

# Scaling Laws of Scientific Discovery with AI and Robot Scientists

Pengsong Zhang<sup>1,\*</sup>, Heng Zhang<sup>2,3,\*</sup>, Huazhe Xu<sup>4</sup>, Renjun Xu<sup>5</sup>, Zhenting Wang<sup>6</sup>, Cong Wang<sup>7</sup>, Animesh Garg<sup>8</sup>, Zhibin Li<sup>9</sup>, Arash Ajoudani<sup>2</sup>, Xinyu Liu<sup>1</sup>

<sup>\*</sup>Equal contribution, <sup>1</sup>University of Toronto, <sup>2</sup>Istituto Italiano di Tecnologia, <sup>3</sup>Università di Genova, <sup>4</sup>Tsinghua University,

<sup>5</sup>Zhejiang University, <sup>6</sup>Rutgers University, <sup>7</sup>Harvard University, <sup>8</sup>Georgia Tech, <sup>9</sup>University College of London

## Abstract

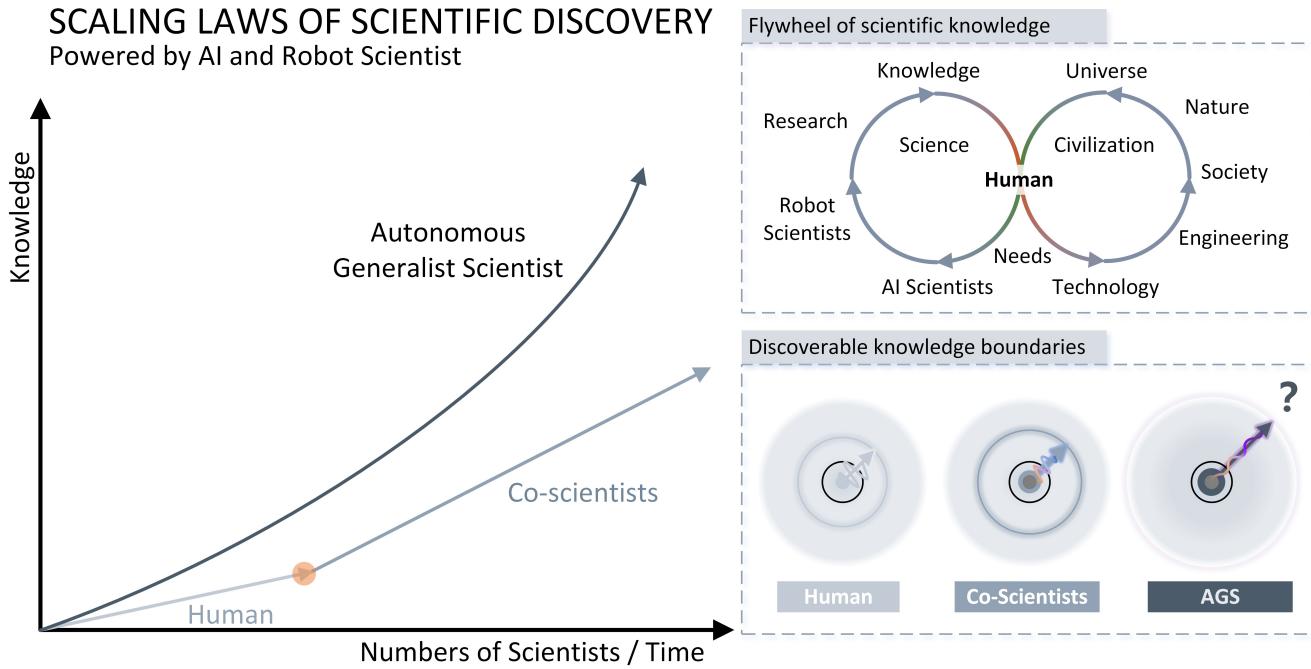
The rapid evolution of scientific inquiry highlights an urgent need for groundbreaking methodologies that transcend the limitations of traditional research. Conventional approaches, bogged down by manual processes and siloed expertise, struggle to keep pace with the demands of modern discovery. We envision an autonomous generalist scientist (AGS) system—a fusion of agentic AI and embodied robotics—that redefines the research lifecycle. This system promises to autonomously navigate physical and digital realms, weaving together insights from disparate disciplines with unprecedented efficiency. By embedding advanced AI and robot technologies into every phase—from hypothesis formulation to peer-ready manuscripts—AGS could slash the time and resources needed for scientific research in diverse field. We foresee a future where scientific discovery follows new scaling laws, driven by the proliferation and sophistication of such systems. As these autonomous agents and robots adapt to extreme environments and leverage a growing reservoir of knowledge, they could spark a paradigm shift, pushing the boundaries of what's possible and ushering in an era of relentless innovation.

## Introduction

Scientific research fuels progress by deepening our understanding of the world, sparking technological breakthroughs, and tackling global challenges like sustainability and health. However, traditional methods are slow and resource-heavy, relying on manual experiments and specialized expertise that limit innovation [33]. Multidisciplinary collaboration, while valuable for blending diverse insights, often stumbles due to coordination challenges and field-specific differences [12].

Recent advances in AI, particularly large language models (LLMs) [38], offer a game-changer. Trained on vast datasets, these tools excel at connecting knowledge across fields, speeding up tasks like literature reviews, idea generation, and writing [10]. Yet, their impact is mostly limited to digital tasks, missing the physical interactions—like lab work—crucial for fields like biology and engineering. Meanwhile, breakthroughs in robotics [4] are bridging this gap, enabling precise physical experiments that complement AI's strengths.

We propose the AGS concept: a fusion of agentic AI and embodied robotics to fully automate research. AGS aims to accelerate discovery by handling everything from hypothesis creation to experiments in both virtual and real-world settings, reducing reliance on niche expertise and boosting reproducibility. By integrating these technologies, AGS could unlock new scaling laws and a flywheel effect in knowledge growth (Fig.1), transforming how science is done. This paper presents a framework for AGS, envisioning a future where autonomous systems drive efficient, innovative research.

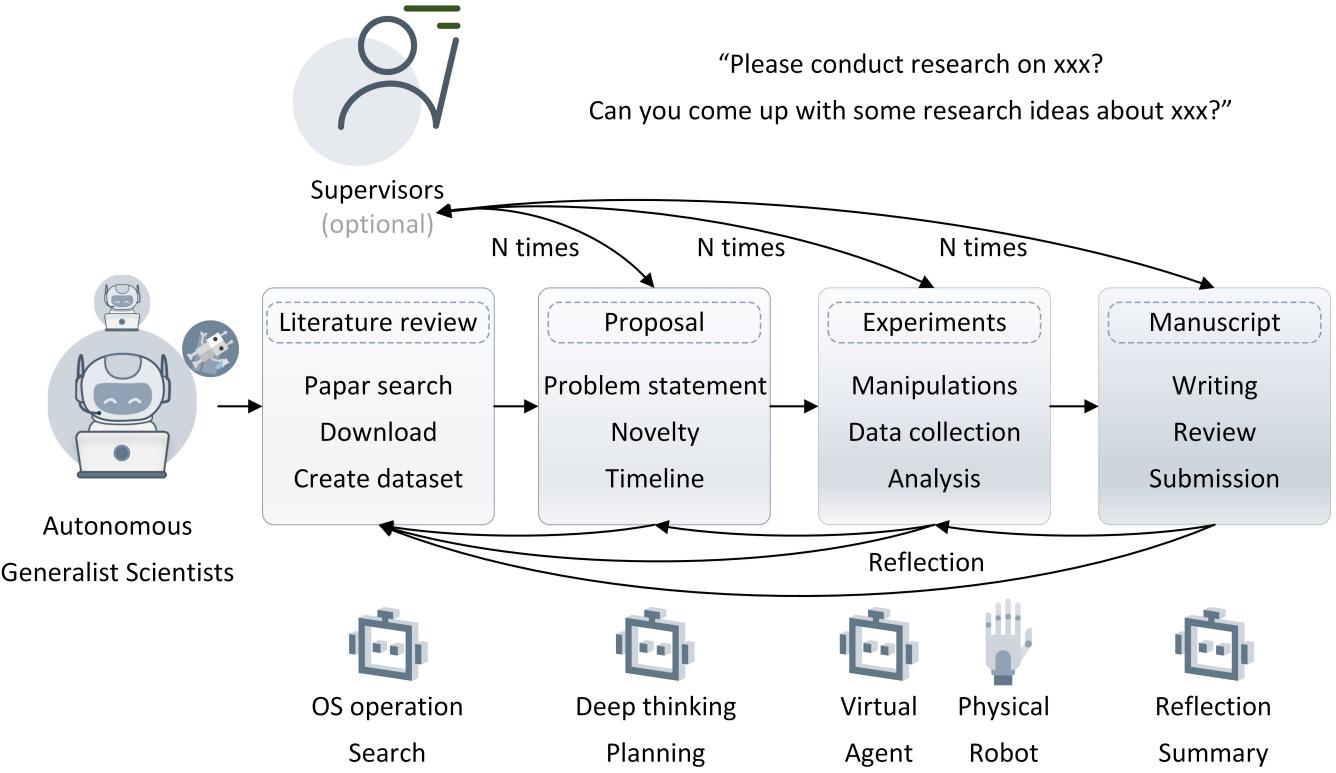


**Figure 1:** Evolution of Scientific Discovery Paradigms: From Human-Centered Research Through Collaborative Systems to Autonomous Generalist Scientists—Expanding Intellectual Horizons and Transcending Physical Limitations.

