physical environments and ultimately realizing the long-term vision of human-like intelligence in robotic form.

Scalability and Resource Efficiency

The future of generative multi-agent systems will depend on developing **scalable architectures** that maintain **efficiency** as the number of agents increases, addressing computational constraints.

Extensive Applications Across Diverse Fields

Generative AI multi-agent systems are expected to expand into **various industrial sectors** (semiconductor, chemicals, E-commerce, healthcare, education, etc.), tackling complex problems and driving advanced computational solutions.

References

[1] T. Sumers, S. Yao, K. Narasimhan, and T. L. Griffiths, “Cognitive Architectures for Language Agents,” Sep. 05, 2023, arXiv: arXiv:2309.02427. doi: 10.48550/arXiv.2309.02427.

[2] S. Franklin and A. Graesser, “Is It an agent, or just a program?: A taxonomy for autonomous agents,” in Intelligent Agents III Agent Theories, Architectures, and Languages, vol. 1193, J. P. Müller, M. J. Wooldridge, and N. R. Jennings, Eds., in Lecture Notes in Computer Science, vol. 1193. , Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 21–35. doi: 10.1007/BFb0013570.

[3] L. Yee, M. Chui, and R. Roberts, “Why AI agents are the next frontier of generative AI | McKinsey,” McKinsey Quarterly. Accessed: Dec. 11, 2024. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-agents-are-the-next-frontier-of-generative-ai

[4] A. Gutowska, “What Are AI Agents? | IBM,” IBM Think. Accessed: Dec. 11, 2024. [Online]. Available: https://www.ibm.com/think/topics/ai-agents

[5] Z. Xi et al., “The Rise and Potential of Large Language Model Based Agents: A Survey,” Sep. 19, 2023, arXiv: arXiv:2309.07864. doi: 10.48550/arXiv.2309.07864.

[6] J. Cook, “OpenAI’s 5 Levels Of ‘Super AI’ (AGI To Outperform Human Capability),” Forbes. Accessed: Dec. 11, 2024. [Online]. Available: https://www.forbes.com/sites/jodiecook/2024/07/16/openais-5-levels-of-super-ai-agi-to-outperform-human-capability/

[7] T. Guo et al., “Large Language Model based Multi-Agents: A Survey of Progress and Challenges,” Apr. 19, 2024, arXiv: arXiv:2402.01680. Accessed: Nov. 18, 2024. [Online]. Available: http://arxiv.org/abs/2402.01680

[8] X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, “A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges,” ResearchGate. Accessed: Nov. 18, 2024. [Online]. Available: https://www.researchgate.net/publication/384732283_A_survey_on_LLM-based_multi-agent_systems_workflow_infrastructure_and_challenges

[9] J. Liu et al., “Large Language Model-Based Agents for Software Engineering: A Survey,” Sep. 04, 2024, arXiv: arXiv:2409.02977. Accessed: Nov. 18, 2024. [Online]. Available: http://arxiv.org/abs/2409.02977

[10] Y. Wang et al., “Agents in Software Engineering: Survey, Landscape, and Vision,” Sep. 23, 2024, arXiv: arXiv:2409.09030. Accessed: Nov. 08, 2024. [Online]. Available: http://arxiv.org/abs/2409.09030

[11] Y. Ye et al., “ProAgent: From Robotic Process Automation to Agentic Process Automation,” Nov. 23, 2023, arXiv: arXiv:2311.10751. doi: 10.48550/arXiv.2311.10751.

[12] L. Dong, Q. Lu, and L. Zhu, “AgentOps: Enabling Observability of LLM Agents,” Nov. 30, 2024, arXiv: arXiv:2411.05285. doi: 10.48550/arXiv.2411.05285.

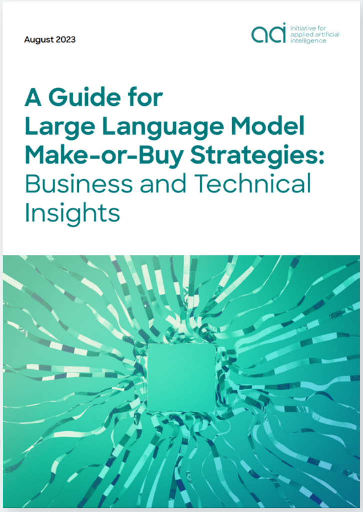
[13] M. Zhuge et al., “Agent-as-a-Judge: Evaluate Agents with Agents,” Oct. 16, 2024, arXiv: arXiv:2410.10934. Accessed: Oct. 21, 2024. [Online]. Available: http://arxiv.org/abs/2410.10934

