OpenQlaw: An Agentic AI Assistant for Analysis of 2D Quantum Materials

Sankalp Pandey[1], **Xuan-Bac Nguyen**[1], **Hoang-Quan Nguyen**[1,4],
Tim Faltermeier[2,4], **Nicholas Borys**[2,4], **Hugh Churchill**[3,4], **Khoa Luu**[1,4]

[1] Quantum AI Lab, University of Arkansas, USA   [2] University of Utah, USA
[3] Department of Physics, University of Arkansas, USA
[4] MonARK NSF Quantum Foundry

[1]{xnguyen, hn016, sankalpp, khoaluu}@uark.edu,
[2]{timfaltermeier@montana.edu,nicholas.borys@utah.edu}, [3]hchurch@uark.edu
https://uark-cviu.github.io/projects/OpenQlaw

Abstract

The transition from optical identification of 2D quantum materials to practical device fabrication requires dynamic reasoning beyond the detection accuracy. While recent domain-specific Multimodal Large Language Models (MLLMs) successfully ground visual features using physics-informed reasoning, their outputs are optimized for step-by-step cognitive transparency. This yields verbose candidate enumerations followed by dense reasoning that, while accurate, may induce cognitive overload and lack immediate utility for real-world interaction with researchers. To address this challenge, we introduce OpenQlaw, an agentic orchestration system for analyzing 2D materials. The architecture is built upon NanoBot, a lightweight agentic framework inspired by OpenClaw, and QuPAINT, one of the first Physics-Aware Instruction Multi-modal platforms for Quantum Material Discovery. This allows accessibility to the lab floor via a variety of messaging channels. OpenQlaw allows the core Large Language Model (LLM) agent to orchestrate a domain-expert MLLM, with QuPAINT, as a specialized node, successfully decoupling visual identification from reasoning and deterministic image rendering. By parsing spatial data from the expert, the agent can dynamically process user queries, such as performing scale-aware physical computation or generating isolated visual annotations, and answer in a naturalistic manner. Crucially, the system features a persistent memory that enables the agent to save physical scale ratios (e.g., 1 pixel = 0.25 μm) for area computations and store sample preparation methods for efficacy comparison. The application of an agentic architecture, together with the extension that uses the core agent as an orchestrator for domain-specific experts, transforms isolated inferences into a context-aware assistant capable of accelerating high-throughput device fabrication.

1   Introduction

Characterization of two-dimensional (2D) quantum materials, such as graphene, MoS_2, and hBN, is foundational to next-generation electronics, quantum technologies [15, 21, 14, 20, 4, 19, 17], and emerging quantum applications [9, 7, 6, 8]. However, due to the subtle visual differences between monolayers, bilayers, and few-layer flakes, it is difficult to rely on standard optical microscopy to reliably determine layer thickness [10]. Researchers must use Atomic Force Microscopy (AFM) for accurate validation, which leads to a slow and labor–intensive manual workflow where each candidate flake must first be identified optically, physically relocated to an AFM, measured at the nanoscale, and then meticulously remapped back to the original optical coordinates [13, 5]. This serves as a major bottleneck for the throughput and scalability of device fabrication.
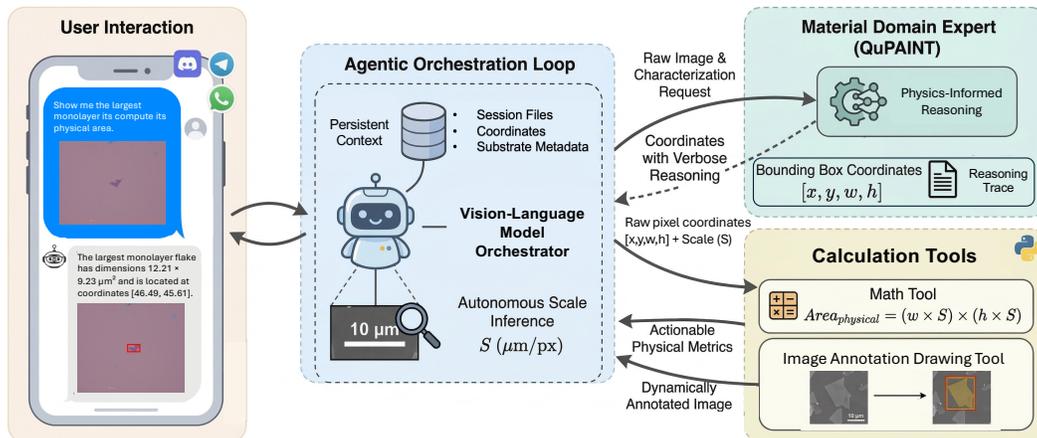
Figure 1: The proposed OpenQlaw framework.