## Definition of Different Automation Levels

The concept is an AI-driven robotic system designed to conduct research across multiple fields, aiming to match and eventually exceed human scientists in speed, scope, and depth. This section defines a framework for categorizing AGS into levels based on autonomy, interaction with virtual and physical environments, and research capabilities (see Table 1). Figures 3 and 4 show its potential evolution.

The framework encompasses five distinct levels of automation. Level 0 represents traditional human-conducted research relying solely on basic computational tools like Excel or Matlab, characterized by discipline-specific focus without AI augmentation. Level 1 introduces basic AI assistance through tools like ChatGPT for literature searches or content generation, though research direction remains entirely human-driven with limited cross-disciplinary integration. Level 2 features AI agents such as Devin or DeepResearch functioning as research assistants capable of autonomously conducting online searches, executing virtual experiments, and synthesizing cross-domain knowledge, while still requiring substantial human guidance. At Level 3, AI systems work collaboratively with human researchers, incorporating robotic capabilities to conduct physical experiments in fields like biology or engineering, handling complex interdisciplinary tasks with precision while maintaining human partnership. Level 4 introduces semi-independent systems (AGIR - AGI Robots) capable of conducting sophisticated research across both virtual and physical domains with minimal oversight, generating novel insights through cross-disciplinary data integration. Finally, Level 5 represents fully autonomous Artificial SuperIntelligence Robots (ASIR) that exceed human capabilities, independently driving groundbreaking research and potentially discovering new scientific principles—though this level presents significant technical and ethical challenges and remains a future aspiration.



**Figure 2:** Integration Architecture for Autonomous Scientific Systems: AI-Robotic Collaboration Platform Enhancing Research Velocity and Facilitating Cross-Disciplinary Knowledge Synthesis.

## Roadmap to Automatic Research with AI Scientist and Robot Scientist

### Overview

The AGS offers a unified framework that blends cutting-edge AI with robotics to fully automate the research process (see Fig.2, Fig.5). Built on a multi-agent system, it pairs AI agents—driven by Large Language Models (LLMs) and machine learning—with robotic systems. The AI handles virtual tasks like coding, hypothesis creation, and data analysis, while robotics takes on physical duties, such as operating lab tools and running precise experiments. This combination speeds up research, improves accuracy, and ensures reproducible results, paving the way for a game-changing shift in multidisciplinary science.

### Literature Review

The literature review underpins research by pinpointing existing knowledge, gaps, and new possibilities. Traditionally, it relies on manual searches and analysis of countless papers—a slow process often limited by outdated data access [37]. This section traces the shift to AI-driven methods, contrasting conventional database/API approaches with advanced OS agent-driven systems that emulate human actions for complex searches and tasks in virtual settings.

Generality	Human Involvement	Interaction	Knowledge Access	Typical Examples	Description
<b>Level 0: No AI</b>	Complete human	Limited tools	Fixed	Excel, Matlab	<i>Humans do everything, no AI.</i>
<b>Level 1: Tool-Assisted</b>	Predominantly human-driven	Limited virtual env	Internal; API search	ChatGPT; Basic ML Models	<i>AI helps with simple tasks, humans lead.</i>
<b>Level 2: Assistant</b>	Significant human oversight	Virtual env	Internal; web search; virtual experiment	OpenDevin; DeepResearch	<i>AI does complex virtual tasks, humans supervise.</i>
<b>Level 3: Partner</b> <i>Reach human-level like undergraduate and master.</i>	Collaborative human-AI/Robot interaction	Virtual and physical envs	Internal; Web search; Physical experiment	Advanced Agent; Advanced Robot	<i>AI/robots handle experiments, humans guide.</i>
<b>Level 4: Researcher</b> <i>Toward human-level like PhD and Prof.</i>	Minimal human supervision	Advanced interaction with virtual and physical envs	Internal; Autonomous search; Creating knowledge	AGI Robot (AGIR)	<i>AI leads research, humans supervise and review.</i>
<b>Level 5: Pioneer</b> <i>Outperform human-level</i>	Fully autonomous	Fully autonomous interaction in all envs	Internal; Autonomous search; Groundbreaking innovations	Artificial SuperIntelligence Robot (ASIR)	<i>AI manages all research, no human needed.</i>

**Table 1:** Definition: Categorization Framework for Computational Research Autonomy.

## Challenges in Traditional Automated Literature Reviews

Conventional automated literature reviews lean on manual effort or restrictive database/API access, narrowing the scope and freshness of data. Database searches lag due to indexing delays, while API-based tools, though faster, miss the latest publications due to dataset constraints. Systems like Survey Agent [39] and AutoSurveyGPT [45] use conversational AI and GPT models to speed up reviews, and specialized tools like the AI Chatbot in Cancer Research [31] aid niche fields. Yet, their reliance on pre-set data sources limits access to cutting-edge findings, a critical issue in fast-moving disciplines.

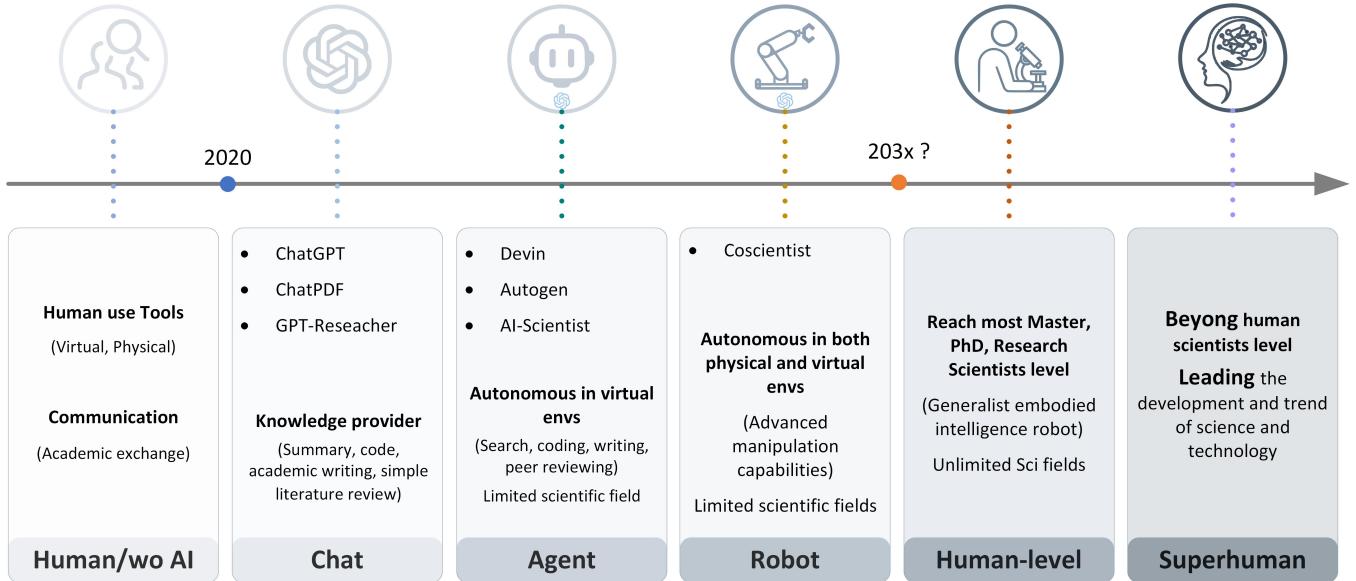
## OS Agents for Enhanced Reviews

To overcome these hurdles, OS agents mimic human-like interactions with digital platforms, moving beyond static API limits. Tools like GPT-4 Vision [53] handle complex web tasks, while OS-Copilot [43] adapts and self-improves in real computer environments. Multimodal agents, such as VisualWebArena [22] and OSWorld [46], navigate sites and fetch up-to-date data directly. These systems dynamically gather comprehensive, current information, strengthening the foundation for research.

