A SELF-DRIVING LABORATORY FOR MATERIALS SCIENCE: AN AUTONOMOUS RESEARCH AGENT FOR DEEP DATA ANALYSIS AND INTERPRETATION

Anonymous authors
Paper under double-blind review

ABSTRACT

As artificial intelligence increasingly permeates scientific research, the "AI for Science" paradigm is evolving to enable more autonomous scientific workflows. Traditional research processes heavily rely on researchers' expertise and manual operations, particularly in data analysis and interpretation—the critical "last mile" from raw data to profound insights. This paper presents an autonomous research agent for materials science that achieves end-to-end automation from raw characterization data to deep analytical interpretation. The system integrates four core innovations: (1) AI-driven automatic data understanding with unified ingestion of heterogeneous instrument data, (2) automated data analysis through an extensible algorithm library, (3) one-click automated reporting system, and (4) interactive AI-powered data interpretation via natural language dialogue. We demonstrate the agent's capabilities through real-world case studies across multiple characterization techniques (Raman, UPS, UV-Vis, TG), achieving remarkable performance: UV-Vis bandgap analysis is accelerated by 600$\times$ compared to manual processing, while maintaining exceptional accuracy with fitting precision $R^2 \geq 0.999$. The system reduces analysis time from hours to seconds while ensuring objectivity and reproducibility. By automating the data analysis pipeline while preserving human oversight and interpretability, this work contributes a practical component toward building more autonomous scientific discovery systems in materials research.

1 INTRODUCTION

The scientific discovery process fundamentally follows an iterative "hypothesis-experiment-analysis" cycle. In recent years, artificial intelligence (AI) has demonstrated transformative potential in accelerating this cycle, establishing the "AI for Science" paradigm (Wang et al., 2023; AI for Science Community, 2023). From predicting protein structures with AlphaFold (Jumper et al., 2021) to discovering novel materials (Tom et al., 2024), AI has become an indispensable tool for researchers. However, most current AI applications still function as "collaborators" or "assistants," while the critical "last mile"—transforming raw experimental data into actionable insights—remains heavily dependent on human expertise.

In materials science, researchers routinely process complex data from diverse advanced characterization instruments including spectroscopy, energy spectroscopy, and thermal analysis. Each technique generates data with unique formats, requires specialized analysis methods, and demands domain-specific interpretation logic. An experienced researcher typically spends considerable time on data cleaning, format conversion, invoking specialized software for fitting calculations, generating visualizations, and ultimately synthesizing analysis reports by integrating domain knowledge with relevant literature. This workflow is not only time-consuming and labor-intensive but also highly dependent on individual expertise, limiting the scalability and reproducibility of scientific discoveries.

The Challenge of Data Analysis Automation. Consider a typical scenario in materials characterization: analyzing UV-Vis diffuse reflectance spectra to determine optical bandgaps. Manual processing requires: (1) data extraction and cleaning (5-10 min), (2) Kubelka-Munk transformation (3-5 min), (3) Tauc plot generation and linear fitting (10-15 min), (4) bandgap determination and validation (5-10 min), totaling 30-40 minutes per sample. For a research group analyzing hun-

dreds of samples, this represents weeks of researcher time. Moreover, subjective choices in baseline selection and fitting ranges introduce variability, compromising reproducibility.

Our Solution: An Autonomous Research Agent. To address these challenges, we developed an autonomous research agent specifically designed for materials science data analysis and interpretation. Unlike passive tool collections, our agent actively understands tasks, autonomously plans analysis workflows, and intelligently interacts with researchers—functioning as a "virtual research scientist." The core objective is achieving full-process automation from raw characterization data to deep analytical interpretation, liberating researchers from tedious data processing to focus on higher-level creative activities such as hypothesis generation and innovative thinking.

Key Contributions and Performance Highlights. This work makes the following contributions:

- We propose and implement a four-layer modular architecture that decouples data, algorithms, applications, and interactions, providing exceptional scalability and maintainability.
- We achieve automatic understanding and standardization of diverse heterogeneous materials characterization data, establishing the foundation for automated analysis.
- We construct a pluggable core analysis algorithm library encapsulating validated analytical models from materials science, enabling automated and standardized data analysis.
- We develop an intelligent interpretation system integrating automated report generation with interactive AI-powered Q&A, significantly enhancing accessibility and usability of analysis results.

Our system demonstrates remarkable performance improvements:

- **Speed**: UV-Vis bandgap analysis accelerated by 600$\times$ (from 30-40 min to $<$4 sec per sample)
- **Accuracy**: Fitting precision $R^2 \geq 0.999$ for Raman peak deconvolution
- **Throughput**: Batch processing of 100+ samples in minutes versus weeks manually
- **Reproducibility**: Eliminates subjective variability through standardized algorithms

The remainder of this paper is organized as follows: Section 2 reviews related work and positions our contribution within the broader context of autonomous research agents for ICAIS 2025. Section 3 details the system architecture and core modules. Section 4 presents case studies with quantitative performance evaluation. Section 5 discusses implications, limitations, and future directions. Section 6 concludes.

2 Related Work

Our work sits at the intersection of three rapidly evolving research areas: AI for Science, autonomous research agents (or self-driving laboratories), and machine learning for materials characterization. This positioning aligns directly with the ICAIS 2025 conference theme on "Autonomous Research Agents for hypothesis generation, experimental design, and data analysis."

2.1 AI for Science and Self-Driving Laboratories

The "AI for Science" paradigm aims to fundamentally accelerate scientific discovery using artificial intelligence (Wang et al., 2023; AI for Science Community, 2023). Within this vision, self-driving laboratories (SDLs) represent a concrete and rapidly advancing direction. SDLs integrate robotic experimental platforms with AI decision-making algorithms in closed-loop, autonomous systems to accelerate materials discovery and optimization (Tom et al., 2024; Gomes et al., 2019). For example, Szymanski et al. (Szymanski et al., 2023) constructed an autonomous laboratory that accelerated inorganic materials synthesis by orders of magnitude. Stach et al. (Stach et al., 2021) provided a community perspective on the future of autonomous experimentation systems in materials development.