[14] H. Jin, L. Huang, H. Cai, J. Yan, B. Li, and H. Chen, “From LLMs to LLM-based Agents for Software Engineering: A Survey of Current, Challenges and Future,” arXiv.org. Accessed: Aug. 30, 2024. [Online]. Available: https://arxiv.org/abs/2408.02479v1

“We see Generative AI agents, embedded into advanced agentic RAGs, replicating complex human analyses and decisions. When guided by clearly defined processes, they automate tasks, increase efficiency and achieve precision, enabling solutions to challenges that were previously out of reach. This capability has the potential to strengthen the German economy by boosting growth and counteracting labour shortages.”

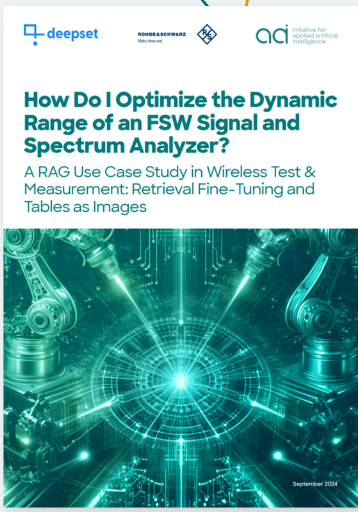
Lukas Wogirz
CEO & Co-Founder
databAIse

Do you want to dive deeper into LLM and RAG?
Start your journey with our white papers.



Firms that employ large language models (LLMs) can create significant value and achieve sustainable competitive advantage. However, the decision of whether to make-or-buy LLMs is a complex one and should be informed by consideration of strategic value, customization, intellectual property, security, costs, talent, legal expertise, data, and trustworthiness. It is also necessary to thoroughly evaluate available open-source and closed-source LLM options, and to understand the advantages and disadvantages of fine-tuning existing models versus pre-training models from scratch.

Our latest whitepaper on Retrieval-Augmented Generation (RAG) offers insights into the advancements and challenges of Retrieval-Augmented Generation (RAG) within the industry. It provides an analysis of industry demands, current methodologies, and the obstacles in developing and evaluating RAG. Additionally, our whitepaper aims to facilitate strategy development and knowledge exchange about practical use cases across various industrial sectors. The whitepaper is the result of extensive studies and discussions conducted with our internal teams and industry partners. It highlights RAG as a cost-effective technique that has significantly improved the trustworthiness and control of Large Language Model (LLM) applications over the past year.



Our RAG use case study on Retrieval-Augmented Generation (RAG) within the test and measurement industry highlights common challenges in the technical domain and explores effective RAG evaluation techniques. We demonstrate how Large Language Models (LLMs) can be leveraged to scale up RAG evaluation reliably, and address industry-specific challenges such as multilingual data, in-domain data, and complex tabular structures. Our vision pipeline and retrieval fine-tuning solutions have significantly improved the accuracy of RAG, proving the value of customized RAG applications for the wireless test and measurement sector.

Authors

Dr. Paul Yu-Chun Chang
Senior AI Expert: Foundation Models –
Large Language Models,
appliedAI Initiative GmbH
p.chang@appliedai.de

Paul Yu-Chun Chang works as an Senior AI Expert specializing in Large Language Models at appliedAI Initiative GmbH. He has over 10 years of interdisciplinary research experience in computational linguistics, cognitive neuroscience, and AI, and more than 6 years of industrial experience in developing AI algorithms in language modeling and image analytics. Paul holds a PhD from LMU Munich, where he integrated NLP and machine learning methods to study brain language cognition.

Mingyang Ma
Principal AI Strategist & Product Manager,
appliedAI Initiative GmbH
m.ma@appliedai.de

Mingyang Ma works as Principal AI Strategist & Product Manager at appliedAI Initiative GmbH, supporting all partner companies' decision making and technical solution identification of various AI use cases, with a particular focus on leveraging LLMs. With over 6 years of expertise in NLP, Mingyang has excelled in the realm of Conversational AI, demonstrating her proficiency in application DevOps and platform development across various processes during her tenure at BMW Group in both Germany and the USA.