1.1 Limitations of Object Detectors

To overcome this restriction, deep learning methods [12, 18, 29, 30, 16, 23], such as object detectors [32, 2, 25] and segmentation models [26, 11, 24], have been used to identify exfoliated flakes directly from optical images. While conventional computer vision approaches excel at localizing objects with strong boundaries or textures, they lack the physical priors necessary for quantum material characterization. Standard models cannot reason about the thin-film interference that dictates optical contrast and can easily confuse physical phases (e.g. monolayer vs. bilayer) when subjected to minor variations in microscope illumination, substrate thickness, or ambient conditions.

1.2 Interpretability with MLLMs

Recent advances like QuPAINT [22] address these representation failures through Physics-Aware Instruction Tuning with the use of a Physics–Informed Attention (PIA) module to fuse visual embeddings with optical priors. This allows specialized Multimodal Large Language Models (MLLMs) to explicitly reason about physical properties and deliver state-of-the-art performance for flake localization and layer characterization. However, these MLLMs are optimized for cognitive transparency to explain step–by–step how they reach their answers rather than practical execution. They output verbose candidate enumerations and dense reasoning traces which, while technically correct, however can induce cognitive overload and lack immediate utility for a researcher.

1.3 Contributions of this Work

The primary contribution of this work is the introduction of a multi-agent framework connecting a domain-expert MLLM to a conversational interface for materials discovery. OpenQlaw decouples the precise reasoning and identification of flakes in optical microscopy images from the deterministic physical calculations, targeted visual rendering, conversational interactions, and memory necessary for real-world decision making and high-throughput device fabrication.

2 The OpenQlaw Framework

We develop OpenQlaw as a modular, agent architecture to resolve the interpretability and execution bottlenecks of multimodal large language models. By building on top of an open-source conversational agent framework, OpenQlaw functions as a central reasoning engine that delegates specialized tasks to domain experts. This allows the system to separate the natural language understanding from complex physical inference and deterministic computation. The architecture is comprised of three primary components: the Agentic Orchestration Loop, the Material Domain Expert, and the Deterministic Execution tools. We visualize the overall architecture in Figure 1.

2.1 Agentic Orchestration Loop

The core of OpenQlaw is a vision-language model, acting as the central orchestrator agent. The agent is capable of managing user interactions through chat interfaces (e.g. Discord or WhatsApp), leveraging the native media attachment capabilities of these platforms for seamless microscopy image uploads. A critical feature of this loop is the access to persistent context, including memory or skills. The orchestrator uses session files as short–term context, maintaining access throughout a conversation to spatial coordinate data after initial inference on an image. Furthermore, the architecture automatically stores important information within its memory, so if a user provides substrate metadata and the specific microscope objective scale, this context can be maintained across a set of samples. Crucially, since the agent itself possesses vision capabilities, it can autonomously infer the appropriate measurement units by analyzing physical scale bars directly from the uploaded images. OpenQlaw allows researchers to continue issuing different natural language constraints, such as isolating a specific flake and calculating its physical area, without requiring redundant image inferences from the Domain Expert.

2.2 Material Domain Expert (QuPAINT)

In standard vision–language pipelines, the MLLM delivers the final output directly to the user. For specialized models such as QuPAINT, the model presents a technically correct but highly verbose enumeration and reasoning trace for its answers. Within the OpenQlaw framework, QuPAINT operates strictly as a specialized Material Domain Expert. When a user query is related to materials, the agent uses its tool–calling capabilities to prompt the Material Domain Expert with the uploaded image and an appropriate characterization request. QuPAINT then performs the heavy physics–aware inference, identifying all the flakes, distinguishing monolayers, and outputting the raw bounding box coordinates ($[x, y, w, h]$) alongside its cognitive trace explaining the selection criteria. The agent intercepts this output, extracts the structured data, and responds to the user's specific query naturally and concisely.