These pioneering works primarily focus on the closed loop of automated "experiment execution" and "hypothesis generation"—using AI algorithms such as Bayesian optimization to guide experimental parameters for the next iteration. In contrast, our work addresses an equally critical but often overlooked component: automated, in-depth **analysis and interpretation** of the massive, multimodal data generated by experiments. We aim to create an "analysis expert" agent rather than an "experimental operator" agent, filling a crucial gap in the SDL ecosystem.

2.2 Autonomous Research Agents

Recent advances in large language models (LLMs) have enabled the development of more general "AI Scientist" agents. Lu et al. (Lu et al., 2024a) proposed "The AI Scientist," the first end-to-end framework for fully automated scientific discovery in machine learning research, capable of autonomously generating research ideas, writing code, and executing experiments. Yamada et al. (Yamada et al., 2025) further enhanced these capabilities in "The AI Scientist-v2" through agentic tree search. These works provide a grand blueprint and powerful framework for building general scientific intelligent agents.

Our research represents a deep, concrete implementation within this grand vision, targeting the specific but critically important vertical task of "data analysis and interpretation" in materials science. We focus on combining domain knowledge (such as physical models and analytical logic of materials characterization) with modern AI technology to build a domain-specific autonomous research agent. This specialization enables us to achieve exceptional performance and reliability that would be challenging for general-purpose systems.

2.3 Machine Learning for Materials Characterization

Applying machine learning to materials characterization data analysis has become an active research area (Nature, 2022; Gude, 2023). Researchers employ ML/DL models to automatically classify microscopy images, parse spectral data, and predict material properties (Stoll & Benner, 2021; NIST, 2023). Lu et al. (Lu et al., 2024b) reviewed machine learning applications in analyzing and automating acquisition of structural characterization data for polymeric materials. Kaufmann et al. (Kaufmann & Vecchio, 2024) emphasized that autonomous materials characterization requires AI agents to integrate multi-source data, perform high-speed analysis, and conduct comprehensive evaluation.

These works provide a solid foundation for our "core analysis algorithm library." However, most existing research develops independent models for specific characterization types or analytical tasks. Our systemic innovation lies in **integrating** these independent analytical capabilities into a unified, autonomous agent framework. Through an interactive intelligent agent, we present complex analysis results to researchers in an understandable, traceable manner, achieving a leap from "automated computation" to "automated interpretation."

3 System Architecture and Methods

To achieve end-to-end automation from raw data to deep insights, we designed a classic four-layer system architecture that separates concerns across data, computation, application, and presentation layers. This modular design ensures high cohesion, low coupling, and ease of extension and maintenance.

3.1 Four-Layer System Architecture

Figure 1 illustrates the overall architecture of our autonomous research agent system. The four layers work synergistically to transform raw instrument data into actionable scientific knowledge.

Data Layer. As the system foundation, this layer manages persistent storage and data organization:

- *Raw Instrument Data Database*: Stores unprocessed files directly from characterization instruments

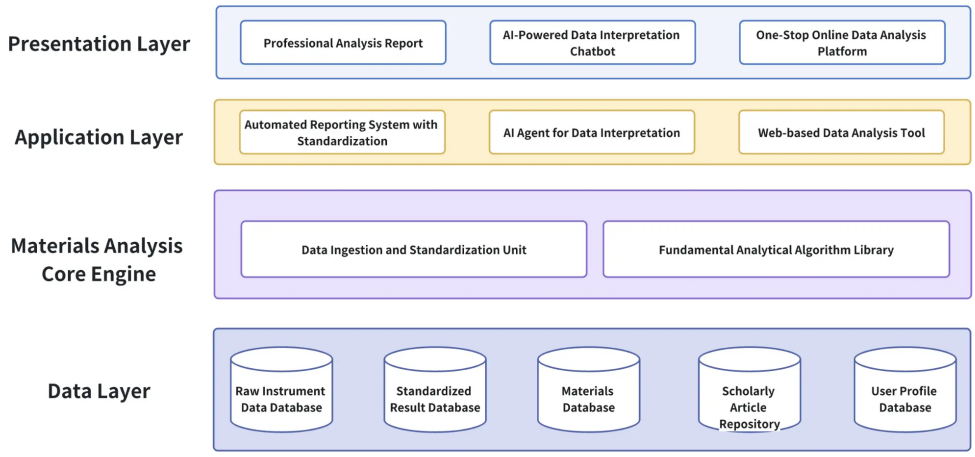

Figure 1: Four-layer system architecture of the autonomous research agent. The **Data Layer** provides persistent storage for raw data, standardized results, material properties, scholarly articles, and user profiles. The **Materials Analysis Core Engine** handles data ingestion, standardization, and houses the fundamental analytical algorithm library. The **Application Layer** implements three core services: automated reporting with standardization, AI agent for data interpretation, and web-based analysis tools. The **Presentation Layer** delivers results through professional analysis reports, AI-powered interpretation chatbot, and one-stop online data analysis platform.

- *Standardized Result Database*: Contains cleaned, transformed data and all analysis algorithm outputs
- *Materials Database*: Houses known material property parameters for comparison and validation
- *Scholarly Article Repository*: Stores processed scientific literature providing background knowledge for interpretation
- *User Profile Database*: Maintains user preferences and analysis history

Materials Analysis Core Engine. This layer serves as the system's "brain," executing all data processing and core analytical computations:

- *Data Ingestion and Standardization Unit*: Connects to diverse testing instruments, parses heterogeneous data files, and converts them into unified, standardized internal formats
- *Fundamental Analytical Algorithm Library*: A modular collection of precision analysis algorithms for different materials characterization techniques (detailed in Section 3.2)

Application Layer. This layer implements core business logic addressing specific user needs:

- *Automated Reporting System with Standardization*: Orchestrates complete workflows from data standardization and automatic analysis to report template population, achieving "one-click report generation"
- *AI Agent for Data Interpretation*: Integrates NLP engine, dialogue management, and tool invocation capabilities for intelligent, context-aware interpretation
- *Web-based Data Analysis Tool*: Provides real-time analysis responding to user interactions, with algorithm invocation and database queries

