Clinical and Translational Science

REVIEW OPEN ACCESS

Agents for Change: Artificial Intelligent Workflows for Quantitative Clinical Pharmacology and Translational Sciences

Mohamed H. Shahin[1] | Srijib Goswami[2] | Sebastian Lobentanzer[3,4] | Brian W. Corrigan[1]

[1]Pfizer Global Research and Development, Groton, Connecticut, USA | [2]Insight Rx, Inc., San Francisco, California, USA | [3]Institute for Computational Biomedicine, Heidelberg University Hospital, Heidelberg, Germany | [4]EBI/Open Targets, EMBL-EBI, Cambridge, UK

Correspondence: Mohamed H. Shahin (mohamed.shahin@pfizer.com)

Received: 19 December 2024 | **Revised:** 12 February 2025 | **Accepted:** 22 February 2025

Funding: The authors received no specific funding for this work.

Keywords: agentic workflows | AI agents | artificial intelligence | explainable machine learning | large language models | machine learning

ABSTRACT

Artificial intelligence (AI) is making a significant impact across various industries, including healthcare, where it is driving innovation and increasing efficiency. In the fields of Quantitative Clinical Pharmacology (QCP) and Translational Sciences (TS), AI offers the potential to transform traditional practices through the use of agentic workflows—systems with different levels of autonomy where specialized AI agents work together to perform complex tasks, while keeping "human in the loop." These workflows can simplify processes, such as data collection, analysis, modeling, and simulation, leading to greater efficiency and consistency. This review explores how these AI-powered agentic workflows can help in addressing some of the current challenges in QCP and TS by streamlining pharmacokinetic and pharmacodynamic analyses, optimizing clinical trial designs, and advancing precision medicine. By integrating domain-specific tools while maintaining data privacy and regulatory standards, well-designed agentic workflows empower scientists to automate routine tasks and make more informed decisions. Herein, we showcase practical examples of AI agents in existing platforms that support QCP and biomedical research and offer recommendations for overcoming potential challenges involved in implementing these innovative workflows. Looking ahead, fostering collaborative efforts, embracing open-source initiatives, and establishing robust regulatory frameworks will be key to unlocking the full potential of agentic workflows in advancing QCP and TS. These efforts hold the promise of speeding up research outcomes and improving the efficiency of drug development and patient care.

JEL Classification: Artificial Intelligence and Machine Learning

1 | Introduction

For more than three decades, artificial intelligence (AI) and machine learning (ML) have been utilized in Quantitative Clinical Pharmacology (QCP) and Translational Sciences (TS) [1–3]. From early computational models assisting in drug metabolism studies to sophisticated algorithms predicting clinical outcomes, AI/ML methodologies have evolved significantly, becoming valuable assets for data analysis, predictive modeling, and informed decision-making [4–6]. Recently, large language models (LLMs) have emerged as powerful AI systems capable of processing and generating human-like text, offering even greater capabilities for data synthesis, analysis, and interpretation [7]. However, despite their potential, LLMs present certain challenges, such as reproducibility issues, concerns over data and information provenance, and critical data privacy considerations for both patient and proprietary information [8]. Additionally, because LLMs are generally

designed for a wide range of tasks, they often lack the domain-specific expertise required in specialized fields like QCP and TS, leading to the necessity for fine-tuning or real-time retrieval-augmented generation (RAG), though these do not always produce optimal results [9].

To address these challenges, the concept of AI agents and agentic workflows has been introduced and explored in various fields [10–16]. AI agents are programs designed to perform specific tasks with varying degrees of autonomy, and when integrated into agentic workflows, they collaborate to achieve complex objectives more efficiently. Prominent figures in AI have underscored the potential of these workflows. For instance, Yann LeCun, Chief AI Scientist at Meta, envisions that "In the future, all human interaction with the digital world will be through AI agents" [17]. Similarly, Andrew Ng, renowned AI expert and founder of DeepLearning.AI, anticipates a dramatic expansion of AI's capabilities due to agentic workflows, stating in 2024 the following, "*I expect that the set of tasks AI could do will expand dramatically this year because of agentic workflows*" [18].

In this review, we look at how agentic workflows can address the challenges of traditional, time-consuming processes in QCP and TS. By combining multiple LLMs with specialized tools and methods, these workflows create a more efficient and reliable system. We will explain what AI agents are and how they function within these agentic workflows, while also sharing examples of how they are being implemented in platforms that support QCP and biomedical research. Our goal is to show how these innovative approaches can empower scientists to work more effectively, automate repetitive tasks, and speed up the pace of research.

2 | Historical Context: Evolution of AI in QCP and TS

AI and ML have been part of QCP and TS since the early 1990s, where they were primarily used for tasks such as pharmacokinetic modeling, dose optimization, and drug interaction predictions [1–3, 19, 20]. Early AI models were often rule-based systems, which relied heavily on predefined algorithms to process structured data and generate outcomes. While these systems provided valuable insights, they were limited by their inability to adapt to new information or context and therefore were not applicable in dynamic clinical environments.

Building upon these earlier models, the 2000s witnessed a major shift with the introduction of deep neural networks and more advanced ML techniques [21]. These models could learn from data, make more accurate predictions, and handle complex datasets, including unstructured data like clinical notes and imaging results. This allowed for better modeling of biological systems and patient responses. However, these advanced models introduced new challenges, particularly around model interpretability, data privacy, and regulatory acceptance [22, 23]. The complexity of neural networks made it difficult for clinicians and regulators to understand the decision-making processes and raised concerns about transparency and trustworthiness.

In recent years, LLMs, which are very large deep learning models, have gained prominence due to their ability to process large amounts of natural language data, making them invaluable for tasks such as literature review, clinical report generation, and patient data synthesis [24]. LLMs like GPT-3.5 [25], (i.e., ChatGPT) have demonstrated remarkable capabilities in understanding and generating human-like text, facilitating the analysis of medical records and scientific publications [26]. Yet, despite their potential, LLMs, applied off the shelf, currently struggle with domain-specific applications in QCP and TS. They lack the precision and reliability required for critical tasks, such as interpreting complex pharmacokinetic data or accurately predicting drug interactions without extensive domain-specific training [5, 27]. Moreover, concerns around data privacy arise when using LLMs, as they may inadvertently expose sensitive patient information during processing.

To address these challenges and meet the specialized knowledge requirements of fields like QCP and TS, agentic workflows have been developed—a new paradigm that shifts from relying on single, general-purpose AI models to orchestrating multiple specialized AI agents. These AI agents are designed to perform specific tasks and can be integrated into a cohesive workflow that mirrors the complex processes of drug development and translational research. By enabling targeted problem-solving and modular adaptability, agentic workflows harness the strengths of AI while mitigating issues related to data privacy and regulatory compliance. This approach embeds domain knowledge directly into AI agents, enhancing precision and reliability. For instance, specialized agents can be fine-tuned with proprietary datasets under strict privacy controls, ensuring compliance with regulatory standards.

