nature machine intelligence

Supplementary information https://doi.org/10.1038/s42256-026-01261-5

An agentic artificially intelligent X-ray scientist

In the format provided by the authors and unedited

Contents

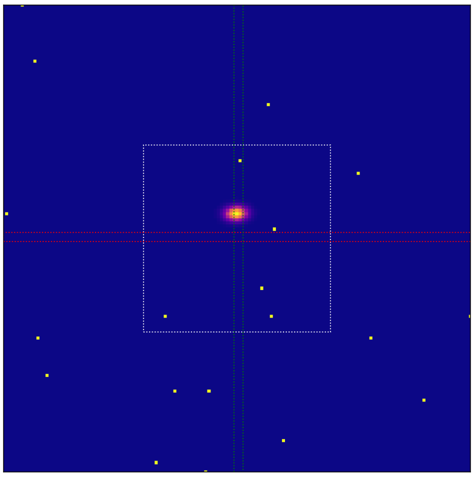
Supplementary Note 1 Prompt materials for AI X-ray Scientists

As referenced in Large language models of Methods, the goal of the short prompt is to provide the AI X-ray Scientist with the essential information while re-emphasizing certain points that might otherwise be overlooked when relying solely on the long prompt. The long prompt, in contrast, contains detailed instructions for performing sample alignment and is available in the shared code repository, as noted in Data Availability and Code Availability. A copy of the short prompt for the real experimental demonstration at SSRL BL17-2 is provided below.

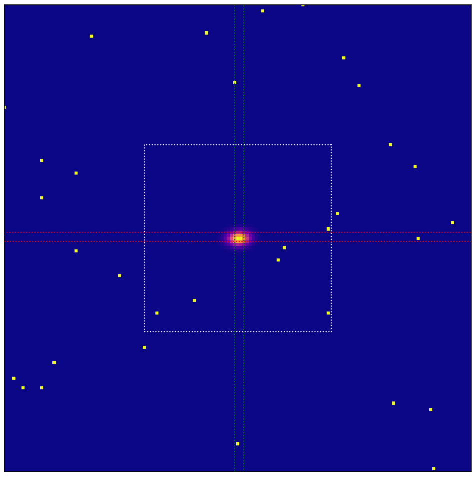
> Could you help align my sample? You may use the .env file located at: 'path/to/env' for accessing necessary information. When creating a new experiment, use the config_path="path/to/config.yaml" as the argument to load correct configuration file.
>
> The experimental operations will completely follow your intention but implemented by human. You will prefix your message Execute Command: [command] to human, and [command] will be typed into the experimental control computer's terminal.
>
> Request one [command] at a time, wait for [command] response before requesting the next [command].
>
> Please pay additional attention to follow instructions, especially for finding the 2nd reference reflection. ** IMPORTANT: You should remember to do fine scans when the scanned peak is poorly sampled and overly sharp.**
>
> Be precise and use enough digits (at least 4 digits) when you want to use 'umv' to move motors, as the peaks can be extremely sharp.
>
> You will work with a vertical reflection setup. Make sure you optimize lattice constants when appropriate; for cubic systems, avoid optimizing the lattice constants upon finding the second reference reflection. Before you setlat or or0/or1, always check the detector image first to make sure the peak is well optimized and centered.
>
> ** CRITICAL: If you only capture the tail of motor scan instead of the entire peak, you MUST extend your scan range to capture the peak accurately. **
>
> ** CRITICAL: If you noticed an offset of any motor, you MUST account for this offset value when you use theoretically calculated motor angles when searching for reference reflections. **
>
> The in-plane orientation of the sample is completely unknown, so you should expect to perform a wider phi scan in the beginning to capture the non-specular peak based on the crystal symmetry. ** CRITICAL: Please avoid requesting overly wide scan such as phi from -180 to 180 **
>
> There might be few degrees offsets in motors, and be careful not to recognize hot pixels as peaks. I am attaching two example detector images with and without a peak for your reference. If you do not see a peak at initial search, try starting with increased scan range.