2.3 Deterministic Execution Tools

Once the orchestrator extracts the raw localization data from the Material Domain Expert, it utilizes deterministic Python-based execution tools to translate pixel coordinates into actionable metrics. Vision models inherently operate in relative space, which is insufficient for physical device fabrication. To address this, if the scale (S, e.g. in $\mu m/px$) is provided by the user or autonomously inferred from an image scale bar by the agent, the bounding box dimensions are routed to a mathematical tool to compute the approximate physical surface area based on $Area_{physical} = (w \times S) \times (h \times S)$. Furthermore, OpenQlaw employs a dedicated image annotation drawing tool. Based on the researcher's specific language constraint (e.g., "Show me the largest monolayer"), the agent utilizes the tool to generate an updated image showing only the requested target. This helps to ensure dynamic image rendering based on what is useful for the researcher.

3 Implementation Details

The OpenQlaw framework is built on top of NanoBot [3], an open-source agentic architecture inspired by OpenClaw [28], allowing for rapid deployment and extensibility. The implementation is divided into the orchestration, the tool-calling, and the deterministic execution pipeline.

3.1 Core Engine

The central reasoning engine is powered by Qwen3-VL-32B-Instruct [31, 1], hosted locally on two NVIDIA Quadro RTX 8000 GPUs. To facilitate accessibility, the agent is integrated with a WhatsApp API bridge and also as a Discord Bot, utilizing the open–source conversational agent framework. This manner of deployment allows researchers to upload microscopy images and receive real-time inference natively on their devices without the overhead of dedicated software.

Figure 2: Baseline output from the Material Domain Expert (QuPAINT) illustrating the verbose cognitive traces and mass raw coordinate enumeration.

3.2 Tool Schema and Prompt Engineering

The Material Domain Expert (QuPAINT) is integrated as an external tool accessible via an API schema. When the agent determines that a user query requires flake characterization, it formulates a payload containing the uploaded image and the natural language query. The orchestrator utilizes a highly constrained system prompt to solve the interpretability bottleneck. This prompt forces the agent to parse QuPAINT's output, extract the structured numerical arrays (i.e., the $[x, y, w, h]$ bounding box coordinates), and suppress the verbose reasoning. This ensures that downstream tools receive clean, standardized data formats rather than raw text streams.

3.3 State Management and Execution

Session context, including conversational memory and skill management, is handled locally by the NanoBot architecture. Each user session generates a conversational history file that can be used to store spatial metadata and user-defined material constraints. This design prevents the system from forgetting the contextual scale (S) during multi-turn conversations. Finally, the execution tools are implemented as standalone Python functions. The mathematical tool calculates physical area by applying the contextual scale (S) directly to the extracted coordinate arrays. For dynamic visual rendering, the system utilizes the Python Imaging Library (Pillow). When called, this tool uses the original image buffer and the target coordinates, drawing isolated bounding boxes and labels as requested by the user constraint. The resulting modified image is returned to the orchestrator, which natively routes the media attachment back to the researcher through the messaging interface.

4 Results and Discussion

We present a qualitative case study comparing the output of the Material Domain Expert (QuPAINT) against the fully orchestrated OpenQlaw framework to evaluate the operational impact. The scenario replicates a standard laboratory workflow where a researcher uploads an optical microscopy image of exfoliated 2D material flakes and submits a natural language query, such as, "Identify the largest monolayer in this image and calculate its area." We use the WhatsApp interface for this task.
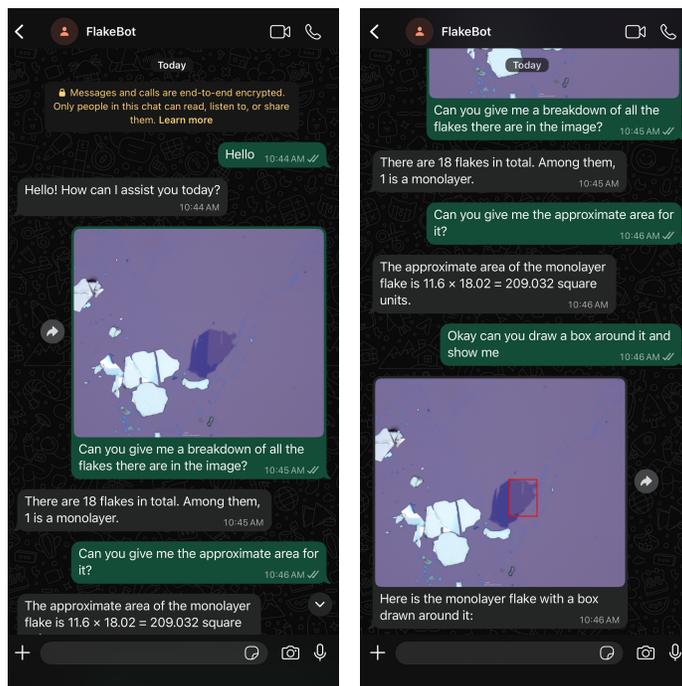
4.1 Baseline Results with QuPAINT

Figure 3: OpenQlaw framework executing a conversational workflow via WhatsApp, showcasing concise natural language responses and deterministic execution.