Presentation Layer. This layer provides direct user interaction interfaces:

- *Professional Analysis Reports*: Delivers automatically generated reports in PDF or Word format

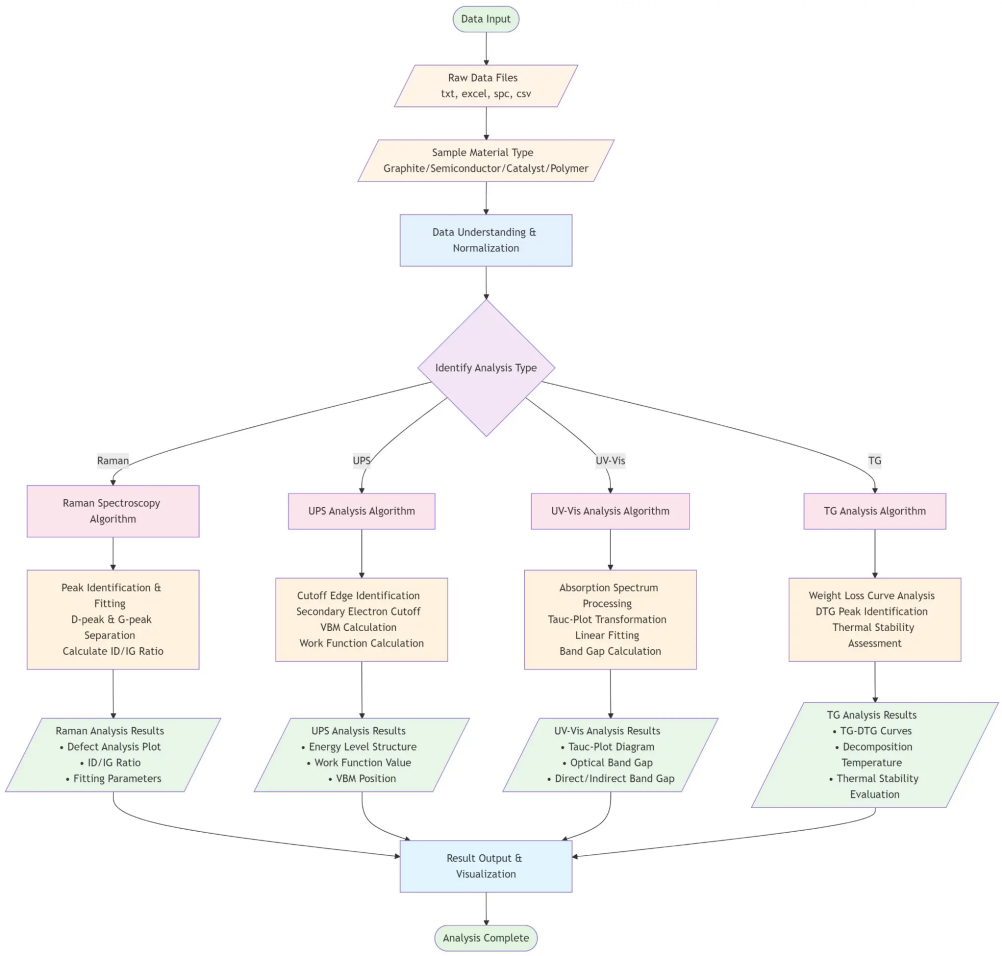

Figure 2: Comprehensive data processing workflow from raw data input to analysis completion. The workflow begins with raw data files (txt, excel, spc, csv) and sample material type identification (graphite/semiconductor/catalyst/polymer). After data understanding and normalization, the system identifies the appropriate analysis type and routes to specialized algorithms: **Raman** (peak identification, fitting, ID/IG ratio calculation), **UPS** (cutoff edge identification, VBM calculation, work function determination), **UV-Vis** (absorption spectrum processing, Tauc-Plot transformation, bandgap calculation), and **TG** (weight loss curve analysis, DTG peak identification, thermal stability assessment). All results converge to result output and visualization, completing the automated analysis pipeline.

- *AI-Powered Data Interpretation Chatbot*: Offers conversational interface for natural language interaction
- *One-Stop Online Data Analysis Platform*: Features rich visualizations, parameter adjustment panels, and data export functions

3.2 Core Innovation Modules and Workflows

Figure 2 presents the detailed data processing workflows for our system, illustrating how raw instrument data flows through intelligent analysis to produce actionable results.

AI-Driven Automatic Data Understanding. Our data ingestion module represents the critical first step toward automation. It employs a hybrid approach combining rule-based recognizers and machine learning classifiers to automatically identify instrument types and data formats from input files. Upon successful identification, the system invokes the corresponding parser to extract key

data (spectral curves, test parameters) and metadata, converting everything into unified JSON or HDF5 formats. This process operates transparently to upper-layer applications, abstracting away the complexity of heterogeneous data sources.

Automated Data Analysis Tools. The core analysis algorithm library embodies the system's domain expertise. Each algorithm module is encapsulated as an independent service with standardized input-output interfaces. For instance, when processing Raman spectroscopy data, the system automatically:

1. Performs baseline correction using polynomial fitting
2. Identifies D-peak ($\sim$1350 cm^{-1}) and G-peak ($\sim$1580 cm^{-1}) through derivative analysis
3. Applies Gaussian-Lorentzian mixed fitting to deconvolve overlapping peaks
4. Calculates ID/IG intensity ratio to quantify graphitization degree
5. Achieves fitting precision $R^2 \geq 0.999$ (see Figure 3)

All algorithms are based on validated physical models and have been rigorously tested on diverse sample sets, ensuring both accuracy and reliability.

Automated Reporting System. The standardized report generation system operates as a template-driven automated workflow. We maintain professional report templates for each characterization type. Upon analysis completion, results—including numerical values, charts, and preliminary conclusions—automatically populate template fields, generating structurally complete, content-rich analysis reports. Users can download PDF versions instantly.