Supplementary Note 2 Example detector images for various alignment accuracies

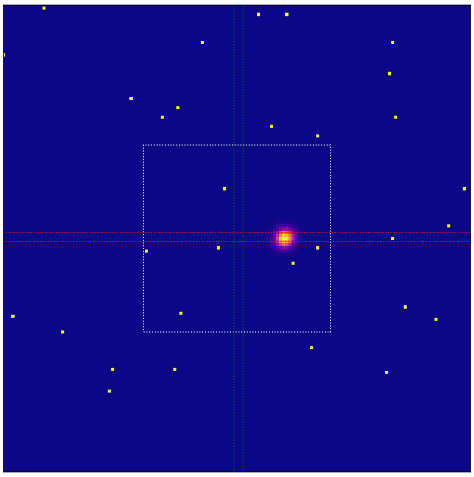
We present two representative detector images for the $(0, 0, 6)$ peak, corresponding to a relatively imperfect alignment and an improved alignment, in Figure S1(a) and (b). An additional example of an imperfect alignment for the $(2, 0, 14)$ peak is shown in Figure S1(c). These cases illustrate the runs with larger offsets highlighted in Figure 2(b). The most common cause of such imperfect alignments is an insufficient scan range of the relevant motors during peak optimization. In the case of Figure S1(a), the scan over the δ motor was not wide enough, causing the AI X-ray Scientist to miss the true peak. For Figure S1(c), further optimization of the χ or ϕ motor is required.



(a) Peak $(0, 0, 6)$



(b) Peak $(0, 0, 6)$



(c) Peak $(2, 0, 14)$

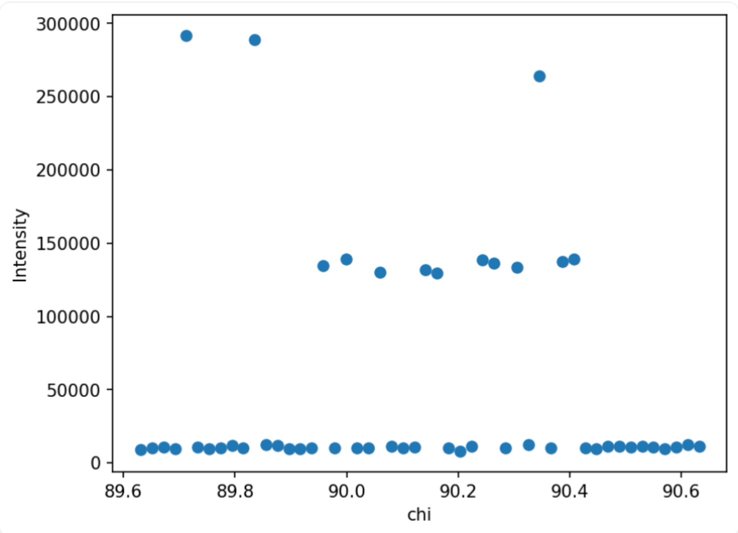
Supplementary Figure S1: Representative detector images of diffraction peaks: **(a)** $(0, 0, 6)$ with a relatively imperfect alignment, **(b)** $(0, 0, 6)$ with an improved alignment, and **(c)** $(2, 0, 14)$ with an imperfect alignment. Detector images are displayed using a false-color intensity map normalized independently for visualization; color indicates relative detector intensity and is not used for quantitative comparison.

We further provide an example of an off-centered peak resulting from an insufficient χ motor scan range in Figures S2 and S3. In this case, although the peak-centering process is iterative, the centering ultimately fails because the LLM does not bring the peak across the corresponding line ROI, leading to a considerable final alignment error of $18.5254°$. Future improvements could involve explicitly prompting the agent to expand the scan range or incorporating warnings about poorly defined peaks into the scan MCP tool.