As shown in Figure 2, when the optical image is processed directly with the QuPAINT model, the output prioritizes cognitive transparency. For instance, when queried to count the flakes, QuPAINT identifies them but returns a dense, raw enumeration of array coordinates. Importantly, due to the inherent limitations of standard LLMs in processing dense numerical strings [27], it miscounts the final number of flakes despite identifying them correctly. Furthermore, for a simple count query, a visual interface with a standardized parsing system wastes compute resources to return an image with uninvited bounding boxes, as it lacks a mechanism to differentiate when visualization is actually required. For a researcher on the lab floor, extracting actionable fabrication metrics from this verbose text and visually crowded image requires manual parsing, creating a significant workflow bottleneck.

4.2 OpenQlaw Results

4.2.1 Delegation and Execution

Conversely, as shown in Figure 3, the OpenQlaw framework resolves this through agentic delegation and deterministic execution. When the researcher queries for a breakdown of the image, the orchestrator delegates the visual inference to QuPAINT and intercepts the dense text to respond naturally and concisely. It subsequently utilizes the mathematical execution tool to perform accurate area calculations using the cached bounding box coordinates. Additionally, because user does not ask to view the flakes in the initial prompt, the agent dynamically conserves resources by omitting the visual rendering. In subsequent requests, when the user applies specific visual constraints, the orchestrator recalls the target coordinates and highlights the appropriate flake(s) on the image.
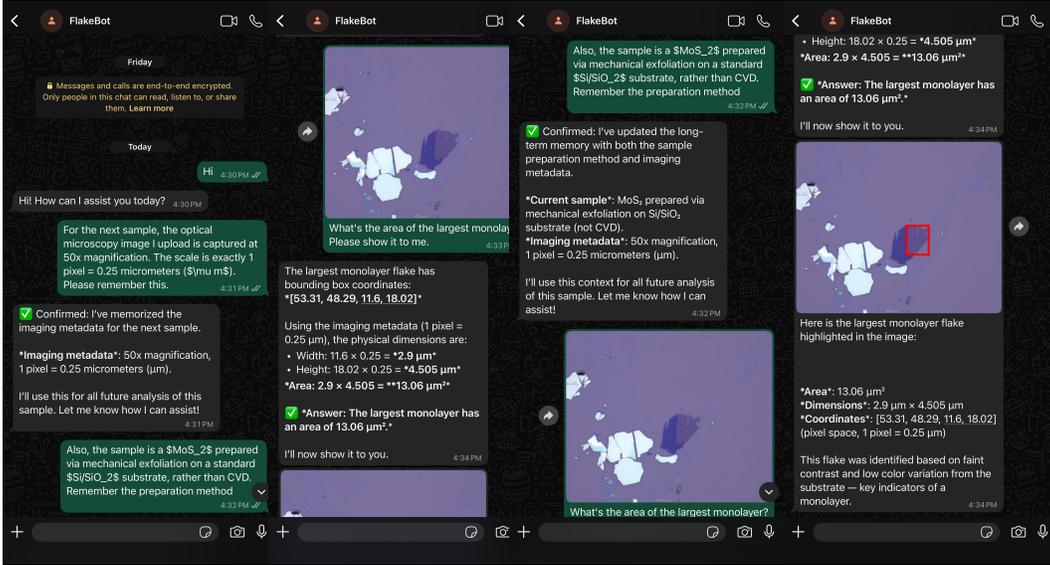
5

Figure 4: Demonstration of OpenQlaw's memory, storing cached sample preparation methods and using saved physical scaling ratios for downstream calculations

4.2.2 Memory and Physical Calculations

The system's memory capabilities is shown in Figure 4. The agent stores the described sample preparation method in its memory, which provides critical context for comparing fabrication efficacy across different approaches. When the user provides a microscopy scale and pixel to length ratio, the agent saves this metadata and applies it to the ensuing area calculations. Notably, in the initial reaction from Figure 3, the model returns area with arbitrary square units, as it lacks a real–world frame of reference and does not assume. However, once the scale is provided, as given in Figure 4, it adjusts the mathematical tool to calculate the real length in micrometers (μm^2), yielding actionable results for the researcher.