3 | Defining AI Agents and Their Anatomy in Agentic Workflows

In this section, we define common terms and characteristics related to AI agents, as summarized in Figure 1.

3.1 | AI Agents

An AI agent is a system that performs tasks by interacting with its environment and takes actions to achieve specific goals. Unlike traditional AI systems that react to predefined instructions, AI agents exhibit a high level of autonomy and adaptability, enabling them to proactively engage in complex and dynamic scenarios. While there is not a universally accepted definition for agents powered by LLMs, they are generally designed to mimic human decision-making and problem-solving capabilities. AI agents can operate with varying degrees of autonomy:

- Autonomous Agents: Fully independent systems that perform tasks without human intervention.
- Semi-autonomous Agents: Systems that operate independently for some tasks but require human input or approval for certain decisions.

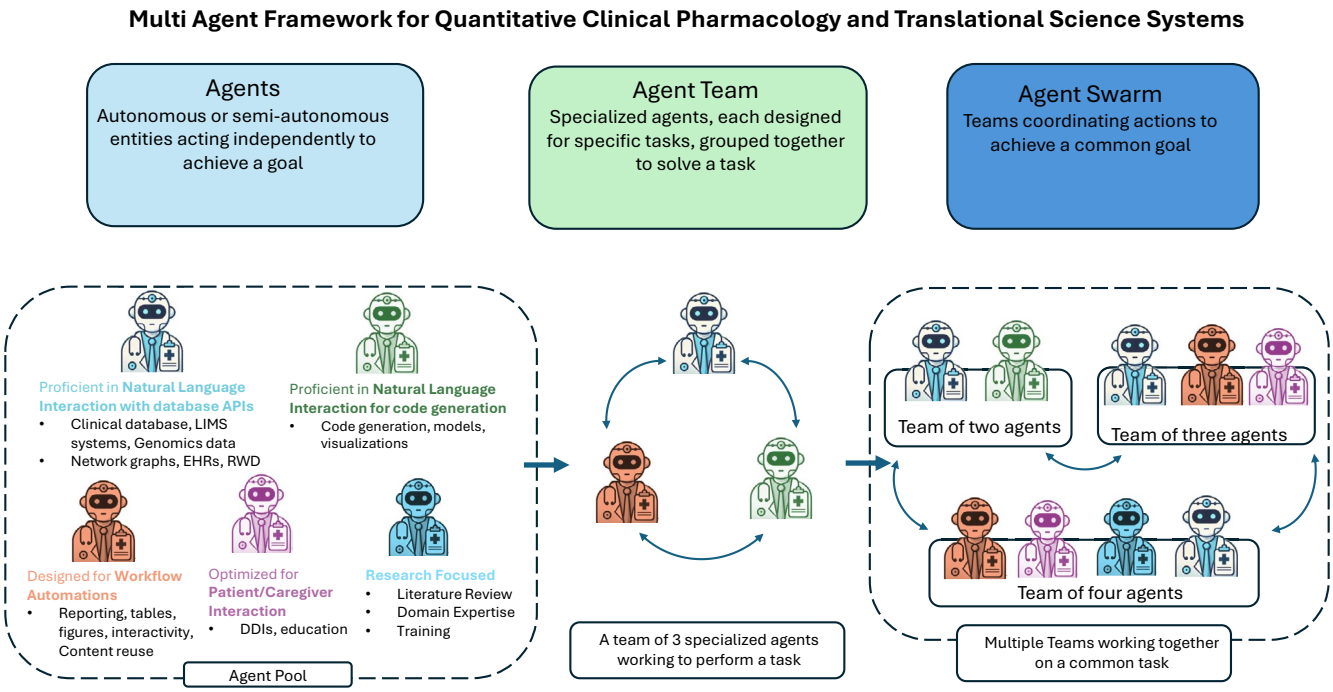

FIGURE 1 | A schematic illustrating how autonomous and semi-autonomous agents collaborate in teams and swarms to perform diverse tasks in quantitative clinical pharmacology and translational science. Individual agents, each with specialized capabilities (e.g., natural language processing, code generation, workflow automation), can be grouped to tackle complex objectives, ultimately forming larger "swarms" that coordinate efforts at scale. DDIs, drug–drug interactions; EHRs, electronic health records; LIMS, laboratory information management system; RWD, real-world data.

- Collaborative Agents: Agents that work alongside humans or other agents, making decisions with human insight or approvals.

AI agents can process structured and unstructured data, adapting their behavior based on changing information. Ideally, they learn from new data and user feedback, updating their models to improve performance over time. In the context of QCP and TS, AI agents can perform various tasks, such as reading from and writing to databases using natural language prompts, generating code or invoking tools based on instructions, summarizing documents and literature, and interacting with team members to provide domain-specific information (e.g., drug–drug interactions, benefit–risk assessments).

3.2 | Agent Teams

An Agent Team is a group of specialized AI agents designed to interact and collaborate to complete complex tasks. Each agent within the team utilizes task-specific or domain-specific LLMs, tailored for particular functions. By coordinating their efforts, agent teams can handle multifaceted challenges more efficiently than individual agents acting alone can. For instance, in PK analysis, one agent might focus on data extraction, another on model development, and a third on result interpretation.

3.3 | Agent Swarms

An Agent Swarm consists of multiple agent teams working together to solve highly complex tasks. This hierarchical structure allows for scalability and the division of labor across numerous specialized agents and teams, mirroring the collaborative nature of human organizations in large projects. In drug development, an agent swarm might coordinate activities across different phases, from target identification and validation to clinical trial management and post-marketing surveillance.

To fully leverage the potential of agentic workflows, it is essential to understand the fundamental structure—or "anatomy"—of the AI agents operating within these systems (Figure 2). An AI agent in this context comprises several interconnected components that enable it to complete tasks with different levels of autonomy, make informed decisions, and continually learn from its environment [28], as described below.

3.4 | The LLM/Foundation Model: The Brain

At the core of every AI agent lies a LLM, or a foundation model, functioning as the "brain" of the system. These models provide foundational intelligence and language comprehension capabilities, allowing the agent to process and generate human-like text, understand complex instructions, and make decisions based on extensive training data. While LLMs have traditionally focused on natural language processing, foundation models extend these capabilities by integrating multimodal data, including images, audio, and video, which enrich the agent's understanding of its environment. In QCP and TS, these capabilities can empower agents to interpret scientific literature, clinical trial protocols, and patient data, and to leverage diverse data sources for advanced analyses and generating valuable insights.