Bernhard Pflugfelder
Head of Generative AI,
appliedAI Initiative GmbH
b.pflugfelder@appliedai.de

Bernhard Pflugfelder works as Head of Generative AI at appliedAI Initiative GmbH. Bernhard has 15 years of experience in the fields of Data Science, Natural Language Processing (NLP), as well as data and AI across different companies such as BMW Group or Volkswagen Group. He is renowned for his expertise especially in the field of AI in general, NLP and generative AI in particular.

Contributors

Antoine Leboyer
Managing Director SW/AI
TUM Venture Labs
antoine.leboyer@unternehmertum.de

Antoine Leboyer is an entrepreneur and the Managing Director of SW/AI at TUM Venture Lab and a Board Member at Hyperganic. Formerly, he served as President and CEO of GSX and held board positions at Geneva Liberal Synagogue and Martello Technologies. He holds an MBA from Harvard, class of '92.

Lukas Wogirz
CEO & Co-Founder
databAIse
ljw@databai.se

Lukas Wogirz is the CEO and Co-Founder of databAIse, an AI-powered platform transforming unstructured text data into actionable insights. With a Master's degree in Electrical Engineering and Information Technology from TUM, Lukas specializes in AI/ML, automation and deep learning. He has previously worked on advanced technologies at MOV.AI, where he developed patented algorithms for industrial automation.

Milos Rusic
CEO and Co-founder
deepset
milos.rusic@deepset.ai

Milos Rusic is the co-founder and CEO of deepset, the company behind Haystack and deepset Cloud—leading solutions for rapid custom LLM and NLP application development. Trusted by NVIDIA, Intel, Airbus, and The Economist, deepset's tools empower enterprises to build and deploy AI solutions tailored to their unique needs and mission-critical use cases. Learn more at deepset.ai.

Dr. Christian Karaschewitz
AI Product Incubation Lead, SAP Business AI
– Product & Partner Management
SAP SE
christian.karaschewitz@sap.com

Christian Karaschewitz is a product manager and innovation leader with over a decade of experience at SAP. Currently, he serves as AI Product Incubation Lead, driving cutting-edge solutions in Business AI. Previously, he led product initiatives for SAP Start-Up initiatives such as Head of Product for Ruum by SAP and Co-Founder and Head of Product of FlexPay by SAP. Beyond his work at SAP, Christian has mentored startups as a Venture Mentor at SAP.iO. He holds a Ph.D. from the University of St. Gallen and a Master's from the University of the Arts Berlin. With a passion for innovation, he excels at delivering impactful technologies and business solutions.

Emre Demirci
Junior AI Engineering LLM
appliedAI Initiative GmbH
E.Demirci@appliedai.de

Emre Demirci is a dedicated Data Engineering Master's student at the Technical University of Munich (TUM). As a working student at AppliedAI, Emre focuses on developing cutting-edge solutions involving large language models (LLMs) and knowledge graphs. Passionate about leveraging technology for impactful solutions, Emre's work bridges the gap between AI innovation and real-world applications.

Joong-Won Seo
Junior LLM & Software Engineer
appliedAI Initiative GmbH
j.seo@appliedai.de

Joong-Won Seo is a Master's student in Computer Science at the Technical University of Munich, specializing in deep learning, generative AI, and full-stack engineering. With extensive experience as a teaching assistant at TUM, he combines a strong research foundation with hands-on engineering skills. He applies LLMs and generative AI to develop prototypes that translate theoretical ideas into practical solutions for real-world challenges.

Ferdy Dermawan Hadiwijaya
Junior Generative AI Engineer
appliedAI Initiative GmbH
f.hadiwijaya@appliedai.de

Ferdy Dermawan Hadiwijaya is a master's student in Computer Science at the Technical University of Munich (TUM), specializing in Natural Language Processing and Generative Models. As a working student at appliedAI, he serves as a Junior GenAI Engineer, bringing three years of professional experience in Large Language Models and software development to bridge cutting-edge academic research with practical industry applications.