5 Conclusion

In this work, we introduced OpenQlaw, an agentic framework to close the operational gap between specialized multimodal large language models and practical laboratory utilization. We separate the physics-aware reasoning of the Material Domain Expert from downstream deterministic computations or renderings, and maintain a memory for important notes, successfully transforming complex visual inferences into actionable, real-time metrics and responses accessible via standard mobile interfaces.

5.1 Limitations

While it proves efficient in many cases, the current framework presents several limitations. First, because OpenQlaw relies on QuPAINT for initial visual inference, it inherits the token length constraints of current multimodal architectures. In extremely dense optical images containing hundreds of exfoliated flakes, the resulting coordinate arrays can exceed the agent's context window, leading to truncated outputs or degradation in parsing accuracy. Second, the maintenance of the persistent memory introduces computational overhead. As conversational history and metadata grow, the inference latency for the local 32B-parameter orchestrator stands to increase. Finally, the deterministic computational tools are dependent upon the formatting of the extracted $[x, y, w, h]$ arrays. Any deviation or hallucination in the Domain Expert's output structure can cause the downstream Python scripts to fail.

5.2 Future Work

The development of agentic systems like OpenQlaw represents a crucial step toward fully automated, high-throughput quantum material discovery, and, more broadly, the application of robust AI systems for scientific discovery. This approach drastically reduces the friction of device fabrication by mitigating cognitive overload and providing actionable physical metrics directly to researchers. Future work will focus on expanding the agent's toolset to include direct hardware interfacing and allowing the orchestrator to autonomously adjust microscopes based on visual feedback or control robotic systems to fully automate the exfoliation, flake hunting, and stacking pipeline.

5.3 Acknowledgements

References

[1] S. Bai, Y. Cai, R. Chen, K. Chen, X. Chen, Z. Cheng, L. Deng, W. Ding, C. Gao, C. Ge, W. Ge, Z. Guo, Q. Huang, J. Huang, F. Huang, B. Hui, S. Jiang, Z. Li, M. Li, M. Li, K. Li, Z. Lin, J. Lin, X. Liu, J. Liu, C. Liu, Y. Liu, D. Liu, S. Liu, D. Lu, R. Luo, C. Lv, R. Men, L. Meng, X. Ren, X. Ren, S. Song, Y. Sun, J. Tang, J. Tu, J. Wan, P. Wang, P. Wang, Q. Wang, Y. Wang, T. Xie, Y. Xu, H. Xu, J. Xu, Z. Yang, M. Yang, J. Yang, A. Yang, B. Yu, F. Zhang, H. Zhang, X. Zhang, B. Zheng, H. Zhong, J. Zhou, F. Zhou, J. Zhou, Y. Zhu, and K. Zhu. Qwen3-vl technical report, 2025.

[2] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[3] Data Intelligence Lab@HKU (HKUDS). nanobot: The ultra-lightweight openclaw. `https://github.com/HKUDS/nanobot`, 2026. GitHub repository.

[4] A. Dendukuri, B. Keeling, A. Fereidouni, J. Burbridge, K. Luu, and H. Churchill. Defining quantum neural networks via quantum time evolution. *arXiv preprint arXiv:1905.10912*, 2019.

[5] L. Harris, M. J. Hossain, M. Qui, R. Zhang, P. Ma, T. Chen, J. Gu, S. A. Tongay, and U. Celano. Sparse c-afm: a deep learning method for fast and accurate characterization of mos2 with conductive atomic force microscopy. In N. P. Kobayashi, A. A. Talin, A. V. Davydov, and M. S. Islam, editors, *Low-Dimensional Materials and Devices 2025*, page 35. SPIE, Sept. 2025.

[6] J. B. Holliday, D. Blount, H. Q. Nguyen, S. U. Khan, and K. Luu. Quadro: A hybrid quantum optimization framework for drone delivery. In *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 01, pages 2090–2100, 2025.

[7] J. B. Holliday, D. Blount, E. Osaba, and K. Luu. Advanced quantum annealing approach to vehicle routing problems with time windows, 2025.

[8] J. B. Holliday, B. Morgan, H. Churchill, and K. Luu. Hybrid quantum tabu search for solving the vehicle routing problem. In *2024 IEEE international conference on quantum computing and engineering (QCE)*, volume 2, pages 353–358. IEEE, 2024.