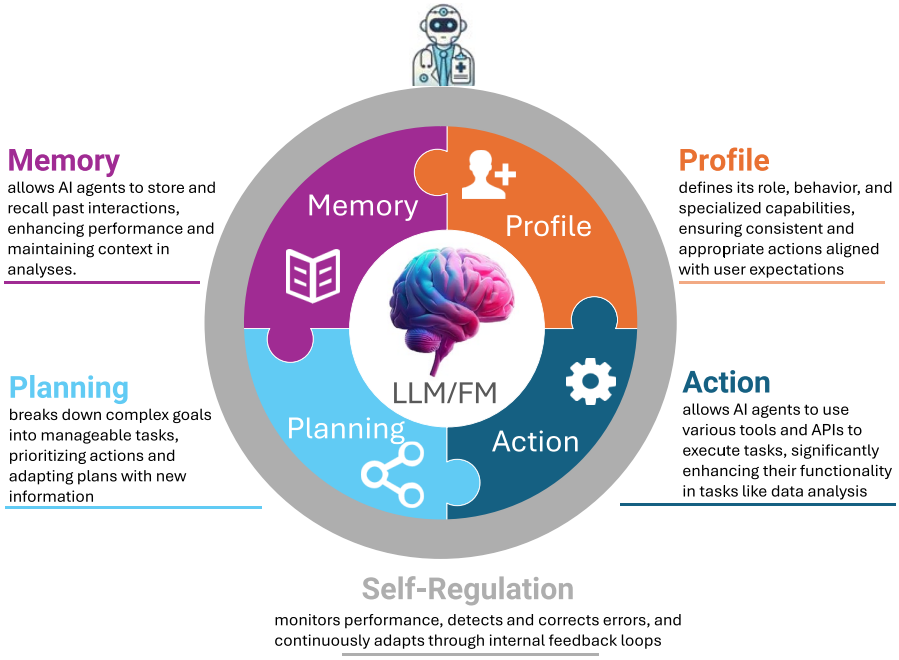

FIGURE 2 | A conceptual illustration of an AI agent's five key components—memory, profile, planning, action, and self-regulation—working in tandem with a large language model (LLM) or foundation model (FM). Each component supports the agent's ability to store context, define its role, break down tasks, execute solutions, and adapt its behavior. API, application programming interface.

3.5 | Memory Module: Storing and Recalling Information

The memory feature allows the AI agent to remember and use information from past interactions, which helps it keep context during ongoing tasks. This means the agent can learn from previous experiences, getting better over time. It also allows for more personalized assistance by recalling what the user has done before, making sure there is consistency in projects that involve multiple steps or take place over a longer period. Memory can be short-term, retaining information during a single task, or long-term, preserving knowledge across multiple interactions and projects.

3.6 | Profile: Defining Agent Behavior and Specialization

An agent's profile outlines its role, behavior, and specific capabilities. This includes its area of specialization (such as PK modeling expert or clinical trial designer), communication style suitable for interactions with scientists and clinicians, adherence to ethical guidelines, and compliance with regulatory standards. The profile ensures that the agent's responses and actions are consistent, appropriate, and aligned with user expectations in both QCP and TS domains.

3.7 | Planning Module: Transforming Objectives Into Tasks

The planning module enables the agent to break down complex goals into smaller, manageable tasks. By looking at the overall objectives, the agent figures out the steps needed to reach them, prioritizes tasks, and adjusts the plan as new information comes in. For instance, when planning a clinical trial, the agent might organize tasks like defining patient inclusion criteria, selecting appropriate endpoints, and determining optimal dosing regimens.

3.8 | Action Module: Utilizing Tools and Resources

The action module allows the AI agent to engage with its environment and carry out tasks, often by using different tools or connecting to application programming interfaces (APIs). This feature broadens the range of tasks the agent can handle. In QCP and TS, for example, the agent might pull data from clinical trial or genomic databases, run simulations using PK/PD modeling software, perform statistical analyses with specialized computational tools, or retrieve real-time data from electronic health records.

3.9 | Self-Regulation Module: Error Detection, Correction, and Continuous Learning

In addition to the core modules described above, many AI agents incorporate a self-regulation module that monitors their performance. This module is responsible for detecting errors and anomalies, initiating corrective measures, and learning from mistakes through multi-step internal checks. Unlike conventional systems that simply throw an error when encountering a problem, the self-regulation module enables AI agents to autonomously resolve issues or propose alternative solutions to continuously improve their performance over time.

4 | Implementing Agentic Workflows in QCP and TS

Agentic workflows bring together AI agents with existing tools, processes, and expert knowledge in a coordinated system. Unlike traditional workflows that sometimes depend on a single LLM, agentic workflows utilize multiple specialized agents working together within a coordinated framework to improve performance and adaptability (see Table 1). A key feature of agentic workflows is the integration of "tool use," where agents are programmed to trigger additional tools or data calls as needed. This capability allows for greater flexibility and accuracy in completing tasks, as agents can leverage the strengths of existing domain-specific tools and data sources. While the concept remains relatively new, these workflows are already being used in hospital systems and simulation environments [11].

An ideal agentic workflow seamlessly integrates across an organization's existing systems, enhancing overall efficiency while leveraging the existing scientific knowledge and expertise within the field to ensure accuracy and relevance. These workflows capitalize on and optimize automated processes that have already been developed, thereby improving operational efficiency. Transparency and reproducibility are crucial, as they play a vital role in establishing trust with users and stakeholders. Furthermore, data privacy is safeguarded by ensuring that data storage and handling are conducted within secure and appropriate systems. A crucial aspect of agentic workflows is the inclusion of a memory component, which allows the workflow to learn and improve over time through continuous use. By adhering to these principles, agentic workflows can significantly enhance the capabilities of QCP and TS, fostering a more integrated, efficient, and trustworthy analytical environment.

An Illustration of the agentic workflow that we envision for QCP and TS is shown in Figure 3. This workflow integrates several key components, including domain experts, task-specific agents, and teams, diverse data sources, and advanced analytic tools.

4.1 | Components of Agentic Workflows for QCP or TS

4.1.1 | Domain Experts

At the heart of QCP/TS agentic workflows are domain experts—such as clinical pharmacologists, translational scientists, and pharmacometricians—who initiate tasks and provide critical oversight. They review the data and code utilized by the agents, ensuring that all analyses and reports meet the highest standards of accuracy and reliability. Their approval is essential for maintaining the integrity of the workflow, as they bring specialized knowledge that guides the agents toward meaningful and relevant outcomes.

4.1.2 | Task-Specific Agents and Teams

To improve efficiency and accuracy, agentic workflows use task-specific agents, each designed for particular tasks. These agents are trained and fine-tuned to handle specific jobs, making them highly skilled at what they do. By selecting agents tailored to the specific requirements of each task, the workflow benefits from a customized approach that enhances overall performance and effectiveness.

4.1.3 | Data Sources

The data sources employed in these workflows are varied and application-dependent. They encompass clinical data from ongoing or previous medicine programs, scientific literature, and other external knowledge bases. Additionally, translational data sources such as pharmacokinetics, biomarkers, "omics," digital, and imaging data are incorporated alongside real-world data (RWD) from electronic health records, medical and prescription claims, and safety reporting systems.