Now that the `eta` motor is at its refined position of 2.0149°, I'll re-optimize `chi` for the (1,0,10) peak.

I'll perform a `chi` scan over a range of ±0.5° around its current position (90.1313°), with 50 intervals and a 0.5-second exposure. I'll continue to use the `pd7` (vertical line) ROI and a `tophat` fit, as this range has been more successful in capturing the peak profile for `chi` previously.

Ran execute_command

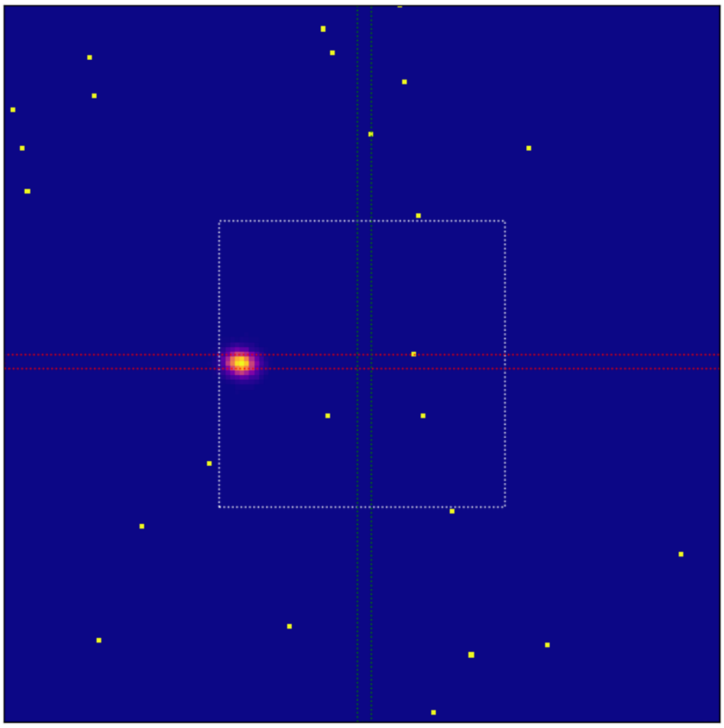


The `chi` scan (chi_1010_scan_7) once again shows a very scattered profile with no clear, well-defined peak. This is consistent with our previous difficulties in optimizing `chi` for this reflection. It's becoming evident that the peak is either extremely broad in `chi` or has a very complex shape, making it difficult to resolve with standard scanning techniques. Also, there are high intensity points at the lower end of the scan range which are likely noise.

Supplementary Figure S2: Representative example of a failed peak-centering process resulting from an insufficient motor scan range (e.g., the χ motor), which leads to unsuccessful peak centering during the scan

Before setting this as our secondary orientation reflection, I will take a detector exposure to visually confirm that the peak is well-centered.

Ran execute_command



The detector image `final_1010_check.raw` confirms that the (1,0,10) peak is visible and well-defined, although it is not perfectly centered in the `pd5` ROI. It is slightly to the left and up from the center.

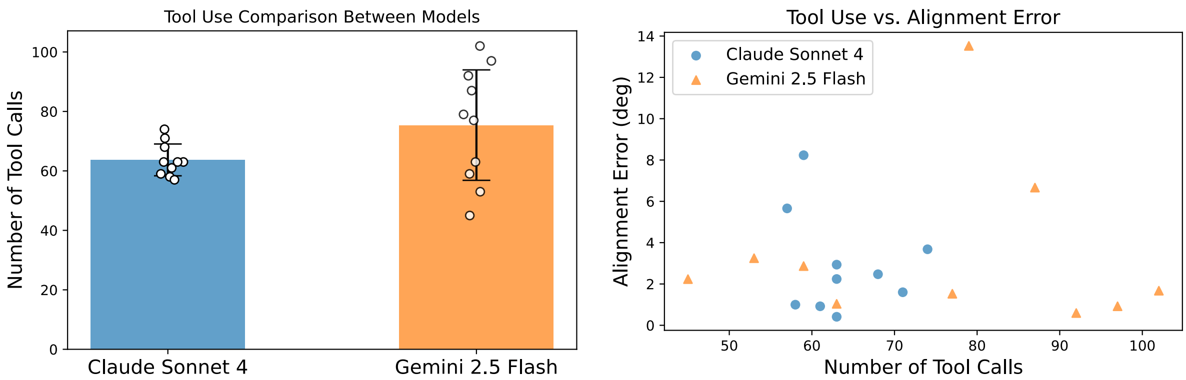
Supplementary Figure S3: Representative example of a failed peak-centering process resulting from an insufficient motor scan range (e.g., the χ motor), which leads to an off-centered peak.