## Comparing Literature Review Approaches

Table 3 contrasts three strategies: using internal databases or model knowledge, querying via search APIs, and simulating human interactions for autonomous searches and manipulations. It showcases their strengths in data retrieval and virtual task performance, highlighting OS agents' edge in flexibility and timeliness.

As shown in Table 3, OS agents excel at accessing up-to-date data and performing tasks like human-like web

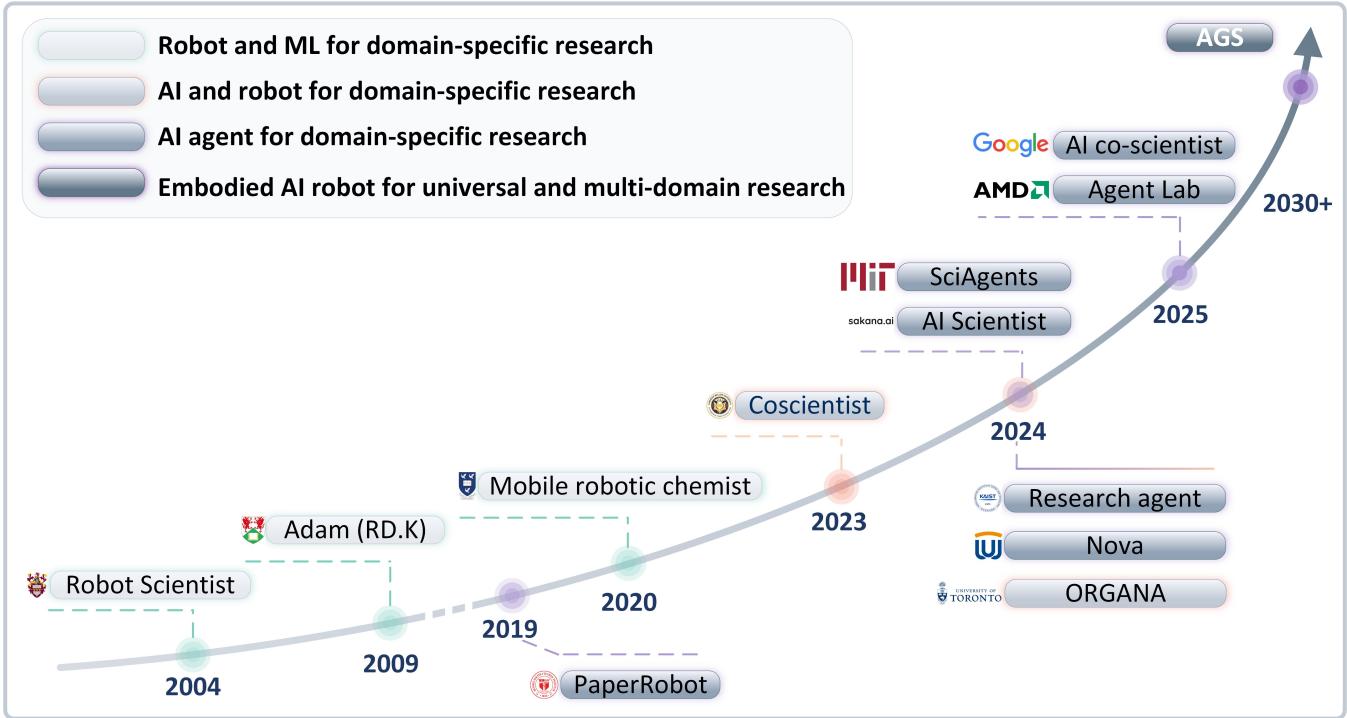


**Figure 3:** Timeline for different automation levels of automatic research.

Project/ Method	AI	Robot	Role	Literature review	Proposal	Virtual Experiment	Physical Experiment	Writing Review	Domin
Adam(2009)	x	Equipments	Autonomous	x	Yes (limited)	x	v	x	Single (Biology)
PaperRobot (2019)	ML	x	Assistant	Reading	v	x	x	x	Single (Biomedical)
Coscientist (2023)	LLM agent	Equipments	Assistant	API search	Experimental design	Coding, browser	v (limited)	x	Single (chemistry)
AI Scientist (2024)	LLM agent	x	Autonomous	API search	v	Coding	x	v	Single (Machine learning)
Organa (2024)	LLM	Robot arm	Assistant	x	x		v (limited)	x	Single (chemistry)
Agent Lab (2025)	LLM agent	x	Assistant	API search	Research plan	Coding	x	v	Single (Machine learning)
AI co-scientist (google, 2025)	LLM agent	x	Assistant	Web search	v	Coding, browser	x	v	Single (Biomedical)
AGS (Future)	Agent	Embodied	Autonomous	v (General)	v (General)	v (General)	v (General)	v (General)	Unlimited

**Table 2:** Analytical Contrast Between current AI and Robot Scientists.

navigation, outpacing API-based methods limited by scholarly resource constraints. This makes them vital for comprehensive, current literature reviews—key to staying ahead in fast-moving fields. By integrating these advanced OS agents, the AGS automates the entire literature review process, retrieving, organizing, and analyzing the latest research to spot gaps and new paths forward. This shift turns the literature review from a slow step into a springboard for rapid, innovative discovery, seamlessly linking it to the full research cycle.



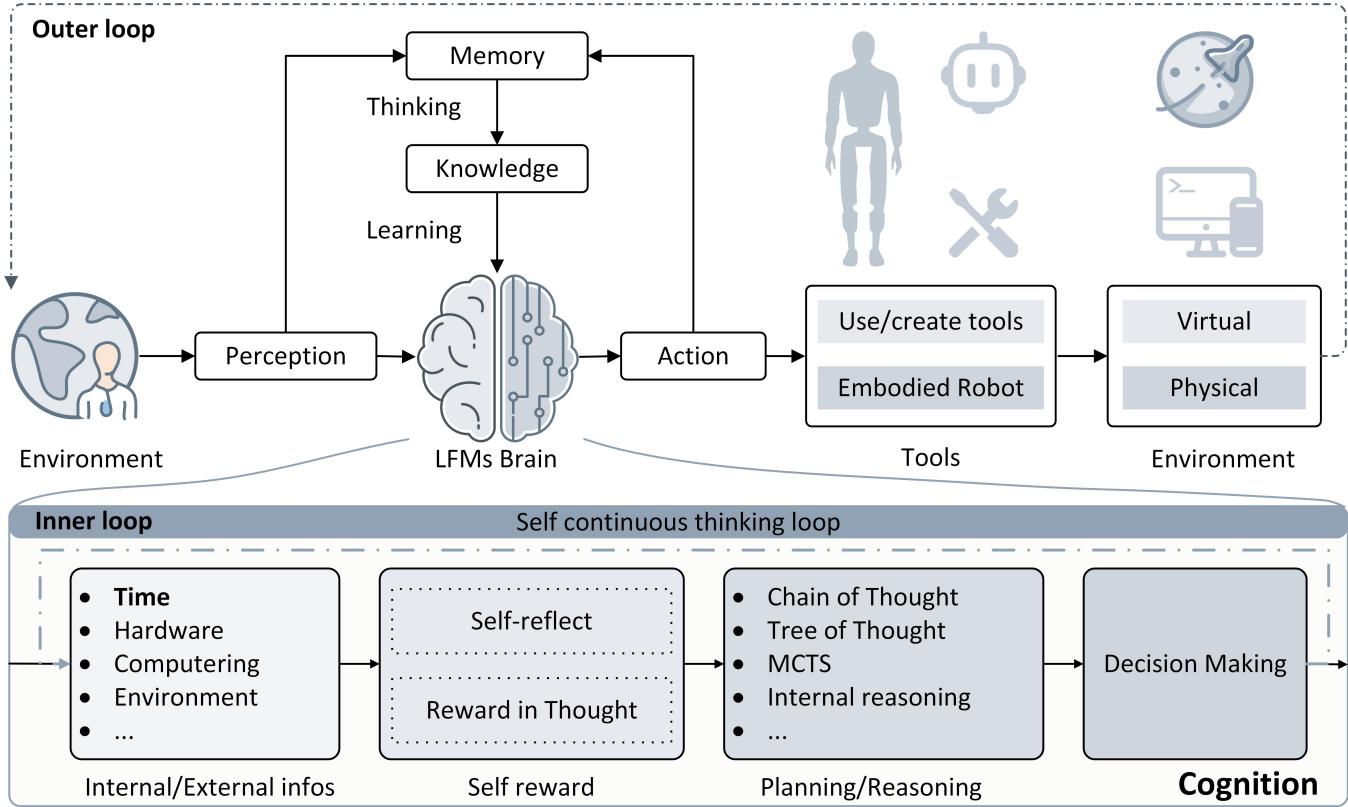
**Figure 4:** Evolution for AI Scientist, Robot Scientist, and AGS .