TABLE 1 | Comparison of agentic workflows and traditional workflows.

Characteristic	AI agentic workflows	Traditional workflows
Autonomy	High; AI agents handle tasks autonomously, semi-autonomously, or collaboratively	Low; manual human-driven processes
Learning	Adaptive; agents learn and improve from feedback	Static; updated manually based on periodic reviews
Speed	Faster; agents work continuously	Slower; constrained by human availability
Complexity	Handles complex tasks via advanced algorithms	Struggles with high complexity without more effort
Scalability	Highly scalable with automation	Limited by human resources and effort
System integration & human collaboration	Employs multiple specialized AI agents, prompt engineering techniques, and generative AI networks, enabling seamless automation with high-level human oversight	Relies on manual workflow management systems (e.g., checklists, flowcharts) with high human involvement and coordination
Innovation	Identifies patterns and opportunities autonomously	Relies on human creativity and expertise
Cost efficiency	High setup cost but lower operational costs	Ongoing labor and management costs
Reproducibility & explainability	Automated processes enhance reproducibility but may require additional measures to improve the explainability of AI decision logic	More transparent and explainable due to human-driven decision making; however, reproducibility may suffer from human variability

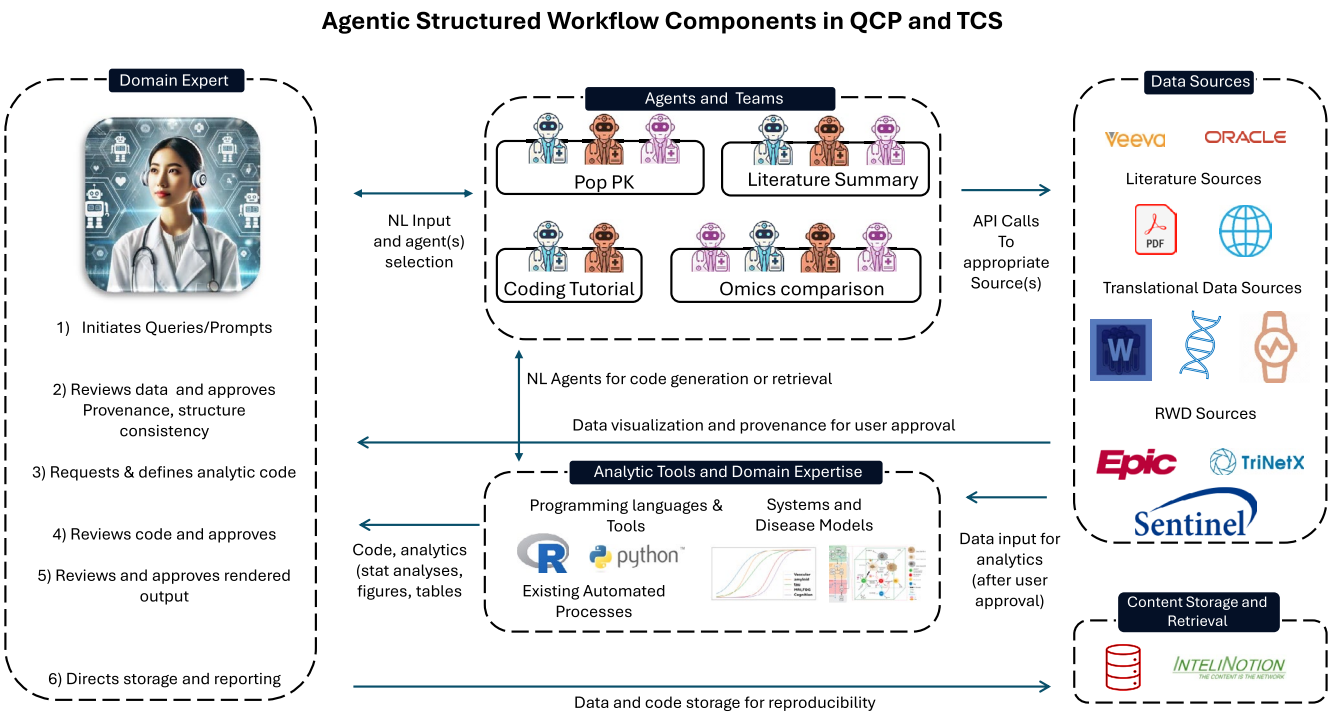

FIGURE 3 | An overview of an agentic workflow in quantitative clinical pharmacology (QCP) and translational clinical science (TCS). It illustrates a multi-step workflow in which a domain expert initiates queries for data or analytics. After that, AI agents are selected based on the task at hand—such as population PK, literature summarization, coding, or omics analysis—and execute API calls to the appropriate data sources. The resulting data are returned for further analysis, code generation, and visualization. At each stage, the domain expert reviews and approves output before final storage and reporting, ensuring reproducible results and maintaining clear provenance. API, application programming interface; Pop PK, population pharmacokinetics; RWD, real-world data.

4.1.4 | Advanced Analytic Tools

Advanced analytic tools and existing domain expertise are seamlessly integrated into the workflow. This includes QSP and drug development tools (DDTs), as well as existing automated analysis and reporting systems. Automated analysis and reporting systems efficiently process data and generate comprehensive reports with minimal human intervention, thereby saving time and reducing the potential for error. Archiving systems provide robust solutions for storing data, analyses, and reports, ensuring reproducibility and facilitating continuous improvement through retrospective review and learning.

By harmoniously combining the expertise of domain specialists, the precision of task-specific agents, the richness of diverse data sources, and the power of advanced analytic tools, these agentic workflows can enhance the efficiency and effectiveness of QCP and TS.

5 | Opportunities and Challenges for Agentic Workflows in QCP and TS

As we noted earlier, the integration of agentic workflows into QCP and TS presents transformative opportunities, alongside challenges that must be thoughtfully addressed. In this section, we highlight some of those potential opportunities and critical challenges.

5.1 | Opportunities

One of the most significant opportunities offered by agentic workflows is the streamlining of complex processes. By automating routine and time-consuming tasks—such as data entry, basic statistical analyses, and initial report drafting—agentic workflows can reduce manual errors and accelerate decision-making. For example, in PK/PD modeling, agents can automatically process patient data to generate preliminary models, allowing researchers to focus on refining and interpreting results rather than data preparation. This automation leads to a more efficient use of resources, enabling teams to handle larger workloads without proportional increases in staffing or time.

Agentic workflows also contribute to enhanced decision-making by augmenting human expertise with real-time data analysis and evidence-based recommendations. These AI agents can quickly go through large amounts of data, find patterns, and point out anything unusual that might be missed during manual reviews. In clinical trial design, for instance, agents can run different simulations to help determine the best dosing strategies or group patients more effectively, which gives researchers helpful insights that lead to informed and accurate decisions.