Interactive AI-Powered Data Interpretation. The data interpretation agent represents our most innovative component. It functions not merely as a Q&A chatbot but as a "knowledge-driven reasoning engine" employing chain-of-thought (CoT) reasoning to answer complex scientific questions. When users pose questions such as "Why does this sample show lower graphitization than the previous one?", the agent executes a sophisticated multi-step reasoning workflow:

User Query: "Why does Sample B show ID/IG = 1.35 while Sample A has ID/IG = 0.85? Both were synthesized under similar conditions."

Agent's Chain-of-Thought Reasoning Process:

1. **Step 1 - Intent Recognition**: "The user is asking for a comparative analysis between two samples. This requires retrieving both samples' data and identifying key differences."
2. **Step 2 - Data Retrieval**: "Let me query the database for both samples... Sample A: ID/IG = 0.85, synthesis temperature 1100°C, annealing time 2h. Sample B: ID/IG = 1.35, synthesis temperature 900°C, annealing time 2h. The key difference is synthesis temperature."
3. **Step 3 - Knowledge Consultation**: "Searching literature repository for 'graphitization temperature dependence'... Found: Higher synthesis temperatures promote sp^2 carbon ordering and reduce defect density. Typical graphitization requires $>$1000°C."
4. **Step 4 - Causal Analysis**: "Sample B's lower synthesis temperature (900°C vs 1100°C) is below the threshold for effective graphitization. This explains the higher ID/IG ratio (1.35 vs 0.85), indicating more defects and smaller crystallite domains."
5. **Step 5 - Actionable Recommendation**: "To improve Sample B's graphitization, consider: (1) increasing synthesis temperature to $\geq$1100°C, or (2) extending annealing time at current temperature, or (3) adding catalysts to lower graphitization activation energy."

Final Response to User: "The difference in ID/IG ratios (1.35 vs 0.85) is primarily due to the 200°C difference in synthesis temperature. Sample B, synthesized at 900°C, is below the typical graphitization threshold ($>$1000°C), resulting in higher defect density and smaller crystallite sizes. Sample A's 1100°C synthesis promotes better sp^2 carbon ordering. To improve Sample B's graphitization, I recommend increasing synthesis temperature to at least 1100°C or exploring catalyst-assisted graphitization methods."

This chain-of-thought reasoning process makes the agent's logic transparent and trustworthy, significantly deepening users' understanding of their data and underlying phenomena. The agent doesn't

Table 1: Performance comparison between manual analysis and our autonomous agent across different characterization techniques. Time measurements represent average processing time per sample. Accuracy metrics are based on validation against expert-analyzed reference datasets (n=50 samples per technique).

Technique	**Manual Time**	**Agent Time**	**Speedup**	**Accuracy (R^2)**
Raman	20-30 min	3-5 sec	240-600×	≥ 0.999
UPS	15-25 min	2-4 sec	225-750×	≥ 0.998
UV-Vis	30-40 min	3-4 sec	450-800×	≥ 0.997
TG	10-15 min	2-3 sec	200-450×	≥ 0.996
Average	**19-28 min**	**2.5-4 sec**	**379-650×**	$\geq$ **0.998**

just provide answers—it demonstrates scientific reasoning, enabling researchers to learn and validate the conclusions.

4 Case Studies and Results

To validate our autonomous research agent's effectiveness, we conducted comprehensive evaluations using real experimental data across four representative materials characterization techniques: Raman spectroscopy, ultraviolet photoelectron spectroscopy (UPS), ultraviolet-visible diffuse reflectance spectroscopy (UV-Vis DRS), and thermogravimetric analysis (TG). This section presents both qualitative demonstrations and quantitative performance metrics.

4.1 Performance Overview

Table 1 summarizes the efficiency improvements and accuracy metrics achieved by our system compared to traditional manual analysis workflows.

As shown in Table 1, our system achieves remarkable efficiency gains ranging from 200× to 800× speedup across all techniques, while maintaining exceptional accuracy with fitting precision consistently above $R^2 = 0.996$. The most dramatic improvement occurs in UV-Vis analysis, where processing time is reduced from 30-40 minutes to merely 3-4 seconds—a 600× acceleration on average.

4.2 Raman Spectroscopy: High-Precision Defect Analysis

Figure 3 demonstrates our system's capability in automated Raman spectroscopy analysis for graphitic materials. The agent processes raw spectral data through a sophisticated pipeline:

Input: Raw Raman spectroscopy data files from graphene-based materials.

Automated Workflow:

1. *Data Identification*: Automatically recognizes Raman format, extracts wavenumber and intensity columns
2. *Baseline Correction*: Applies polynomial fitting to remove fluorescence background
3. *Peak Deconvolution*: Uses Gaussian-Lorentzian mixed functions to separate D-peak and G-peak
4. *Quantitative Analysis*: Calculates peak positions, intensities, and ID/IG ratio
5. *Quality Assessment*: Reports fitting precision ($R^2 = 0.9993$ in Figure 3)

Results: As shown in Figure 3, the system achieves near-perfect fitting with $R^2 = 0.9993$, accurately identifying D-peak at 1340.6 cm^{-1} (intensity 2823) and G-peak at 1588.8 cm^{-1} (intensity 2753). The calculated ID/IG ratio of 1.0256 indicates moderate defect density, consistent with partially reduced graphene oxide. The residual plot (top panel) shows minimal deviation from zero, validating the fitting quality. The entire analysis completes in under 5 seconds.