Methods	Searching abilities	Manipulation task
Database	Outdate data	✗
API	Limitation latest data from providers with open api	✗
OS agent	Latest data from virtual environment	v (Virtual environment) Manipulate all kinds of softwares like human. (use mouse, keyboard, etc.)
AGS (agent and robot-based)	Latest data from virtual and physical environments	v (Virtual and <b>Physical</b> environments) Embodied Intelligence

**Table 3:** Methodological Comparison: Approaches to Information Retrieval and Manipulations Across Research Contexts

## Proposal Generation

A research proposal maps out a study, pinpointing the problem and detailing a plan to tackle it. In NLP, LLM-generated ideas often outshine human ones in novelty [34]. Within the AGS, advanced AI automates this process, iteratively refining it beyond mere idea generation [5]. The system crafts problem statements, hypotheses, methodologies, and project plans by analyzing literature gaps, designing robust approaches, and incorporating feedback. This ensures proposals are innovative and aligned with current research needs.



**Figure 5:** The framework of AGS.

## Problem Statement

Crafting a research proposal starts with a sharp problem statement. The AI system scans literature review results, spotting key gaps in knowledge. By blending insights from various sources, it shapes a statement that tackles these gaps with a bold, relevant question. Using topic modeling and clustering, the system narrows broad topics into focused queries, ensuring the research is both timely and set to advance the field.

## Hypothesis and Methodology

Following problem definition, the AI system generates hypotheses by evaluating research concepts against existing literature to identify originality and unexplored relationships. The system then develops testing methodologies, selecting appropriate research designs and analytical approaches that ensure scientific rigor and feasibility. By incorporating best practices from its literature analysis, the AI enhances the research plan's reliability and structural soundness.

## Planning

The AI system leverages literature review data to shape research design and project development, extracting valuable methodologies while identifying and incorporating mitigation strategies for potential challenges. The planning module establishes a detailed timeline with efficient resource allocation and clear research milestones. Through simulation of various research trajectories and their potential outcomes, the system enables a flexible

project management approach that maintains progress while adapting to emerging discoveries.

## **Iterative Refinement by Communication, Feedbacks**

The AI system facilitates effective communication within the proposal development process, presenting ideas clearly to supervisors, users, or peer agents for evaluation. By collecting and integrating feedback, the system refines the proposal, incorporating valuable insights and addressing identified weaknesses. Through its multi-agent peer review mechanism, multiple AI agents critically evaluate the proposal, challenge assumptions, and suggest enhancements. This collaborative refinement ensures the final research proposal is comprehensive, rigorously examined, and submission-ready.

## **Innovation and Research Gap Alignment**

The AI system continuously evaluates proposal innovation by assessing originality, feasibility, and knowledge advancement potential. By directly connecting proposed research to gaps identified in the literature review, the system enhances the proposal's relevance and potential impact. This strategic positioning ensures research initiatives are both innovative and targeted toward addressing critical disciplinary questions.

## **Experimentation**

### **The Challenges in Scientific Experimentation**

Research extends beyond computational analysis into physical experimentation with real-world materials as demonstrated in Table 4. Traditional methods depend on human expertise for experimental design and execution—a resource-intensive approach. While AI has advanced significantly in data-centric applications, physical experimentation automation remains underdeveloped. AI tools for data science [17] and LLM-based scientific discovery systems [27] operate primarily in virtual environments, limiting their real-world experimental utility. These constraints highlight the need for embodied intelligent robots capable of performing complex physical tasks with human-like precision. Current robotic platforms struggle with generalization, typically handling only narrow, predefined tasks without the flexibility to adapt across different experimental contexts [48].

### **Current Advances and Remaining Challenges**

Existing robotic systems in autonomous chemical research [26] demonstrate progress in AI-guided experimentation but lack cross-domain flexibility. While LLM-driven platforms show potential for automating research processes [18], they remain limited in physical manipulation capabilities.

LLM-based embodied AI represents a promising path toward real-world AGI by connecting high-level reasoning with physical interaction. These systems leverage textual knowledge to enhance robot comprehension of complex instructions and contextual decision-making. Platforms like OpenDevin [11] offer glimpses of future generalist agents, though their application to physical sciences remains challenging. Vision-language models [28] and public robotics datasets [30] support this development but face generalizability limitations.

Concurrently, world model research focuses on environmental interaction learning, enabling robots to form internal representations of surroundings and predict action outcomes. These models, using sensor data and machine learning techniques, allow robots to navigate complex environments by simulating scenarios and selecting optimal strategies. Works like [41] demonstrate progress in this area.

Main	Research field	Subfields	Virtual manipulation	Physical manipulation	○V/●P
Natural Science	Physics	Classical Physics, Quantum Physics, Particle Physics, Astrophysics	Theoretical modeling, numerical simulation, data analysis	Equipment operation, experimental measurement, sample preparation	○ ○ ○ ○ ● ●
	Chemistry	Physical Chemistry, Organic Chemistry, Inorganic Chemistry, Materials Chemistry	Molecular modeling, reaction prediction, data analysis	Synthesis, characterization, property testing	○ ● ● ● ●
	Biology	Molecular Biology, Cell Biology, Genetics, Ecology	Bioinformatics analysis, system modeling, data processing	Laboratory experiments, microscopy, field work	○ ● ● ● ●
	Earth	Geology, Meteorology, Oceanography, Environmental Science	Environmental modeling, data analysis, system simulation	Field investigation, sample analysis, monitoring	○ ○ ● ● ●
Formal Sciences	Mathematics	Pure Mathematics, Applied Mathematics, Statistics, Mathematical Physics	Theoretical derivation, numerical computation, modeling	Data collection, verification, demonstration	○ ○ ○ ○ ○ ●
	Computer	Theoretical CS, AI/ML, Software Engineering, Cybersecurity	Algorithm development, system design, software programming	Hardware testing, system deployment, maintenance	○ ○ ● ● ●
Applied Sciences	Engineering	Mechanical, Electrical, Chemical, Civil Engineering	Design modeling, simulation, optimization	Manufacturing, testing, system integration	○ ○ ● ● ●
	Medical	Clinical Medicine, Biomedical Engineering, Pharmacy	Medical imaging, data analysis, treatment planning	Clinical practice, laboratory testing, patient care	○ ● ● ● ●
	Agricultural	Agronomy, Horticulture, Animal Science	Growth modeling, system simulation, data analysis	Field experiments, breeding, cultivation	○ ● ● ● ●
Social Sciences & Humanities	Social	Economics, Sociology, Psychology, Political Science	Data analysis, behavioral modeling, simulation	Field research, experiments, surveys	○ ○ ○ ○ ● ●
	Humanities	Philosophy, History, Literature, Arts	Digital analysis, data mining, archival processing	Field investigation, artifact analysis, creation	○ ○ ○ ● ● ●
Interdisciplinary Science	Bioinformatics	Genomics, Systems Biology, Computational Biology	Computational analysis, modeling, prediction	Experimental validation, data collection	○ ○ ○ ○ ○ ●
	Cognitive	Neuroscience, Psychology, AI, Linguistics	Cognitive modeling, data analysis, simulation	Brain imaging, behavioral experiments, testing	○ ○ ○ ● ● ●
	Environmental	Environmental Science, Sustainability, Climate Studies	Environmental modeling, impact assessment, analysis	Field monitoring, sampling, experimental studies	○ ○ ○ ○ ● ●
	Nanotech	Nanomaterials, Nanoelectronics, Nanobiotechnology	Nanostructure simulation, process modeling	Material synthesis, characterization, testing	○ ● ● ● ●

**Table 4:** Virtual and Physical Manipulation Needs for Scientific Research. This table illustrates the characteristic requirements for virtual and physical manipulation across diverse scientific disciplines. The V/P ratio (rightmost column) represents general tendencies rather than precise quantitative measurements, highlighting the relative emphasis typically placed on computational versus experimental approaches in each field.