Furthermore, the ability of agentic workflows to speed up research can transform how new therapies and treatments are developed. For example, AI agents can automate literature reviews by scanning and summarizing the latest publications related to a

study or a disease, saving researchers a lot of time. They can also help generate new ideas by pulling together data from different sources, like genomic databases and clinical trial results, to uncover potential drug targets or biomarkers. This faster pace not only shortens research timelines but also increases the chances of making groundbreaking discoveries.

5.2 | Challenges

Although agentic workflows bring exciting possibilities, there are several challenges to address to effectively implement agentic workflows in QCP and TS. One significant obstacle is how to integrate them with the systems that organizations have been using for years. Many companies rely on well-established tools and processes, and introducing agentic workflows could mean making major changes, like updating software, retraining staff, and adjusting standard procedures. This process can be complicated, take up resources, and temporarily disrupt ongoing projects. To overcome this, careful planning is needed, along with involvement from stakeholders and potentially a step-by-step approach to avoid too much disruption all at once.

Data quality and provenance are critical concerns when deploying agentic workflows. These systems need good data to work correctly, and if the data is not accurate, it can lead to mistakes, which can be a big problem in areas like patient care or drug development. Ensuring data provenance—the detailed history of data origins and transformations—is essential for traceability and accountability. Implementing robust data governance frameworks, including validation checks and metadata documentation, is necessary to maintain the integrity of the workflows.

Maintaining reproducibility poses another significant challenge. Agentic workflows, by design, are dynamic; agents continuously learn and update their models based on new data and interactions. While this adaptability is advantageous, it can make reproducing specific results difficult if the system's state changes between analyses. Reproducibility is a cornerstone of scientific research, and its absence can undermine confidence in the results generated by agentic workflows. To address this, mechanisms such as version control for models, detailed logging of agent activities, and the ability to snapshot the system state at specific points are vital.

Building trust among users and stakeholders is perhaps one of the most critical challenges. For agentic workflows to be embraced, clinicians, researchers, and regulatory bodies must have confidence in the system's outputs. To make that happen, it is important to be open about how the agents work, how they make decisions, and how they handle data. This includes providing easy-to-understand explanations, showing how the algorithms work, and making sure the agents' recommendations are clear and can be explained. Getting users involved in the development process and giving them the training they need can also help build trust and make it easier for people to accept and use these new systems.

Data privacy is another major concern, especially when dealing with sensitive patient information or proprietary data. Traditional language models can sometimes struggle with ensuring high levels of data security, particularly when working with external data or operating in cloud environments that may be vulnerable. To address this, agentic workflows must ensure that all data storage and management happen in secure systems that comply with regulations like HIPAA and GDPR. Additionally, for AI tools to be widely accepted in fields like QCP and TS, they must meet strict regulatory requirements that prioritize patient safety as well as data integrity. The dynamic and opaque nature of LLMs makes it difficult to meet regulatory requirements, as regulators need clear explanations of how decisions are reached. To address this, agentic workflows need to thoroughly document all data processing and ensure they follow the necessary regulatory standards.

5.3 | Strategies to Overcome Challenges

To effectively address these challenges, agentic workflows should enhance the abilities of domain experts by working collaboratively with them, augmenting human expertise rather than replacing it. Additionally, built-in quality assurance steps are essential, including human reviews and approvals of data provenance and structure, as well as the analytic code before use, to ensure the integrity and reliability of the workflow. Moreover, making sure that results from these systems can be consistently reproduced is really important. To do this, the data and code generated by agentic workflows need to be stored in secure systems that meet regulatory standards. This helps protect data privacy while also ensuring that everything complies with global regulatory requirements.

It is unlikely that these developments can be achieved by the often-siloed efforts of individual actors in academia and industry. Particularly with respect to the rapidly evolving ecosystem of LLMs, LLM frameworks, auxiliary technologies, and open-source libraries for their integration and programmatic access, the clinical pharmacology and translational research communities need to federate efforts for the resource-effective application of these new technologies. We suggest that it is most efficient to aggregate generic functions for the use of LLMs as described above in open-source frameworks [29], in order to federate the maintenance required to remain up-to-date in the face of the breakneck pace of current developments. In addition to removing redundancies in the technical requirements of deploying agent-based systems, this will also introduce much-needed diversity into the ways we approach these intriguing but brittle technologies. Wherever possible, these contributions should be made in a collaborative, pre-competitive setting between academia and industry.

Furthermore, we propose that it is crucial to go beyond the current LLM-driven developments for a sustainable implementation of the goals outlined above. LLMs have systematic biases and technical limitations that cannot currently be addressed due to their black-box nature, such as confabulation and lack of long-term attention. Guaranteeing semantic stability of the maintained data while upholding privacy-related and regulatory constraints requires robust, transparent knowledge management and continuous monitoring. Injecting domain expertise into agentic workflows, which often rely on generically trained LLMs without domain-specific behavioral training, requires efficient communication between knowledge management

systems and the acquisition of knowledge by the LLMs. The open frameworks we propose [29] should have intimate native integration with dedicated, semantically enriched knowledge stores [30]. Only in this manner can robust knowledge retrieval of information relevant to the agentic workflows be upheld and monitored effectively.

For monitoring the performance of agentic workflows, it is essential that significant domain expertise and engineering effort go into establishing and maintaining highly specific benchmarking suites that cover a wide range of tasks in the spectrum of QCP and TS. The need to federate benchmarking and monitoring is apparent for both performance optimization and regulatory compliance. Dedicated benchmarking tasks should be a cornerstone of open-source implementations of agentic workflows in all biomedical disciplines, allowing maximum transparency, oversight, and stakeholder access to these transformative technologies. The most logical place for implementing benchmarks is alongside the open-source libraries that deploy the agentic workflows [29].

6 | Case Studies and Practical Applications

6.1 | Example #1: InsightRX Apollo-AI

InsightRX Apollo-AI is a practical example of how agentic workflows can support quantitative clinical pharmacologists. Currently under development, Apollo-AI aims to enhance the analytical capabilities of QCP and TS experts by offering tools for PK and PD analyses. The system addresses several limitations of traditional LLM-based tools, such as the risk of hallucinations—where models generate incorrect or nonsensical information—and challenges associated with user interface and workflow design.