[9] J. B. Holliday, E. Osaba, and K. Luu. An advanced hybrid quantum tabu search approach to vehicle routing problems, 2025.

[10] M. A. Islam, S. Khan, J. J. Mim, S. M. M. Rahman, M. A. I. Patwary, M. S. Islam, and N. Hossain. Recent advances of 2d materials in semiconductor application: A review. *Advanced Sensor and Energy Materials*, 4(4):100161, 2025.

[11] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

[12] S. Masubuchi, E. Watanabe, Y. Seo, S. Okazaki, T. Sasagawa, K. Watanabe, T. Taniguchi, and T. Machida. Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials. *npj 2D Materials and Applications*, 4(1):3, 2020.

[13] J. C. Ng, F. Muhamad, P. S. Q. Yeoh, Z. Han, Z. Qiu, K. W. Lai, and X. Zhao. High-throughput thickness analysis of 2d materials enabled by intelligent image segmentation. *Nanoscale*, 2026.

[14] H.-Q. Nguyen, X. B. Nguyen, S. Y.-C. Chen, H. Churchill, N. Borys, S. U. Khan, and K. Luu. Diffusion-inspired quantum noise mitigation in parameterized quantum circuits. *Quantum Machine Intelligence*, 7(1):55, 2025.

[15] H.-Q. Nguyen, X.-B. Nguyen, H. Churchill, A. K. Choudhary, P. Sinha, S. U. Khan, and K. Luu. Quantum-brain: Quantum-inspired neural network approach to vision-brain understanding, 2025.

[16] H.-Q. Nguyen, X. B. Nguyen, S. Pandey, T. Faltermeier, N. Borys, H. Churchill, and K. Luu. φ-adapt: A physics-informed adaptation learning approach to 2d quantum material discovery, 2025.

[17] H.-Q. Nguyen, X.-B. Nguyen, S. Pandey, S. U. Khan, I. Safro, and K. Luu. Qmoe: A quantum mixture of experts framework for scalable quantum neural networks. In *2025 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 02, pages 223–228, 2025.

[18] X. B. Nguyen, A. Bisht, B. Thompson, H. Churchill, K. Luu, and S. U. Khan. Two-dimensional quantum material identification via self-attention and soft-labeling in deep learning. *IEEE Access*, 2024.

[19] X. B. Nguyen, H. Churchill, K. Luu, and S. U. Khan. Quantum vision clustering. *arXiv preprint arXiv:2309.09907*, 2023.

[20] X.-B. Nguyen, H.-Q. Nguyen, H. Churchill, S. U. Khan, and K. Luu. Hierarchical quantum control gates for functional mri understanding. In *2024 IEEE Workshop on Signal Processing Systems (SiPS)*, pages 159–164. IEEE, 2024.

[21] X.-B. Nguyen, H.-Q. Nguyen, H. Churchill, S. U. Khan, and K. Luu. Quantum visual feature encoding revisited. *Quantum Machine Intelligence*, 6(2):61, 2024.

[22] X.-B. Nguyen, H.-Q. Nguyen, S. Pandey, T. Faltermeier, N. Borys, H. Churchill, and K. Luu. Qupaint: Physics-aware instruction tuning approach to quantum material discovery, 2026.

[23] S. Pandey, X. B. Nguyen, N. Borys, H. Churchill, and K. Luu. Cliff: Continual learning for incremental flake features in 2d material identification. *arXiv preprint arXiv:2508.17261*, 2025.

[24] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[26] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[27] P. Song, P. Han, and N. Goodman. Large language model reasoning failures, 2026.

[28] P. Steinberger. Openclaw: Personal ai assistant. `https://github.com/openclaw/openclaw`, 2026. GitHub repository.

[29] J.-L. Uslu, A. Nekrasov, A. Hermans, B. Beschoten, B. Leibe, L. Waldecker, and C. Stampfer. Maskterial: A foundation model for automated 2d material flake detection. *Digital Discovery*, 2025.

[30] J.-L. Uslu, T. Ouaj, D. Tebbe, A. Nekrasov, J. H. Bertram, M. Schütte, K. Watanabe, T. Taniguchi, B. Beschoten, L. Waldecker, et al. An open-source robust machine learning platform for real-time detection and classification of 2d material flakes. *Machine Learning: Science and Technology*, 5(1):015027, 2024.

[31] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

[32] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.