Combining world models with LLM-based embodied AI marks a significant step forward for general-purpose robotics, as it unites structured environmental representations with sophisticated cognitive capabilities. As shown in [3], this powerful connection allows robots to anchor their decisions within physical constraints while effectively processing complex language inputs.

The primary obstacles facing general robotic systems in physical environments include:

- **Robotic Perception and Manipulation.** General-purpose robots require sophisticated environmental

Consideration Factor	Description
<b>Diversity &amp; Flexibility</b>	The robot requires versatile manipulation capabilities, enabling it to seamlessly switch between diverse end-effectors or utilize human-like tools to perform a broad spectrum of experimental tasks across various disciplines.
<b>Precision &amp; Reliability</b>	Scientific experiments require exceptional precision and dependability. The robot needs to deliver consistent manipulation performance, reducing the likelihood of errors and malfunctions.
<b>Adaptability &amp; Learning Ability</b>	The robot must be capable of acquiring new skills and adjusting to different environments, utilizing reinforcement learning or other continue learning methods to fulfill the changing demands across multiple fields.
<b>Human-Robot Collaboration</b>	The robot should be able to collaborate with human scientists, demanding proficient communication and interaction abilities, such as comprehending instructions and offering feedback.
<b>Integration of Virtual &amp; Physical Manipulation</b>	The robot should effectively combine virtual manipulation skills, such as simulations and data analysis, with physical manipulation tasks to ensure smooth operation in both digital and physical settings.
<b>Safety</b>	When dealing with dangerous substances or conducting high-risk activities, the robot needs to be equipped with strong safety features, including emergency shutdowns, fault detection systems, and responsive mechanisms.

**Table 5:** Comprehensive Considerations for Manipulation Needs.

awareness and interaction capabilities [44, 52]. This encompasses accurate object recognition, spatial localization, and precise manipulation. Effective robotic systems depend on integrated sensor arrays and advanced actuator mechanisms that enable detailed environmental perception and fine-grained control precision.

- **Self-Directed Operation and Reasoning.** Effective robotic systems must demonstrate independent reasoning and task execution capabilities [42]. This necessitates sophisticated planning algorithms, contextual reasoning frameworks, and adaptive learning mechanisms. Modern robots must navigate dynamic environments, identify and circumvent obstacles, and respond appropriately to changing operational conditions. Research initiatives like [51] are advancing autonomous decision-making frameworks that enable robots to independently plan and execute complex task sequences.
- **Knowledge Transfer and Environmental Responsiveness.** Next-generation robots must transfer knowledge between domains and apply previous learning to unfamiliar scenarios [21]. This requires sophisticated learning architectures capable of cross-domain knowledge application. Truly versatile robotic platforms must demonstrate flexibility across diverse operational environments and task requirements. Contemporary research such as [8] focuses on developing learning frameworks that maximize generalization from limited training examples and enhance adaptation to novel contexts.
- **Operational Security.** Human-robot collaboration introduces potential safety concerns, particularly in unstructured environments [42]. Ensuring robots operate safely while manipulating objects near humans remains a critical priority. Research initiatives like [51] emphasize developing safety-oriented behaviors through real-time environmental sensing and risk-aware learning models. Robots operating in shared spaces must make rapid safety assessments during task execution. Advanced systems including DeepMind’s AutoRT [1] implement comprehensive safety protocols, such as force limitation mechanisms and human-proximity operational constraints. The SafeVLA framework [50] integrates safety considerations

into vision-language architectures to protect environmental elements, hardware systems, and human collaborators.

- **Human Engagement.** Creating natural robot-human communication remains technically challenging [2], requiring advanced natural language processing [32], emotional recognition capabilities, and non-verbal communication understanding. Robots must adapt to established social conventions and interaction protocols. Successful embodied AI depends on seamless human-robot engagement. This includes interpreting emotional states, understanding physical gestures, and recognizing social dynamics—all representing active research challenges.
- **Ethical and Legal Frameworks.** Increasing robot autonomy raises significant ethical questions regarding decision processes and potential harm risks [35]. Critical considerations include responsibility allocation, privacy protection, and ethical data utilization. Robots interacting with humans must demonstrate sound ethical and moral reasoning. This becomes particularly significant in sensitive contexts like healthcare and eldercare where human wellbeing is directly impacted [16].

## AI Agent And Embodied Robot for Virtual and Physical Experimentation

The integrated AGS artificial intelligence with advanced robotics to streamline experimental processes across computational experiments and laboratory environments. Drawing upon contemporary innovations, including language model applications for experimental parameter optimization [27] and breakthroughs in precision robotic handling of research materials [23], this comprehensive framework aims to develop a highly versatile and resource-efficient approach to scientific investigation.

The generalist scientist integrates AI with robotics to streamline experimental processes across virtual and physical domains. Building upon recent advances in LLM-based experimental optimization [27] and robotic manipulation of scientific materials [23], this framework enhances experimental adaptability and efficiency.

- **Experimental Design and Implementation:** The system interprets research proposals, identifies experimental requirements, and develops execution strategies. Beyond methodological selection, it orchestrates resource allocation and task scheduling to maximize efficiency. The framework incorporates Chain of Thought reasoning [40] to break down complex experimental procedures into logical, sequential operations, ensuring methodological coherence throughout the process.
- **Advanced Robotics Physical Manipulation for lab work:** The framework deploys intelligent robotic systems capable of sophisticated real-time adjustments during experiments. These systems extend current capabilities demonstrated in platforms like ORGANIA [13], which automates chemical experimentation while facilitating researcher collaboration to refine protocols and adapt to emerging results.
- **Computational Simulation:** Pre-experimental virtual modeling, similar to approaches in [19], enables scenario testing before physical implementation. This computational approach reduces resource expenditure while enhancing outcome reliability through proactive identification of experimental challenges.
- **Adaptive Resource Coordination:** Efficient experimental resource utilization, combined with feedback-based protocol adjustment capabilities, underpins successful scientific investigation. Integration frameworks such as ROS-LLM [29] demonstrate how language models can enhance robotic systems for structured experimental reasoning and dynamic decision processes.
- **Scientific Reproducibility Assurance:** The framework prioritizes experimental reproducibility and accuracy as scientific cornerstones. Through calibrated robotic systems and adaptive AI models, it

maintains result consistency across varied experimental conditions, reflecting the validation approaches utilized in Chemistry3D [23].

Methods	Environments	Tasks	Areas
<b>Agent</b>	Virtual	Computer use: coding, data analysis, writing, plot, etc.	Computer science, Math, Bioinformatics, etc.
<b>Robot</b>	Physical and Virtual	Virtual and physical manipulation abilities: creating and using tools.	Engineering like medical, biology, chemistry, space, etc.

**Table 6:** Contrasting AI Agents and Embodied Robot for Scientific Investigation

## Manuscript Preparation

### The Role of Manuscript Preparation in Research

The documentation and presentation of research findings represents a crucial component in the scientific workflow. This stage involves synthesizing experimental outcomes, organizing information logically, and communicating discoveries effectively to the academic community. Researchers traditionally face numerous challenges during this process, including ensuring factual precision, complying with disciplinary conventions, and articulating complex concepts in accessible language.

### Automated Manuscript Drafting

Within our autonomous generalist scientist framework, we propose utilizing state-of-the-art AI systems to bridge the gap between experimental results and scholarly documentation. This approach aims to create intelligent systems capable of producing preliminary manuscript drafts that organize research findings into structured sections following established academic conventions.

- **Experimental Data Processing:** The proposed system initially aggregates experimental outcomes, generates analytical summaries, and develops visual representations through figures and tables. Drawing inspiration from tools like MatPlotAgent [47] and Data Interpreter [36], the framework emphasizes result presentation that accentuates significant research discoveries. The system optimizes data visualization to enhance clarity and impact of experimental findings.
- **Diversified Content Integration:** The framework incorporates capabilities for producing varied scientific content formats, including quantitative tables, analytical graphs, procedural illustrations, and visual demonstrations. It generates professional-grade visualizations using multiple representational approaches, creates procedural demonstrations for complex methodologies, and develops interactive visual tools when applicable. This comprehensive approach enables multifaceted presentation of research outcomes across complementary formats.
- **Citation Coordination:** Bibliography management represents another key function where the system demonstrates particular utility. Through integration with citation management platforms, the AI ensures proper formatting of scholarly references while maintaining comprehensive and current bibliographic records [9]. The system verifies citation accuracy throughout the documentation process.