To address these challenges, the design of the agent-based analysis system was guided by several key principles: clearly defining agent roles and responsibilities, ensuring that each agent's tasks were narrowly focused, and maintaining clear human-agent interaction throughout the analysis process. The application was developed with a customized user interface (UI) and backend infrastructure. A well-designed UI/UX is essential not only for enhancing the platform's overall usability and ensuring reliable code output but also for understanding human intent throughout the analysis process. A pure chat-based UI like ChatGPT is likely to be suboptimal for PKPD analysis. For example, user workflows for QCP/TCS analysis will require a user interface that can accommodate multiple analysis tasks such as data visualization, user collaboration, analysis management, and code editing. For modeling tasks, users should be able to develop and diagnose models in an iterative manner as well as submit multiple jobs simultaneously. While low-level APIs to LLMs are available to develop a robust analysis system, they often present similar workflow challenges and are generally beyond the technical expertise of most users. Additionally, the underlying software infrastructure was customized to ensure robust data security and compliance throughout the analysis process.

The Apollo-AI system employs a variety of specialized AI agents, each with distinct roles (Figure 4), as defined below, that contribute to a cohesive and efficient analytical workflow. Central to this architecture is the Agent–Computer Interface (ACI), which enhances the functionality and efficiency of these agents.

6.1.1 | Conversational Agent

The Conversational Agent acts as the primary interface between the user and the system. It is specific role is to process user input, such as natural language queries, and translates them into tasks for other agents to execute. By leveraging example queries, analysis plans, and code snippets, the conversational agent ensures that the user's requests are accurately interpreted and carried out effectively.

For example, if a clinical pharmacologist wants to model a patient's drug concentration levels, the Conversational Agent will first confirm that the request pertains to population PK modeling with some preliminary analysis requirements before passing

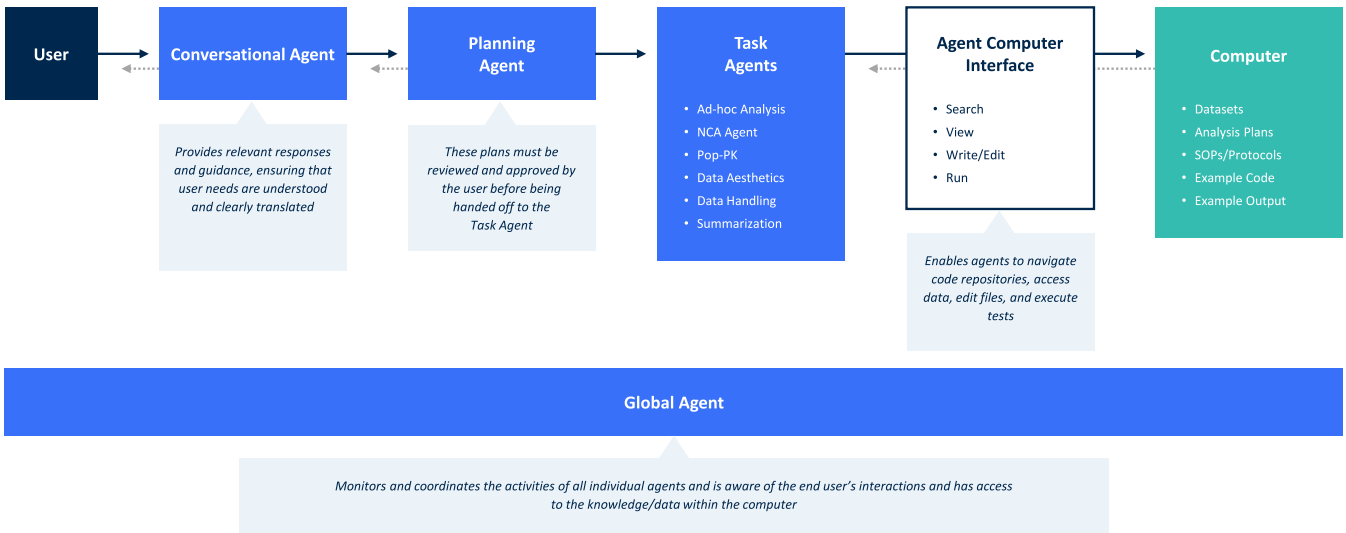

FIGURE 4 | Overview of the Apollo-AI system agentic workflow.

it on to a planning agent. This step is crucial for capturing user intent accurately and serves as a safeguard against downstream errors or hallucinations.

6.1.2 | Planning Agent

The Planning Agent organizes the steps necessary to fulfill a user's analysis request, ensuring that everything is aligned with the user's objectives and any predefined study requirements. Before a plan is executed, the user has a chance to review, modify, and approve the plan developed by the agent, which serves as an important quality check. Similar to the conversational agent, the planning agent keeps the human in the loop by understanding user intent and making the underlying analysis process transparent. For example, an analysis plan for a basic NCA could outline the data variables to be used, provide a step-by-step guide for the analysis process (including which PK parameters to include and the method for calculating terminal half-life), and specify how to handle data below the limit of quantification (BLQs), among other considerations. The user will then have the ability to directly modify the plan before proceeding with the analysis.

6.1.3 | Task Agents

Task Agents are tools designed to perform tasks throughout the analysis process, like finding data outliers, excluding data, running analysis (e.g., exploratory analysis, pop-PK, NCA), making aesthetic modifications to plots/tables, and managing the analysis workflow. Task agents will follow the plan created by the Planning Agent and use the system's resources through the ACI to complete their tasks. For example, a Task Agent might flag and remove unusual data points that could throw off the results, helping to keep the analysis accurate. For an analysis task like NCA, the task agent could invoke specific R libraries or other computational packages to fulfill the analysis request. With access to example code, outputs, and the ACI, Task Agents are able to do their tasks reliably.

6.1.4 | Global Agent

The Global Agent monitors and coordinates the activities of all individual agents, is aware of the end user's interactions, and has access to the knowledge/data within the computer. Its primary objective is to offer timely recommendations and orchestrate agent actions to achieve optimal outcomes. For example, during model development, the Global Agent will track all prior modeling runs, remain aware of the study context and data constraints, and offer suggestions to the end user throughout their workflow. These recommendations may include changes to the model structure, covariance matrix, or error model.

6.1.5 | Agent–Computer Interface

The ACI is a crucial component of the Apollo-AI system designed to enhance overall system performance by providing agents with an environment similar to the tools used by software engineers. This interface enables agents to navigate code repositories, access data, edit files, and execute tests. The ACI enables the retrieval of accurate and relevant knowledge to supplement an agent's response to help prevent downstream hallucinations. Specifically tailored to the operational characteristics of LLMs, the ACI mimics the interactive features of integrated development environments (IDEs) used by developers. Both Task and Planning Agents within Apollo-AI leverage the ACI to search files, write code, view and edit data, as well as run analysis code.

6.1.6 | Computational Infrastructure

Referred to as the "Computer," the underlying computational infrastructure contains all the necessary data, files, PK/PD software, and code and output examples. It interacts with the agents through the ACI, supplying the necessary resources for analysis and code execution. This part of the system acts as a repository for the AI agents, designed to have all of the necessary components required to perform clinical pharmacology analysis.