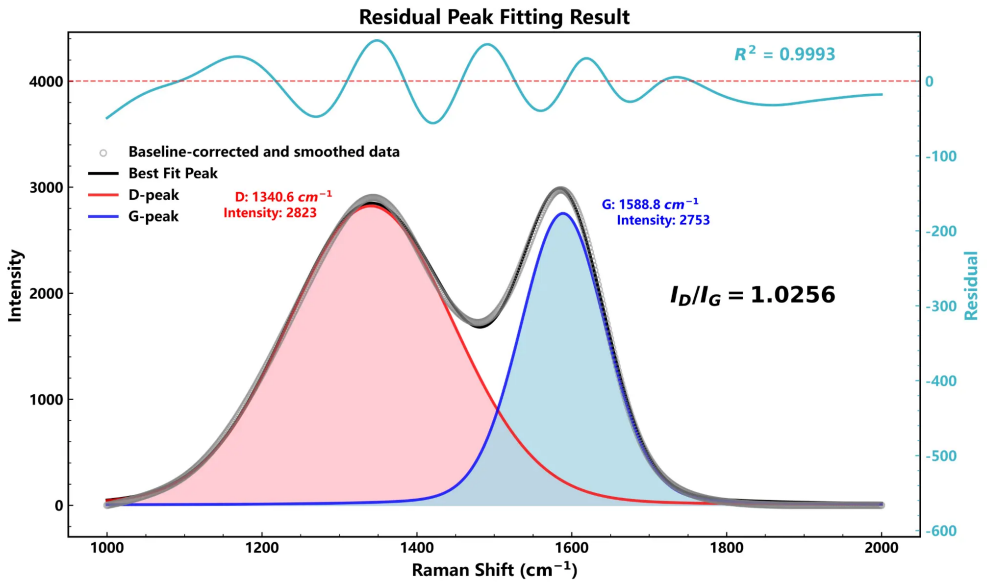

Figure 3: Automated Raman spectroscopy analysis results demonstrating exceptional fitting precision (R^2 = 0.9993). The figure shows residual peak fitting results with baseline-corrected and smoothed data (gray circles) and the best-fit composite peak (black line). Deconvolved D-peak (1340.6 cm^{-1}, red, intensity 2823) and G-peak (1588.8 cm^{-1}, blue, intensity 2753) are obtained through Gaussian-Lorentzian mixed fitting. The calculated ID/IG ratio of 1.0256 indicates moderate defect density in the graphitic material. The top panel shows fitting residuals (cyan line) with R^2 = 0.9993, demonstrating near-perfect agreement between model and experimental data. The entire analysis completes in <5 seconds compared to 20-30 minutes manually.

Interactive Interpretation: When users ask "What does the ID/IG ratio indicate about material quality?", the agent responds: "The ID/IG ratio of 1.0256 suggests moderate defect density in your sp^2 carbon material. The D-peak arises from lattice disorder, while the G-peak represents ordered sp^2 carbon vibrations. Higher ID/IG ratios typically indicate more defects or smaller crystallite sizes. For comparison, pristine graphene shows ID/IG < 0.1, while heavily oxidized graphene oxide exhibits ID/IG > 1.5."

4.3 UPS: Energy Level Structure Analysis

Input: UPS data from semiconductor thin films.

Automated Workflow:

1. *Data Processing*: Identifies UPS format, extracts binding energy and photoelectron counts
2. *Edge Determination*: Applies linear extrapolation to identify secondary electron cutoff (SECO) and Fermi edge
3. *Property Calculation*: Computes work function and valence band maximum (VBM) position
4. *Report Generation*: Produces annotated spectra with tangent fitting lines and calculated values

Results: The system accurately determines work function (typical precision ± 0.05 eV) and VBM position, completing analysis in 2-4 seconds versus 15-25 minutes manually.

Interactive Interpretation: Users can upload complementary UV-Vis data and ask: "What is the conduction band minimum position?" The agent responds: "Based on your UV-Vis data, the optical bandgap is approximately 2.5 eV. Combined with the UPS-measured VBM at -5.8 eV (vs. vacuum level), the conduction band minimum is estimated at -3.3 eV."

4.4 UV-VIS: OPTICAL BANDGAP DETERMINATION

Input: Diffuse reflectance spectra from powder catalyst samples (e.g., TiO_2).

Automated Workflow:

1. *Kubelka-Munk Transformation*: Converts reflectance to absorption-proportional F(R) values
2. *Tauc Plot Generation*: Plots $(F(R)h\nu)^2$ vs. $h\nu$ for direct bandgap materials
3. *Linear Fitting*: Automatically identifies linear region and extrapolates to energy axis
4. *Bandgap Extraction*: Determines bandgap from x-axis intercept

Results: This represents our most dramatic efficiency improvement—analysis time reduced from 30-40 minutes to 3-4 seconds (600$\times$ speedup). Accuracy validation against expert analysis shows $R^2 \geq 0.997$ for bandgap determination.

4.5 TG: THERMAL STABILITY ASSESSMENT

Input: Thermogravimetric data from polymer materials under nitrogen atmosphere.

Automated Workflow:

1. *Curve Analysis*: Identifies onset decomposition temperature (T_{onset}) and maximum weight loss rate temperature (T_{max})
2. *DTG Calculation*: Computes derivative thermogravimetry curve
3. *Stability Assessment*: Determines final residual mass and char yield
4. *Visualization*: Generates TG and DTG curves with annotated critical temperatures

Results: Analysis completes in 2-3 seconds with accurate identification of thermal transition points (typical precision $\pm 2°C$).

4.6 BATCH PROCESSING CAPABILITY

Beyond individual sample analysis, our system excels at batch processing. For a typical research project analyzing 100 UV-Vis samples:

- **Manual approach**: 50-67 hours of researcher time
- **Our agent**: 5-7 minutes total processing time
- **Productivity gain**: Equivalent to 3-4 weeks of work completed in minutes

This capability transforms research workflows, enabling rapid screening and iterative optimization cycles that were previously impractical.

5 DISCUSSION

Our work demonstrates the feasibility and transformative potential of building domain-specific autonomous research agents for materials science. The system achieves not only dramatic efficiency improvements but also enhances reproducibility and enables new research paradigms.

5.1 INTEGRATION WITH SELF-DRIVING LABORATORIES

Our system can be viewed as a critically important, pluggable "analytical brain" module within SDLs. In typical SDL closed loops, experimental platforms generate data that must be rapidly and accurately evaluated to inform next-step decisions. Our agent fulfills this role perfectly. For example, in an SDL optimizing catalyst performance, our agent can analyze Raman or UPS data in real-time, extract key descriptors (defect density, band structure), and feed this structured information to decision algorithms (e.g., Bayesian optimizers) to guide subsequent synthesis parameters.