- **Documentation Support:** Following data analysis, the AI facilitates manuscript development. It generates section-specific content for introductory context, methodological procedures, empirical outcomes, and interpretative discussion. The system additionally supports mathematical formula representation and ensures compliance with publication-specific formatting requirements for document consistency and scholarly presentation.

## Peer Review

To strengthen manuscripts before submission, the framework incorporates comprehensive evaluation protocols to identify and address potential weaknesses, ensuring adherence to scholarly standards.

- **Internal Review Mechanisms:** The system employs dual evaluation approaches through reflexive assessment and collaborative agent critique. Its reflexive components evaluate argumentative structure, methodological robustness, and expositional clarity. Concurrently, specialized evaluation agents scrutinize distinct manuscript elements—including statistical methodology, experimental protocols, and theoretical frameworks—offering multifaceted improvement perspectives [24].
- **Independent Assessment:** The framework facilitates engagement with external evaluators, including both AI systems and human specialists, to secure objective manuscript evaluation. These external review mechanisms simulate journal evaluation procedures, providing comprehensive feedback on scientific contribution, originality, and research significance. The system additionally facilitates human expert collaboration, integrating specialized knowledge to enhance document quality [14].
- **Research Ethics Framework:** AI integration in scientific documentation raises important considerations regarding attribution and content authenticity. While language models demonstrate significant writing assistance capabilities, they present potential risks of generating inaccurate information or "hallucinations" [25]. The proposed framework incorporates governance protocols ensuring human researchers maintain appropriate oversight and responsibility for content development, preserving scholarly integrity throughout the documentation process.

## Finalization and Submission

The culminating phase involves comprehensive review, formatting refinement, and submission coordination. The framework streamlines these final processes to facilitate efficient manuscript publication.

- **Journal-Specific Formatting:** The system implements precise formatting protocols according to target publication guidelines. This ensures all document elements—from typographical specifications to visual content placement—conform to journal requirements. The system applies appropriate reference styles, section organization, and visual presentation standards to meet publication criteria.
- **Submission Process Management:** The framework facilitates publication submission workflows. It manages submission documentation, coordinates file transfers, and processes editorial communications including revision requests. This automated approach streamlines interactions with publishing platforms while maintaining document integrity throughout the submission process.

## **Advantages of Automated Manuscript Preparation**

The integration of AI in scientific documentation aims to accelerate publication timelines while maintaining quality standards, making academic publishing more accessible across research communities.

- **Temporal Optimization:** Automation significantly reduces documentation development cycles—from initial drafting through refinement to submission—allowing researchers to redirect efforts toward core scientific activities and intellectual development.
- **Quality Standardization:** By employing AI for reference management, document formatting, and visualization generation, the framework minimizes technical inconsistencies while ensuring compliance with established academic conventions. This systematic approach enhances document reliability and professional presentation.
- **Expanded Participation:** The framework's automation capabilities democratize access to professional document preparation, particularly benefiting researchers with limited institutional resources or editorial support. This technological approach helps level the academic publishing landscape by reducing technical barriers to manuscript development.

## **Reflection and Feedback**

### **The Need for Communication and Reflection**

Within the generalist scientist architecture, comprehensive information exchange and analytical self-assessment serve as critical components ensuring cohesive research progression. These mechanisms facilitate knowledge transfer between research phases, similar to collaborative dynamics in human research teams. Strategic module interaction coupled with systematic process evaluation enhances hypothesis formulation, methodological precision, and scientific output quality. This section examines how the automated generalist scientist incorporates these interactive elements to maximize research effectiveness and innovation potential.

### **Self-Reflection Mechanism**

Self-reflection is a key component in the generalist scientist auto-research framework, enabling the system to evaluate its performance and make necessary adjustments to improve future research outcomes.

Analytical self-assessment represents a fundamental element within the integrated research automation architecture, allowing continuous monitoring and iterative enhancement of scientific processes. This systematic evaluation facilitates performance analysis and strategic adjustments to strengthen subsequent investigative outcomes.

- **Analytical Self-Assessment:** Drawing from contemporary AI self-evaluation research, the framework implements comprehensive performance monitoring protocols. This evaluation encompasses output accuracy verification, data relevance examination, and research objective alignment analysis. Through these systematic assessments, the system identifies and addresses potential inaccuracies, including content generation errors or factual misrepresentations [20].
- **Iterative Enhancement Mechanisms:** Through systematic self-evaluation integration, the research platform progressively refines its operational capabilities. This iterative approach incorporates insights from previous research cycles, adjusts methodological approaches, and enhances computational processes to generate increasingly reliable and scientifically sound outcomes [15].

## **Inter-module Communication**

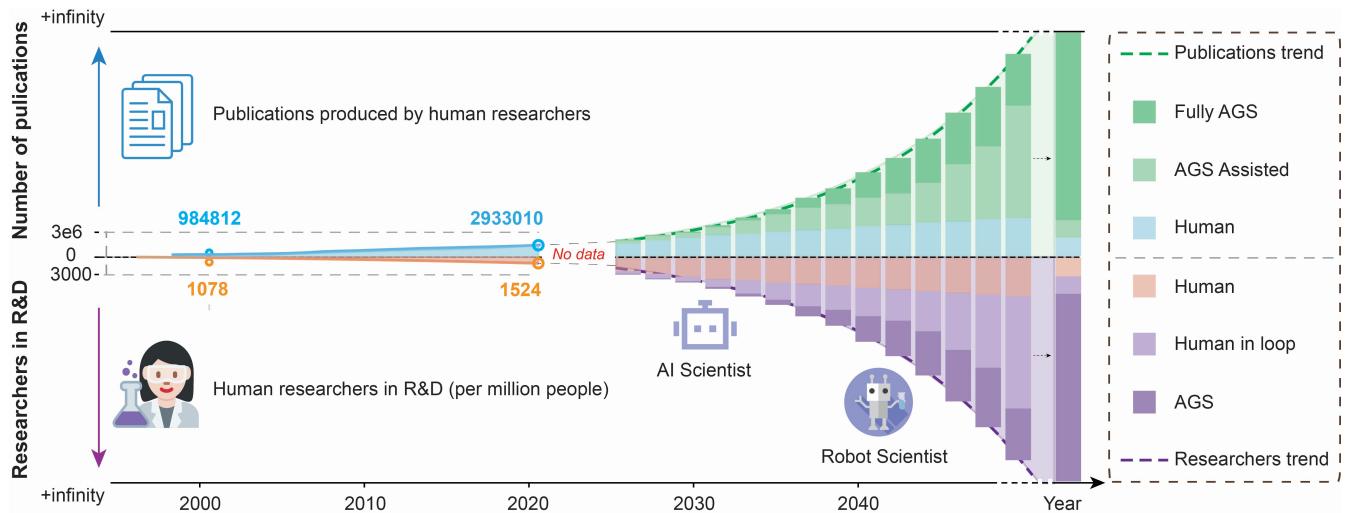
The framework architecture mirrors collaborative scientific teams, facilitating information flow between research phases to enhance overall investigative quality.

- **Bidirectional Information Exchange:** The system establishes interconnected pathways between functional components (literature analysis, proposal development, experimentation, documentation), creating comprehensive feedback networks. Experimental outcomes inform documentation priorities and emphasis, while literature discoveries trigger proposal refinements. This dynamic communication structure ensures research coherence through continuous information updates between specialized modules.
- **Progressive Development Cycles:** The framework implements cyclical refinement processes, systematically enhancing research hypotheses, methodological approaches, and scientific outputs based on emerging data. This structured iterative approach maintains research relevance through progressive improvement, ensuring each development stage builds upon accumulated knowledge from previous phases.

## **External Perspective Integration**

The incorporation of diverse external viewpoints represents an essential research component, introducing alternative analytical frameworks and identifying potential enhancement opportunities.

- **Human Supervision Framework:** The system implements structured oversight mechanisms enabling researchers to monitor development and provide directional guidance. This human-augmented approach maintains alignment between system operations and investigator objectives while preserving scientific integrity. The collaborative interface allows researchers to shape the investigative process while leveraging computational capabilities.
- **Academic Evaluation Simulation:** The framework incorporates publication assessment modeling that replicates scholarly review processes. This virtual evaluation generates constructive critiques that inform manuscript refinement prior to formal submission, addressing potential methodological or structural weaknesses [49]. The simulation enables preemptive quality enhancement based on anticipated reviewer perspectives.



**Figure 6:** Historical Patterns[6, 7] and Projected Developments in Global Scientific Research Output and Workforce. (Note: Official World Bank Group data for the 2020-2024 period remains pending release.)