While still in development, the Apollo-AI system exemplifies how a well-coordinated agentic workflow could be built for QCP. By giving each agent a specific role and making sure they work smoothly with the available technology, the system aims to address many of the limitations associated with traditional workflows.

6.2 | Example 2: BioChatter Reflection Agent

The BioChatter Reflexion Agent showcases the application of agentic workflows in the biomedical field. Given the importance of retrieving accurate and relevant knowledge to supplement LLM responses, we implemented an agent capable of querying semantically structured knowledge graphs (KGs). The BioChatter library [29], an open-source framework for the application of LLMs in biomedical research, connects natively to semantically grounded KGs built by the BioCypher framework [30]. Through this connection, we facilitate the knowledge retrieval from a given KG via the generation of dedicated queries. While LLMs generally perform well in translating natural language queries into structured formats, they often struggle with context, especially in zero-shot scenarios; that is, when they lack the possibility of correcting erroneous queries. To address this issue, we can introduce a reflexion workflow that can consider the quality of a query result and decide whether to pass on the result to the user or engage in a round of corrections. For complex questions, this can lead to a significant increase in robustness, especially when combined with an instruction to consider the quality of the answer with respect to the question by the user. A more detailed and technical description of the implementation can be found in the BioChatter documentation (https://biochatter.org/features/reflexion-agent/).

A practical example of the BioChatter Reflexion Agent in action is its integration into the DECIDER ovarian cancer project (https://www.deciderproject.eu). In this project, we are building a molecular tumor board application that allows physicians and clinical geneticists to discuss and stratify patients based on a holistic view of their integrated clinical and molecular data,

enriched by knowledge from relevant databases. We build a knowledge graph with patient information (clinical history, genetic variants, biological processes, druggability information) and a vector database with relevant literature; we also allow connecting to relevant web APIs through a tool-calling module. Those capabilities are available to the users via two distinct web applications with distinct purposes. The modular design of the developer toolkit can accommodate diverse resources and workflows. A vignette describing this use case, with links to live web applications, is available in the BioChatter documentation (https://biochatter.org/vignettes/custom-decider-use-case/).

This generic framework can be extended to workflows involving arbitrarily many interconnected LLM- and non-LLM nodes with complex logic. Collecting these workflows in an open-source library has the advantage of federating work on commonly used workflows and sharing expertise and experience across actors in the space of QCP and TS. Particularly when adding tool use capabilities (https://biochatter.org/features/rag/#api-calling), the potential of such workflows for saving time and nerves is considerable. Combined with customizable graphical user interfaces, which we provide open-source and with high flexibility to account for different workflows, we can facilitate access to these technologies for a wide range of clinical and translational scientists.

Importantly, this workflow is not without limitations. It is becoming clear that our current LLMs are not fully reliable; even for the simplest tasks, success cannot be guaranteed [31]. For this reason, we include an extensive benchmarking suite with the BioChatter library, including a living benchmark that is continuously expanded based on real-world applications. We recommend building such a benchmark for every application and all its use cases: knowledge graph or database queries, vector database retrieval, API calling, handover of tasks in agentic workflows, comprehension of the user's wishes, and many more. Particularly in solutions that are deployed in the real world, monitoring of the agentic systems' performance is elementary for building trust in these systems in the early phase of adoption. Otherwise, these systems may cause a wave of disillusionment and lack of adoption after the initial euphoria subsides [31].

7 | Future Directions and Considerations

The future of agentic workflows in QCP and TS will be shaped by the development of domain-specific agents finely tuned to operate with precision and relevance within these fields. To achieve this, these agents will need to be trained on specific datasets and integrated seamlessly with tools and processes, like electronic health records and laboratory systems. The success of this endeavor hinges on collaboration between experts in the field, AI researchers, clinicians, and industry professionals. Domain experts offer deep knowledge of pharmacological data and clinical practices, which is essential for creating scientifically accurate models. AI researchers bring expertise in ML and data processing, helping to adapt the technology for QCP and TS needs. Meanwhile, pharmaceutical companies and healthcare organizations provide practical insights for implementing and scaling these systems, as well as the resources needed for widespread use. By fostering interdisciplinary collaboration and pooling resources, the QCP and TS communities can accelerate the development of these systems, addressing challenges such as data interoperability and model validation.

Additionally, embracing collaborative development and open-source frameworks, particularly those focused on agentic workflows, can give researchers a solid starting point for building new ideas. These tools make it easier to share progress, solve problems together, and improve transparency. A great example of this is how open-source projects like TensorFlow and PyTorch have helped AI development by being available to everyone and allowing people to contribute. Adopting this kind of approach can ensure that advancements in agentic workflow implementation in QCP and TS are shared, avoid repeating work, and make faster progress.

As these workflows become more common in QCP and TS, it is important to tackle potential biases in AI models, especially those involving patient data and treatment decisions. Bias can happen when the data used to train models do not fully represent all groups, which could lead to unequal care for some populations. To prevent this, it is important to regularly check, test, and adjust the models to keep things fair. One way to do this is by including more diverse datasets, using fairness techniques in the algorithms, and bringing ethicists into the development process. By focusing on these ethical standards, we can make sure patient care is fair and build trust in AI systems among clinicians and patients alike.

As these workflows evolve, regulatory considerations will play an increasingly important role in balancing innovation with patient safety and data integrity. Recent regulatory initiatives, such as the FDA's January 2025 draft guidance, "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products," [32] underscore the critical importance of AI model risk assessment. The guidance emphasizes that a thorough understanding of the model's training data, performance metrics, and decision-making processes is necessary to ensure reliability and transparency. In the context of agentic workflows in QCP and TS, it is imperative to integrate systematic risk assessment methodologies. Such methodologies should involve rigorous validation and benchmarking, detailed documentation of model development and updates, and the implementation of robust risk mitigation strategies. These measures not only align with regulatory expectations but also enhance trust and reliability, ultimately ensuring that AI-driven analyses meet the high standards required for clinical decision-making and regulatory acceptance.

Furthermore, as agentic workflows gain prevalence, there will be an increasing demand for regulatory frameworks that can keep pace with technological advancements while ensuring patient safety and data integrity. Engaging proactively with regulatory bodies like the FDA or EMA can help shape policies that balance innovation with oversight. Establishing standardized benchmarking protocols and compliance guidelines will be vital for sustaining the performance and reliability of these dynamic systems. Long-term monitoring, coupled with periodic reviews and updates, will ensure that agentic workflows continue to meet regulatory requirements and adapt to new scientific insights.

8 | Conclusion

The integration of AI agents and agentic workflows into QCP and TS represents a significant leap forward in processing complex datasets, making informed decisions, and accelerating research and development. However, the path to widespread adoption is accompanied by challenges, including ensuring data privacy and security, overcoming technical limitations, and achieving regulatory acceptance. Overcoming these challenges will take a team effort, bringing together experts from different fields to collaborate. Additionally, developing specialized agents and setting up strong regulatory guidelines are also key steps. By working with domain experts, AI researchers, clinicians, and industry leaders, we can create agentic workflows that are both scientifically sound and practical for real-world use. Moreover, supporting open-source projects and sharing resources will help speed up innovation and ensure that progress is accessible to everyone in the field.