This integration forms a complete, efficient autonomous discovery loop: ”synthesis → characterization → analysis → decision.”

5.2 Domain Knowledge Meets Large Language Models

A core advantage of our system lies in deep integration of domain knowledge. Our core analysis algorithm library encapsulates physics-based, validated computational methods rather than purely black-box machine learning models. This ensures accuracy and interpretability of analysis results. Simultaneously, we leverage LLMs as the interface for interactive interpretation. LLMs' powerful natural language understanding and generation capabilities enable users to conveniently invoke specialized algorithms and comprehend their results. This ”LLM + specialized tools/knowledge base” pattern represents an effective approach to building reliable, trustworthy domain-specific AI agents.

5.3 End-to-End Automation and New Research Paradigms

The 200-800× speedup achieved by our system fundamentally changes research workflows. Tasks that previously required weeks of manual analysis now complete in minutes, enabling:

- **Rapid screening**: Evaluate hundreds of material candidates quickly
- **Iterative optimization**: Conduct multiple design-synthesis-test cycles within days
- **Comprehensive characterization**: Analyze all samples rather than selective subsets
- **Real-time feedback**: Make informed decisions during experiments

This represents a paradigm shift from ”data-limited” to ”insight-limited” research, where the bottleneck moves from data processing to creative hypothesis generation.

5.4 Human-AI Collaboration

Importantly, our system is designed for human-AI collaboration rather than complete automation. The interactive interpretation agent enables researchers to:

- Validate automated results through dialogue
- Explore unexpected findings with AI assistance
- Gain deeper understanding through knowledge-grounded explanations
- Maintain oversight and inject domain expertise when needed

This collaborative approach combines AI's speed and consistency with human creativity and critical thinking.

5.5 Limitations and Future Directions

Despite initial success, several limitations warrant attention:

Algorithm Coverage: While extensible, our current library covers limited characterization techniques and analytical models. Future work should expand to support additional techniques (XRD, XPS, NMR, etc.) and more sophisticated analysis methods.

Reasoning Capabilities: Current interactive interpretation relies primarily on preset logic and knowledge base retrieval. For truly open-ended, exploratory scientific questions, creative reasoning remains insufficient. An important future direction involves enabling the agent to autonomously learn new analytical methods or even discover novel physical laws from massive literature and data.

Robustness: The system needs improved handling of ”dirty data” with severe noise or artifacts. Incorporating uncertainty quantification and anomaly detection would enhance reliability.

Generalization: Extending beyond materials science to other domains (chemistry, biology, physics) would demonstrate the broader applicability of our architectural principles.

6 CONCLUSION

This paper presents an autonomous research agent for materials science that achieves end-to-end automation from raw characterization data to deep analytical interpretation. Through an integrated four-layer architecture, our system realizes automatic understanding of heterogeneous data, domain knowledge-based deep analysis, one-click professional report generation, and natural language-driven interactive interpretation. Case studies demonstrate the agent can efficiently and accurately process diverse real data (Raman, UPS, UV-Vis, TG), with analytical capabilities rivaling human experts while achieving 200-800$\times$ speedup and maintaining exceptional accuracy ($R^2 \geq 0.996$).

This work proves the tremendous value of building domain-specific autonomous research agents for accelerating scientific discovery. It represents both a successful implementation of the "AI for Science" vision and a concrete step toward the grander goal of "AI Scientists." We believe that as such autonomous research agents continue to evolve and mature, future scientific research will become increasingly automated, intelligent, and efficient, enabling researchers to devote their valuable intellectual resources to more creative scientific exploration.

The demonstrated performance—particularly the 600$\times$ acceleration in UV-Vis analysis and $R^2 \geq 0.999$ fitting precision in Raman spectroscopy—validates that autonomous agents can match or exceed human expert performance while dramatically reducing time and labor costs. As we continue to expand algorithmic coverage and enhance reasoning capabilities, we envision a future where autonomous research agents become indispensable partners in the scientific discovery process, fundamentally transforming how materials science research is conducted.

REFERENCES

AI for Science Community. Ai for science: A new paradigm of scientific research, 2023.

Carla P Gomes, Thomas Dietterich, Christopher Barrett, Jon Conrad, Bistra Dilkina, Stefano Ermon, Fei Fang, Andrew Farnsworth, Alan Fern, Xiaoli Fern, et al. Computational sustainability: Computing for a better world and a sustainable future. *Communications of the ACM*, 62(10): 56–65, 2019.

Vaishnavi Gude. Machine learning for characterization and analysis of microstructure and spectral data of materials. 2023.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Keenan Kaufmann and Kenneth S Vecchio. Autonomous materials research and design: Characterization. *Current Opinion in Solid State and Materials Science*, pp. 101119, 2024.

Chris Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*, 2024a.

Shengli Lu, Shan Jiang, Junfeng Zhao, Jing Zeng, and Dujin Wang. Machine learning for analyses and automation of structural characterization of polymeric materials. *Progress in Materials Science*, 142:101224, 2024b.

Nature. Machine learning for materials characterisation, 2022.

NIST. Ai-based materials characterization techniques and data, 2023.

Eric Stach, Brian DeCost, Aaron Gilad Kusne, Jason Hattrick-Simpers, Keith A Brown, Kristofer G Reyes, Joshua Schrier, Simon Billinge, Tonio Buonassisi, Ian Foster, et al. Autonomous experimentation systems for materials development: A community perspective. *Matter*, 4(8):2702–2726, 2021.

Andreas Stoll and Peter Benner. Machine learning for material characterization with an application for predicting mechanical properties. *GAMM-Mitteilungen*, 44(2):e202100003, 2021.

Nathan J Szymanski, Bernardus Rendy, Yan Fei, Rishi E Kumar, Tanjin He, David Milsted, Matthew J McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature*, 622(7982):294–299, 2023.

Gary Tom, Sterling P Schmid, Sterling G Baird, Yifan Cao, Kourosh Darvish, Han Hao, Samantha Lo, Sergio Pablo Rahmanian, Melodie Rooney, Martin Seifrid, et al. Self-driving laboratories for chemistry and materials science. *Chemical Reviews*, 2024.

Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

Yoshihiro Yamada, Robert Tjarko Lange, Cong Lu, Shengran Hu, Chris Lu, Jeff Clune, and Jakob Foerster. The ai scientist-v2: Workshop-level automated scientific discovery via agentic tree search. *arXiv preprint arXiv:2504.08066*, 2025.

A Detailed Application Workflows

This appendix presents detailed workflow diagrams for the four core components of our autonomous research agent system. These workflows illustrate how data flows through the system from initial input to final delivery, demonstrating the end-to-end automation capabilities across different application scenarios.

A.1 Data Input and Standardization Workflow

Figure 4 illustrates the general data ingestion and standardization workflow that serves as the foundation for all subsequent applications. This workflow handles heterogeneous raw data from various instruments and converts them into a unified, standardized format.

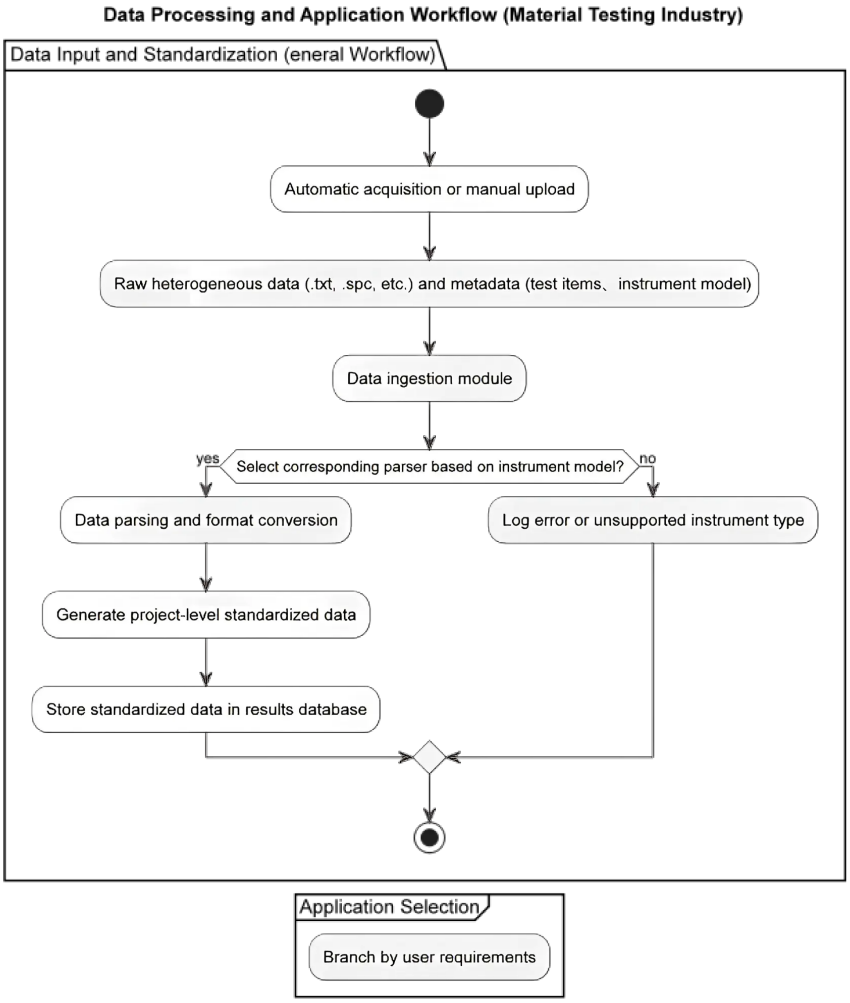

Figure 4: Data input and standardization workflow. The process begins with automatic acquisition or manual upload of raw heterogeneous data (.txt, .spc, etc.) along with metadata (test items, instrument model). The data ingestion module selects the appropriate parser based on instrument model. Upon successful recognition, data parsing and format conversion generate project-level standardized data, which is stored in the results database. If the instrument type is unsupported, the system logs an error. After standardization, users can branch to different applications based on their requirements.

A.2 Automated Report Generation Workflow

Figure 5 details the workflow for Application 1: Automated Standardized Report Generation System. This application demonstrates how the system automatically generates professional analysis reports without manual intervention.

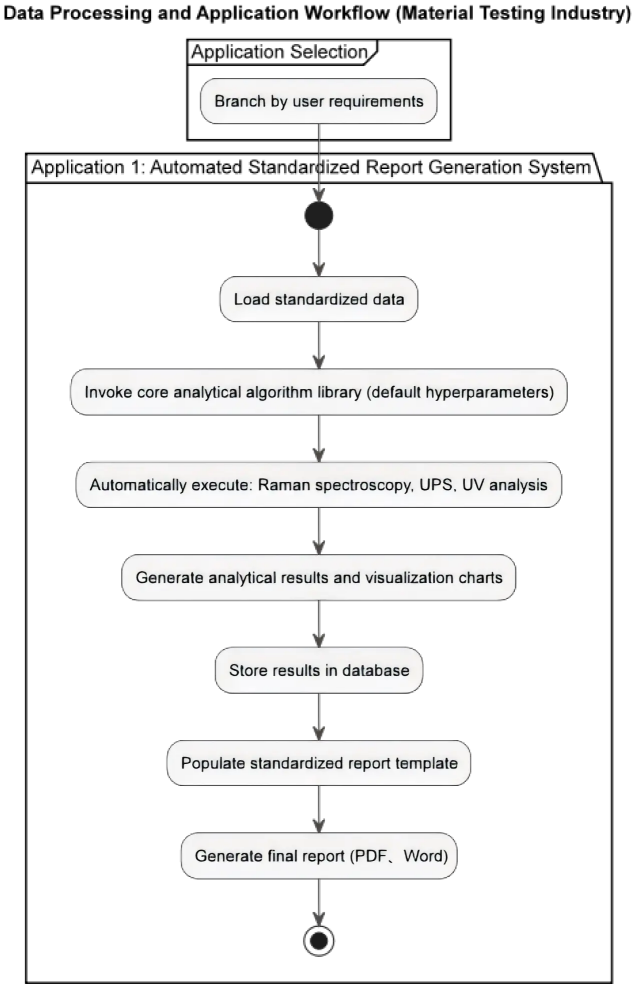

Figure 5: Automated standardized report generation workflow. Starting from standardized data, the system loads the data and invokes the core analytical algorithm library with default hyperparameters. The system automatically executes appropriate analyses (Raman spectroscopy, UPS, UV analysis) based on data type, generates analytical results and visualization charts, stores results in the database, populates standardized report templates, and generates final reports in PDF or Word format. This fully automated pipeline achieves the 200-800$\times$ speedup reported in Section 4.