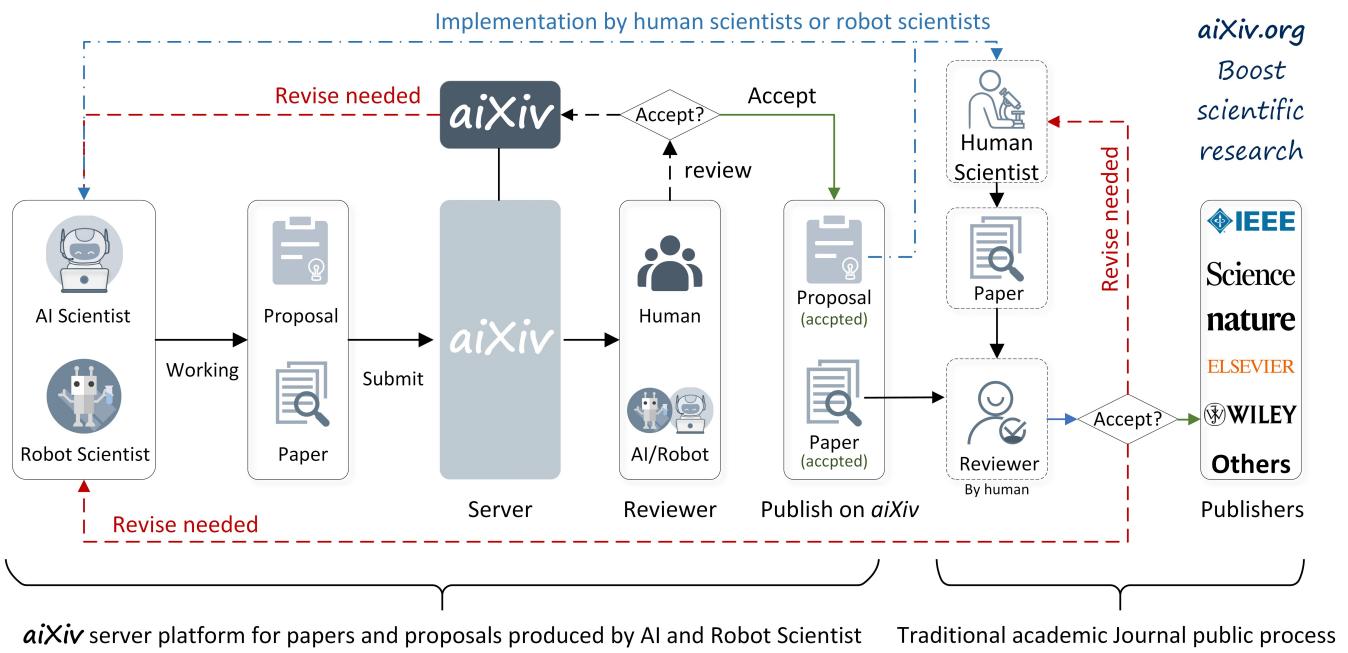
## Open Research Questions

### How to Manage Publications from AI/Robot Scientists: Do We Need an Open Platform for Preprints?

The integration of AI and robotic scientists into academic research introduces novel publication challenges beyond conventional human-centered systems. AIXIV (Fig. 7) represents a potential intermediary framework connecting autonomous research systems with established scientific publishers. This specialized platform would coordinate proposal submissions while implementing customized evaluation protocols for machine-generated research. Evaluation would involve collaborative assessment between human reviewers and AI systems examining practical feasibility, scientific novelty, and methodological soundness, while upholding research integrity standards. Though this approach could potentially enhance discovery rates, substantial obstacles persist regarding proper attribution, scientific responsibility, verification processes, and mainstream acceptance. Implementation challenges include establishing objective assessment criteria, efficient resource distribution, and platform adaptability. Human oversight likely remains essential for ensuring quality alignment with established scientific benchmarks in this transformative publication landscape.

### Is Human-Like Physical Configuration Necessary for Scientific Automation Platforms?

The necessity of humanoid design for AGS presents an intriguing technological consideration. While not fundamentally required for effective scientific function, humanoid configurations offer distinct advantages in research environments designed for human physiology and interaction patterns. Laboratory spaces and scientific equipment primarily accommodate human ergonomics, allowing humanoid robots to navigate and utilize existing instrumentation without extensive facility modifications. Advanced bilateral manipulation capabilities enable precise experimental procedures across diverse scientific disciplines, while anthropomorphic design facilitates intuitive human-robot collaboration through familiar physical characteristics and



**Figure 7:** Framework of aiXiv server platform for papers and proposals produced by AI and Robot Scientist.

movement patterns. Nevertheless, non-humanoid architectures may offer superior performance in specific contexts—providing enhanced precision, stability, or operational capability in extreme environments unsuitable for human researchers. In summary, while humanoid design represents one viable approach rather than a necessity, it offers significant advantages in laboratory integration, operational versatility, and human collaboration potential, particularly in conventional research settings.

## Can Robot Scientists Conduct Independent Scientific Inquiry in Extreme Environments Beyond Human Physical Limitations?

Robot Scientists offer significant potential to extend scientific investigation beyond Earth's boundaries into space exploration (Fig. 1). Initially establishing research capabilities on the Moon and Mars, these autonomous systems could methodically expand operations throughout our Solar System and potentially beyond. Their operational advantages—functioning in harsh environments, making independent decisions despite communication delays, and conducting continuous experiments—position them as valuable assets for space-based research. Future development might enable robotic research networks operating across multiple celestial bodies, facilitating detailed study of distant astronomical phenomena. While formidable challenges exist in developing systems with sufficient radiation protection, long-term operational reliability, and effective distant communication protocols, Robot Scientists represent a promising approach to expanding humanity's scientific reach. With continued technological advancement, these systems could substantially enhance our understanding of the cosmos through methodical exploration and discovery beyond Earth.

## **Impact Statement**

This paper establishes a structured classification framework for AI-driven autonomous scientific systems. The proposed taxonomy facilitates precise communication between scientific researchers, technological innovators, and regulatory authorities. Through detailed categorization of autonomy levels in scientific investigation, this framework offers methodological guidance for multidisciplinary tool creation, enhances collaboration across scientific domains, and addresses critical ethical implications in autonomous research deployment. System development must adhere to established ethical guidelines, like the 23 Asilomar AI Principles, while maintaining compliance with applicable regulatory structures and international standards for advanced AI applications. This disciplined approach to classification supports the responsible evolution of autonomous scientific platforms that enhance research capabilities while incorporating appropriate safeguards against potential complications.

## **Conclusion**

The autonomous generalist scientist presents a groundbreaking framework harnessing the comprehensive interdisciplinary knowledge within foundation models, while synergizing AI agent capabilities with embodied robotic systems to automate scientific inquiry across digital and physical domains. This unified approach facilitates not only swift information processing, hypothesis formulation, and automated virtual experimentation but also enables sophisticated physical research implementations, effectively connecting computational simulations with tangible laboratory applications. Crucially, this methodology operates beyond conventional scientific boundaries, facilitating investigations throughout established disciplines while potentially catalyzing entirely novel research directions. The systematic reproducibility of computational and robotic platforms potentially establishes new scaling laws for knowledge discovery, substantially enhancing research productivity compared to traditional human-centered approaches. As development of this integrated paradigm progresses, the collaborative fusion between artificial intelligence and robotics will fundamentally reconfigure academic investigation practices, fostering innovations with far-reaching positive societal implications.