In conclusion, looking ahead, it is important to balance technological innovation with ethical considerations and regulatory standards to advance QCP and TS. It is crucial to understand potential biases in AI models and collaborate with regulators as well as domain experts and AI researchers to ensure fairness, patient safety, and data integrity while using AI models. Agentic workflows hold a lot of promise for unlocking new innovations to our workflows that could lead to better patient care and more efficient drug development processes.

Conflicts of Interest

Mohamed H. Shahin and Brian W. Corrigan are employees of, and may own stock/options in, Pfizer, and Srijib Goswami is an employee of, and may own stock/options in, InsightRx Inc. As an Associate Editor for *Clinical & Translational Science*, Mohamed Shahin was not involved in the review or decision process for this paper.

References

1. B. W. Corrigan, P. R. Mayo, and F. Jamali, "Application of a Neural Network for Gentamicin Concentration Prediction in a General Hospital Population," *Therapeutic Drug Monitoring* 19 (1997): 25–28.

2. P. Szolovits, R. S. Patil, and W. B. Schwartz, "Artificial Intelligence in Medical Diagnosis," *Annals of Internal Medicine* 108 (1988): 80–87.

3. J. V. Gobburu and E. P. Chen, "Artificial Neural Networks as a Novel Approach to Integrated Pharmacokinetic—Pharmacodynamic Analysis," *Journal of Pharmaceutical Sciences* 85 (1996): 505–510.

4. S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, et al., "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice," *BMC Medical Education* 23 (2023): 689.

5. M. H. Shahin, A. Barth, J. T. Podichetty, et al., "Artificial Intelligence: From Buzzword to Useful Tool in Clinical Pharmacology," *Clinical Pharmacology and Therapeutics* 115 (2024): 698–709.

6. N. Terranova, D. Renard, M. H. Shahin, et al., "Artificial Intelligence for Quantitative Modeling in Drug Discovery and Development: An Innovation and Quality Consortium Perspective on Use Cases and Best Practices," *Clinical Pharmacology and Therapeutics* 115 (2024): 658–672.

7. J. Clusmann, F. R. Kolbinger, H. S. Muti, et al., "The Future Landscape of Large Language Models in Medicine," *Communications Medicine* 3 (2023): 141.

8. H. Zhou, F. Liu, B. Gu, et al., "A Survey of Large Language Models in Medicine: Progress, Application, and Challenge," arXiv preprint arXiv:2311.05112, 2023.

9. Y. Gao, Y. Xiong, X. Gao, et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," arXiv. abs/2312.10997, 2023.

10. J. Li, Y. Lai, W. Li, et al., "Agent Hospital: A Simulacrum of Hospital With Evolvable Medical Agents," arXiv:2405.02957, 2024.

11. RISA Labs, I, "RISA—Multi-Agent System," 2024.

12. Q. Wu, G. Bansal, J. Zhang, et al., "AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation," arXiv:2308.08155, 2023.

13. Z. Durante, Q. Huang, N. Wake, et al., "Agent AI: Surveying the Horizons of Multimodal Interaction," arXiv:2401.03568, 2024.

14. K. Swanson, W. Wu, N. L. Bulaong, J. E. Pak, and J. Zou, "The Virtual Lab: AI Agents Design New SARS-CoV-2 Nanobodies With Experimental Validation," bioRxiv. 2024.2011.2011.623004, 2024.

15. A. Fourney, G. Bansal, H. Mozannar, et al., "Magentic-One: A Generalist Multi-Agent System for Solving Complex Tasks," arXiv preprint arXiv:2411.04468, 2024.

16. M. D'Arcy, T. Hope, L. Birnbaum, and D. Downey, "Marg: Multi-Agent Review Generation for Scientific Papers," arXiv preprint arXiv:2401.04259, 2024.

17. P. Akkiraju, S. Beshar, and H. Korn, "AI Agents Are Disrupting Automation: Current Approaches, Market Solutions and Recommendations," 2024.

18. "Andrew Ng's Luminary Talk: A Look At AI Agentic Workflows," 2024, https://landing.ai/videos/andrew-ng-a-look-at-ai-agentic-workflows-and-their-potential-for-driving-ai-progress.

19. P. Veng-Pedersen and N. B. Modi, "Neural Networks in Pharmacodynamic Modeling. Is Current Modeling Practice of Complex Kinetic Systems at a Dead End?," *Journal of Pharmacokinetics and Biopharmaceutics* 20 (1992): 397–412.

20. A. S. Hussain, R. D. Johnson, N. N. Vachharajani, and W. A. Ritschel, "Feasibility of Developing a Neural Network for Prediction of Human Pharmacokinetic Parameters From Animal Data," *Pharmaceutical Research* 10 (1993): 466–469.

21. V. Kaul, S. Enslin, and S. A. Gross, "History of Artificial Intelligence in Medicine," *Gastrointestinal Endoscopy* 92 (2020): 807–812.

22. R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep Learning for Healthcare: Review, Opportunities and Challenges," *Briefings in Bioinformatics* 19 (2018): 1236–1246.

23. E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Transactions on Neural Networks and Learning Systems* 32 (2021): 4793–4813.

24. Z. A. Nazi and W. Peng, "Large Language Models in Healthcare and Medical Domain: A Review," *Informatics* 3 (2024): 57.

25. T. B. Brown, "Language Models Are Few-Shot Learners," arXiv preprint arXiv:2005.14165, 2020.

26. C. Peng, X. Yang, A. Chen, et al., "A Study of Generative Large Language Model for Medical Research and Healthcare," *npj Digital Medicine* 6 (2023): 210.

27. C. Ling, X. Zhao, J. Lu, et al., "Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey," arXiv preprint arXiv:2305.18703, 2023.

28. Z. Xi, W. Chen, X. Guo, et al., "The Rise and Potential of Large Language Model Based Agents: A Survey," ArXiv. abs/2309.07864, 2023.

29. S. Lobentanzer, S. Feng, N. Bruderer, et al., "A Platform for the Biomedical Application of Large Language Models," arXiv preprint arXiv:2305.06488, 2023.

30. S. Lobentanzer, P. Aloy, J. Baumbach, et al., "Democratizing Knowledge Representation With BioCypher," *Nature Biotechnology* 41 (2023): 1056–1059.

31. L. Zhou, W. Schellaert, F. Martínez-Plumed, Y. Moros-Daval, C. Ferri, and J. Hernández-Orallo, "Larger and More Instructable Language Models Become Less Reliable," *Nature* 634 (2024): 61–68.

32. FDA, "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products," 2025.