About appliedAI Initiative GmbH

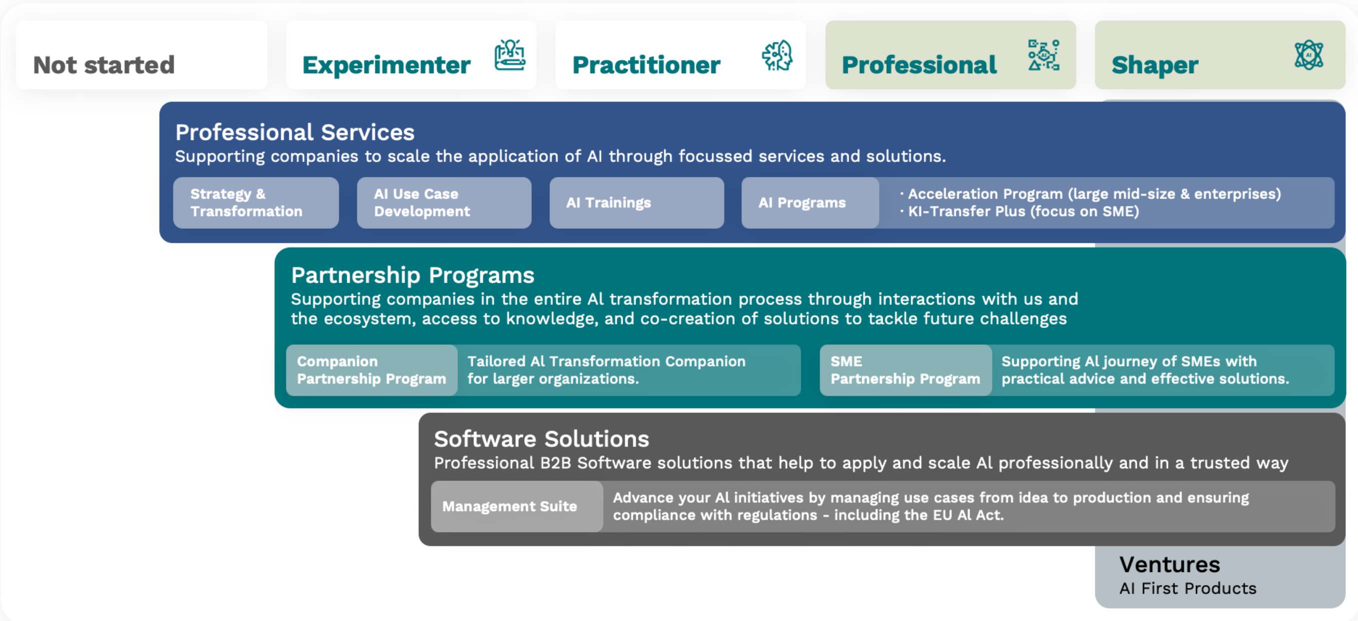
appliedAI is Europe's largest initiative for the application of trusted AI technology. The initiative was established in 2017 by Dr. Andreas Liebl as a division of UnternehmerTUM Munich and transferred to a joint venture with Innovation Park Artificial Intelligence (IPAI) Heilbronn in 2022.

At the Munich and Heilbronn offices, more than 100 employees pursue the goal of making European businesses a shaper in the AI era in order to maintain Europe's competitiveness and actively shape the future.

appliedAI holistically supports international corporations, including BMW and Siemens, as well as medium-sized companies in their AI transformation. This is accomplished through partnership-based exchange and joint knowledge building, comprehensive accelerator programs, and specific solutions and services, such as strategy consulting and Use-Case development.

For more information, please visit
https://www.appliedai.de/en/

We offer a unique set of offerings to help companies on their way to becoming AI shapers



Acknowledgement

We express our sincere appreciation for the excellent work carried out by the authors, reviewers, and designer of this white paper. Their great expertise, motivation, and dedication to every detail made this white paper an exceptional contribution to the AI community.

Furthermore, we extend our gratitude to all contributors for their valuable contributions throughout this joint case study. Their profound expertise and commitment have played a pivotal role in shaping the ideas, knowledge and results presented in this paper.

Our sincere thanks go to our appliedAI industry partners Atruvia, Linde, Giesecke & Devrient, Rohde & Schwarz, SAP and Siemens as well as our technology partner deepset for their openness, commitment, and contributions to this case study. We highly appreciate their willingness to share insights and results.

The collective expertise, exchange, and dedication to advancing the knowledge in generative AI and gentic applications were great inspirations throughout the process of creating this white paper.

initiative for applied artificial intelligence

Generative AI Agents in Action: Revolutionizing Software Development Testing

appliedAI Initiative GmbH

August-Everding-Straße 25
81671 München
Germany
www.appliedai.de