Supplementary Note 3 Tool-use calls

We report the number of tool-use calls issued by the agent in Figure S4a. Each tool-use call corresponds to an explicit interaction between the LLM and the experimental control interface (e.g., motor motion, detector acquisition, or data query), and thus serves as a coarse measure of the interaction complexity of the agentic workflow.

As shown, the two models require comparable numbers of tool-use calls to complete the alignment task, indicating that the overall structure of the agentic workflow is similar across models. We note that the absolute number of tool calls depends on model-specific reasoning behavior and is not intended as a fine-grained efficiency benchmark.

To further examine whether the number of tool-use calls is directly related to alignment performance, we plot the final alignment error as a function of the total number of tool-use calls (Figure S4b). Across the tested cases and models, we do not observe a strong or systematic correlation between the number of tool-use calls and the final alignment errors.



(a) Number of tool-use calls (b) Tool-use calls vs. alignment error

Supplementary Figure S4: **(a)** Comparison of numbers of tool-use calls by two models. The bar heights indicate mean values across runs, error bars denote ± 1 standard deviation, and circular markers represent individual runs. The results are obtained from $n = 10$ independent simulated experiments. **(b)** Final alignment error as a function of the total number of tool-use calls.

Supplementary Note 4 Robustness benchmarks under non-ideal experimental conditions

Here, we discuss the performance of the Gemini 2.5 Flash-based AI X-ray Scientist under a set of challenging experimental scenarios as shown in Figures 2(d) and 2(e). These include cases with (1) higher motor offsets, where the nominal motor readouts differ more significantly from the true physical positions, modeled as $\Delta\eta \sim \mathcal{N}(0.0°, 1.0°)$ and $\Delta\phi \sim \mathcal{N}(0.0°, 3.0°)$, (2) diffraction from a hypothetical multi-grained domain, (3) reduced reasoning with a fixed 512-token thinking budget, compared to the default setting in which the model dynamically allocates its reasoning budget, and (4) reduced model temperature from the default value $T = 1.0$ to $T = 0.5$. Note that some large alignment errors or aborted runs were associated with unmodeled motor-limit behavior during wide scans, in which a motor did not return to its nominal starting position after reaching a position boundary. While this motor behavior was unintended, such edge cases introduce additional imperfections that require adaptive reasoning rather than scripted recovery, and therefore serve as stress tests for the closed-loop agentic workflow.