## **Bibliography Notes**

## **References**

## **Bibliography Notes**

- [1] Ahn M, Dwibedi D, Finn C, et al (2024) Autort: Embodied foundation models for large scale orchestration of robotic agents. arXiv preprint arXiv:240112963
- [2] Ajoudani A, Zanchettin AM, Ivaldi S, et al (2018) Progress and prospects of the human–robot collaboration. Autonomous robots 42:957–975
- [3] Aljalbout E, Sotirakis N, van der Smagt P, et al (2024) Limt: Language-informed multi-task visual world models. arXiv preprint arXiv:240713466
- [4] Angelopoulos A, Cahoon JF, Alterovitz R (2024) Transforming science labs into automated factories of discovery. Science Robotics 9(95):eadm6991

- [5] Baek J, Jauhar SK, Cucerzan S, et al (2024) Researchagent: Iterative research idea generation over scientific literature with large language models. arXiv preprint arXiv:240407738
- [6] Bank W (2024) Researchers in research and development (per million people). URL <https://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6?end=2022&start=1996&view=chart&year=2009>
- [7] Bank W (2024) Scientific and technical journal articles. URL <https://data.worldbank.org/indicator/IP.JRN.ARTC.SC?view=chart>
- [8] Bharadhwaj H, Vakil J, Sharma M, et al (2024) Roboagent: Generalization and efficiency in robot manipulation via semantic augmentations and action chunking. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp 4788–4795
- [9] Bom HSH (2023) Exploring the opportunities and challenges of chatgpt in academic writing: a roundtable discussion. Nuclear medicine and molecular imaging 57(4):165–167
- [10] Bommasani R, Hudson DA, Adeli E, et al (2021) On the opportunities and risks of foundation models. arXiv preprint arXiv:210807258
- [11] Brohan A, Brown N, Carbajal J, et al (2023) Rt-2: Vision-language-action models transfer web knowledge to robotic control. arXiv preprint arXiv:230715818
- [12] Daniel KL, McConnell M, Schuchardt A, et al (2022) Challenges facing interdisciplinary researchers: Findings from a professional development workshop. Plos one 17(4):e0267234
- [13] Darvish K, Skreta M, Zhao Y, et al (2024) Organa: A robotic assistant for automated chemistry experimentation and characterization. arXiv preprint arXiv:240106949
- [14] Dergaa I, Chamari K, Zmijewski P, et al (2023) From human writing to artificial intelligence generated text: examining the prospects and potential threats of chatgpt in academic writing. Biology of sport 40(2):615–622
- [15] Dong Y (2024) The multi-agent system based on llm for online discussions. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, pp 2731–2733
- [16] Elendu C, Amaechi DC, Elendu TC, et al (2023) Ethical implications of ai and robotics in healthcare: A review. Medicine 102(50):e36671
- [17] Hong S, Lin Y, Liu B, et al (2024) Data interpreter: An llm agent for data science. arXiv preprint arXiv:240218679
- [18] Ifargan T, Hafner L, Kern M, et al (2024) Autonomous llm-driven research from data to human-verifiable research papers. arXiv preprint arXiv:240417605
- [19] Jablonka KM, Ai Q, Al-Feghali A, et al (2023) 14 examples of how llms can transform materials science and chemistry: a reflection on a large language model hackathon. Digital Discovery 2(5):1233–1250
- [20] Ji Z, Yu T, Xu Y, et al (2023) Towards mitigating llm hallucination via self reflection. In: Findings of the Association for Computational Linguistics: EMNLP 2023, pp 1827–1843

- [21] Ju Y, Hu K, Zhang G, et al (2024) Robo-abc: Affordance generalization beyond categories via semantic correspondence for robot manipulation. In: European Conference on Computer Vision, Springer, pp 222–239
- [22] Koh JY, Lo R, Jang L, et al (2024) Visualwebarena: Evaluating multimodal agents on realistic visual web tasks. arXiv preprint arXiv:240113649
- [23] Li S, Huang Y, Guo C, et al (2024) Chemistry3d: Robotic interaction benchmark for chemistry experiments. arXiv preprint arXiv:240608160
- [24] Lingard L (2023) Writing with chatgpt: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on medical education* 12(1):261
- [25] Liu H, Azam M, Bin Naeem S, et al (2023) An overview of the capabilities of chatgpt for medical writing and its implications for academic integrity. *Health Information & Libraries Journal* 40(4):440–446
- [26] M. Bran A, Cox S, Schilter O, et al (2024) Augmenting large language models with chemistry tools. *Nature Machine Intelligence* 6(5):525–535
- [27] Ma P, Wang TH, Guo M, et al (2024) Llm and simulation as bilevel optimizers: A new paradigm to advance physical scientific discovery. arXiv preprint arXiv:240509783
- [28] Ma Y, Song Z, Zhuang Y, et al (2024) A survey on vision-language-action models for embodied ai. arXiv preprint arXiv:240514093
- [29] Mower CE, Wan Y, Yu H, et al (2024) Ros-llm: A ros framework for embodied ai with task feedback and structured reasoning. arXiv preprint arXiv:240619741
- [30] Padalkar A, Pooley A, Jain A, et al (2023) Open x-embodiment: Robotic learning datasets and rt-x models. arXiv preprint arXiv:231008864
- [31] Pan A, Musheyev D, Bockelman D, et al (2023) Assessment of artificial intelligence chatbot responses to top searched queries about cancer. *JAMA oncology* 9(10):1437–1440
- [32] Ren AZ, Dixit A, Bodrova A, et al (2023) Robots that ask for help: Uncertainty alignment for large language model planners. arXiv preprint arXiv:230701928
- [33] Rossoni AL, de Vasconcellos EPG, de Castilho Rossoni RL (2023) Barriers and facilitators of university-industry collaboration for research, development and innovation: a systematic review. *Management Review Quarterly* pp 1–37
- [34] Si C, Yang D, Hashimoto T (2024) Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers. arXiv
- [35] Valluri DD, et al (2024) Exploring cognitive reflection for decision-making in robots: Insights and implications. *International Journal of Science and Research Archive* 11(2):518–530
- [36] Vázquez PP (2024) Are llms ready for visualization? In: 2024 IEEE 17th Pacific Visualization Conference (PacificVis), IEEE, pp 343–352
- [37] Wagner G, Lukyanenko R, Paré G (2022) Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology* 37(2):209–226

- [38] Wang L, Ma C, Feng X, et al (2024) A survey on large language model based autonomous agents. *Frontiers of Computer Science* 18(6):186345
- [39] Wang X, Chen J, Li N, et al (2024) Surveyagent: A conversational system for personalized and efficient research survey. arXiv preprint arXiv:240406364
- [40] Wei J, Wang X, Schuurmans D, et al (2022) Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35:24824–24837
- [41] Wu P, Escontrela A, Hafner D, et al (2023) Daydreamer: World models for physical robot learning. In: *Conference on robot learning*, PMLR, pp 2226–2240
- [42] Wu X, Xian R, Guan T, et al (2024) On the safety concerns of deploying llms/vlms in robotics: Highlighting the risks and vulnerabilities. arXiv preprint arXiv:240210340
- [43] Wu Z, Han C, Ding Z, et al (2024) Os-copilot: Towards generalist computer agents with self-improvement. arXiv preprint arXiv:240207456
- [44] Xia Z, Deng Z, Fang B, et al (2022) A review on sensory perception for dexterous robotic manipulation. *International Journal of Advanced Robotic Systems* 19(2):17298806221095974
- [45] Xiao C (2023) Autosurveygpt: Gpt-enhanced automated literature discovery. In: *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp 1–3
- [46] Xie T, Zhang D, Chen J, et al (2024) Osworld: Benchmarking multimodal agents for open-ended tasks in real computer environments. arXiv preprint arXiv:240407972
- [47] Yang Z, Zhou Z, Wang S, et al (2024) Matplotagent: Method and evaluation for llm-based agentic scientific data visualization. arXiv preprint arXiv:240211453
- [48] Yoshikawa N, Skreta M, Darvish K, et al (2023) Large language models for chemistry robotics. *Autonomous Robots* 47(8):1057–1086
- [49] Yu J, Ding Z, Tan J, et al (2024) Automated peer reviewing in paper sea: Standardization, evaluation, and analysis. arXiv preprint arXiv:240712857
- [50] Zhang B, Zhang Y, Ji J, et al (2025) Safevla: Towards safety alignment of vision-language-action model via safe reinforcement learning. arXiv preprint arXiv:250303480
- [51] Zhang H, Solak G, Lahr GJG, et al (2024) Srl-vic: A variable stiffness-based safe reinforcement learning for contact-rich robotic tasks. *IEEE Robotics and Automation Letters* 9(6):5631–5638.  
<https://doi.org/10.1109/LRA.2024.3396368>
- [52] Zhang H, Solak G, Hjorth S, et al (2025) Towards passive safe reinforcement learning: A comparative study on contact-rich robotic manipulation. arXiv preprint arXiv:250300287
- [53] Zheng B, Gou B, Kil J, et al (2024) Gpt-4v (ision) is a generalist web agent, if grounded. arXiv preprint arXiv:240101614

**Acknowledgments** We express our appreciation to Xiaoshuang Wang, Yinjun Jia, and Zhenzhong Lan for their insightful discussions and contributions that enriched this work. Special acknowledgment to Yajuan Shi for her constructive feedback and technical assistance with graphical enhancements.

**Competing interests** The authors declare no competing interests.

**Author contributions** All authors contributed to the conceptual development, textual composition, research formulation, supplied evaluative assessment, discussions and comments.