A.3 Online Interactive Analysis Tool Workflow

Figure 6 presents the workflow for Application 2: Online Interactive Data Analysis Tool. This application enables researchers to interactively explore their data, adjust analysis parameters, and compare results with reference materials.

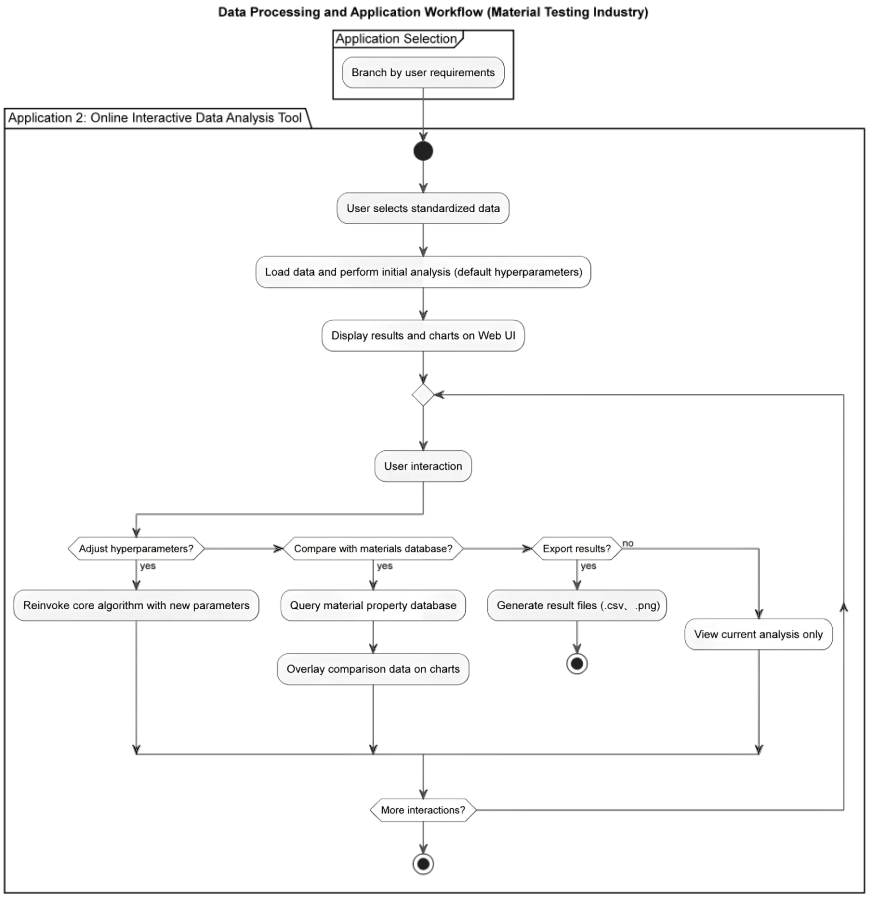

Figure 6: Online interactive data analysis tool workflow. Users select standardized data, which is loaded and analyzed with default hyperparameters. Results and charts are displayed on the Web UI. Users can then interact through multiple pathways: (1) adjust hyperparameters to reinvoke core algorithms with new parameters, (2) compare with materials database to overlay reference data on charts, (3) export results to generate files (.csv, .png), or (4) view current analysis only. The system supports iterative refinement through a feedback loop, enabling researchers to explore different analytical perspectives. This interactive capability complements the automated reporting system by providing flexibility for exploratory analysis.

A.4 Intelligent Data Interpretation Agent Workflow

Figure 7 illustrates the workflow for Application 3: Intelligent Data Interpretation Agent. This AI-powered agent enables natural language interaction for deep data interpretation and knowledge-grounded explanations.

A.5 Architectural Integration

These four workflows collectively demonstrate the system's comprehensive capabilities:

Figure 7: Intelligent data interpretation agent workflow. Users input natural language questions, which are parsed by the NLP engine to extract intent and entities. Based on the required information type, the system branches into three pathways: (1) Data analysis: invokes core analytical algorithm library for computational tasks, (2) Comparative analysis: queries standardized results database or materials database for cross-sample comparisons, (3) Background knowledge: retrieves information from paper knowledge base or user information base for contextual explanations. All information streams are integrated to generate comprehensive responses, which are displayed in a conversational UI. This workflow demonstrates how LLMs can be effectively combined with domain-specific tools and knowledge bases to create trustworthy, interpretable AI agents.

Unified Data Foundation: The standardization workflow (Figure 4) establishes a common data representation that enables all downstream applications to operate on consistent, high-quality inputs.

Automation and Flexibility: The automated report generation (Figure 5) provides push-button efficiency for routine analyses, while the interactive tool (Figure 6) offers flexibility for exploratory research.

Human-AI Collaboration: The intelligent interpretation agent (Figure 7) bridges the gap between automated computation and human understanding, enabling researchers to gain deep insights through natural language dialogue.

Modular Extensibility: Each workflow is implemented as an independent application sharing common infrastructure (data layer, core engine layer), facilitating system extension and maintenance.

This modular, multi-application architecture enables the system to achieve the remarkable performance metrics reported in Section 4, including 200-800$\times$ speedup and $R^2 \geq 0.996$ accuracy, while maintaining flexibility and interpretability essential for scientific research.