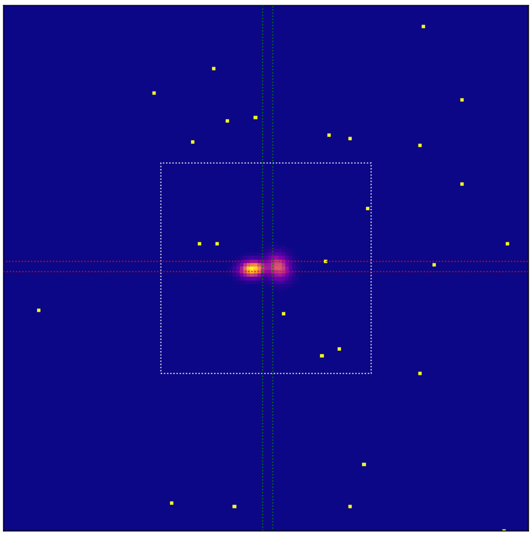
Across these scenarios, the agent was able to determine the sample orientation and lattice parameter c with reasonable accuracy in most successful runs, despite increasingly non-ideal experimental conditions. Performance, however, was found to degrade in more adversarial cases, particularly under reduced reasoning budgets, where the agent exhibited bimodal behavior: it either completed the task smoothly or failed abruptly when the reasoning depth was insufficient. Meanwhile, increasing motor offsets introduce more intrinsic ambiguities between nominal and actual motor positions, which can confuse the alignment process and lead to larger errors. These results suggest that, provided the experimental instrumentation is reasonably calibrated and sufficient reasoning capacity is available, the AI X-ray Scientist can maintain effective closed-loop operation such as its displayed robustness on multi-grained samples. We also show a representative detector image and the corresponding motor scan in Supplementary Note 5, which illustrate limitations of the current peak-fitting tools and hints at potential improvement by preparing more robust analysis tools for AI.

We emphasize that these benchmark results are not intended as absolute “scores” comparing different LLMs on related experimental tasks. Because the guidance prompts were primarily developed with Claude models during exploration, the setup may be somewhat more favorable to them. The results are presented merely to illustrate that our approach is broadly applicable across different agentic AI models. Overall, the benchmarks demonstrate the robustness and consistency of our agentic AI X-ray Scientist, supporting its potential for reliable, fully automated experimental workflows.

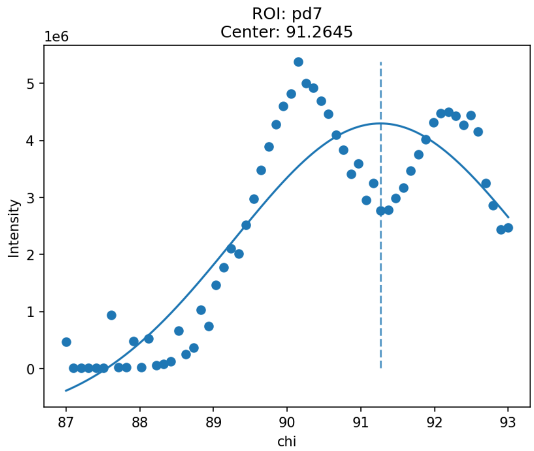
Supplementary Note 5 Alignment of multi-grain samples

We show a representative simulated detector image and the corresponding χ motor scan in Figure S5, which were used to construct the multi-grain test case reported in Figure 2. While the visual disturbance introduced by the presence of multiple grains is relatively mild for the AI agent at the detector-image level, the peak-fitting tool provided to the agent fails to robustly isolate a single diffraction peak. Instead, the fitted peak center is biased toward the midpoint between the two nearby peaks, causing the motor to be aligned to an incorrect position.

We note that this limitation arises from the simplicity of the current peak-fitting implementation and could be mitigated in future work by incorporating additional fitting criteria, such as prioritizing the highest-intensity peak or explicitly detecting multi-peak structures. More robust and informative MCP tools for peak identification and confidence assessment are therefore expected to further improve the reliability of AI-driven autonomous experimentation in such non-ideal scenarios.



(a) Detector image of the $(0, 0, 6)$ peak

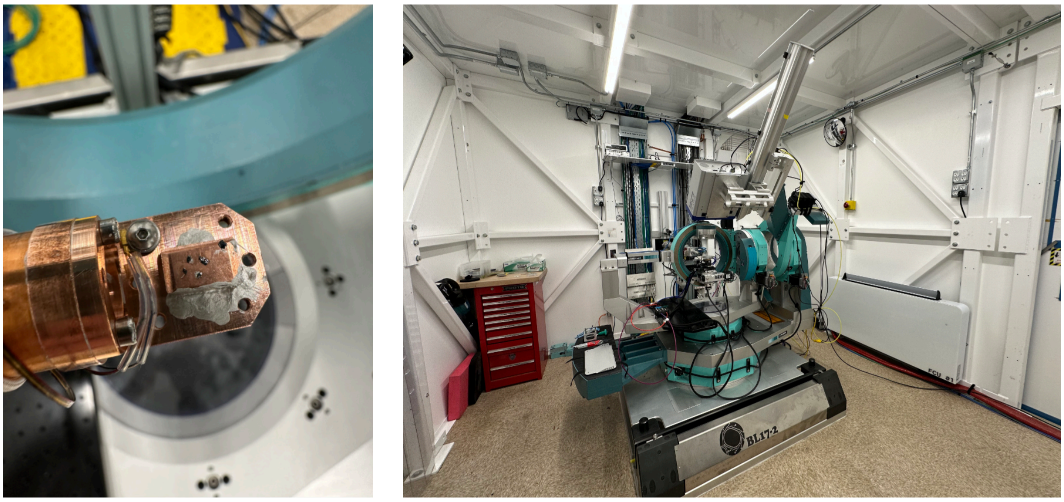


(b) Motor χ scan the of $(0, 0, 6)$ peak

Supplementary Figure S5: **(a)** Detector image of the $(0, 0, 6)$ reflection showing two diffraction peaks arising from two crystalline grains with a small relative orientation offset. **(b)** The corresponding χ motor scan exhibiting a clear double-peak structure consistent with the multi-grain diffraction observed in the detector image. Detector images are displayed using a false-color intensity map normalized independently for visualization; color indicates relative detector intensity and is not used for quantitative comparison.

Supplementary Note 6 Sample and six-circle diffractometer images

Photographs of the $Co_3Sn_2S_2$ sample and the six-circle diffractometer are presented in Figure S6.



(a) (b)

Supplementary Figure S6: Photographs of **(a)** the $Co_3Sn_2S_2$ sample mounted on a copper holder and **(b)** the six-circle diffractometer used in the synchrotron X-ray experiment.

Supplementary Note 7 Verification peak from the real experiment

We present the verification peak from the real experiment demonstration described in the section Deploying the AI X-ray Scientist on a real synchrotron experiment, shown in Figure S7. The terminal output of the `wh` command reports the calculated reciprocal lattice coordinates (H, K, L), see below.

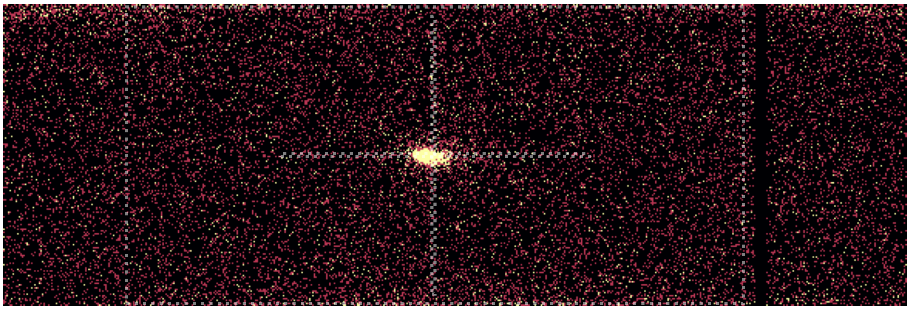
```
1463.PSIC> wh

H K L =  2.0006  0.035186  13.968  Lambda = 0.7749 Two Theta = 52.753

    Alpha      Beta        Psi        Tau        Qaz       Naz      Omega
  30.6042  -19.6516  118.7511   78.7855   90.0000 -177.6443  -23.3986

       del         eta         chi         phi          nu          mu
  52.75289     2.97789    89.01500    -2.63000     0.00000     0.00000
```

The close agreement with the expected $(2, 0, 14)$ peak confirms the high alignment accuracy achieved by the AI X-ray Scientist.

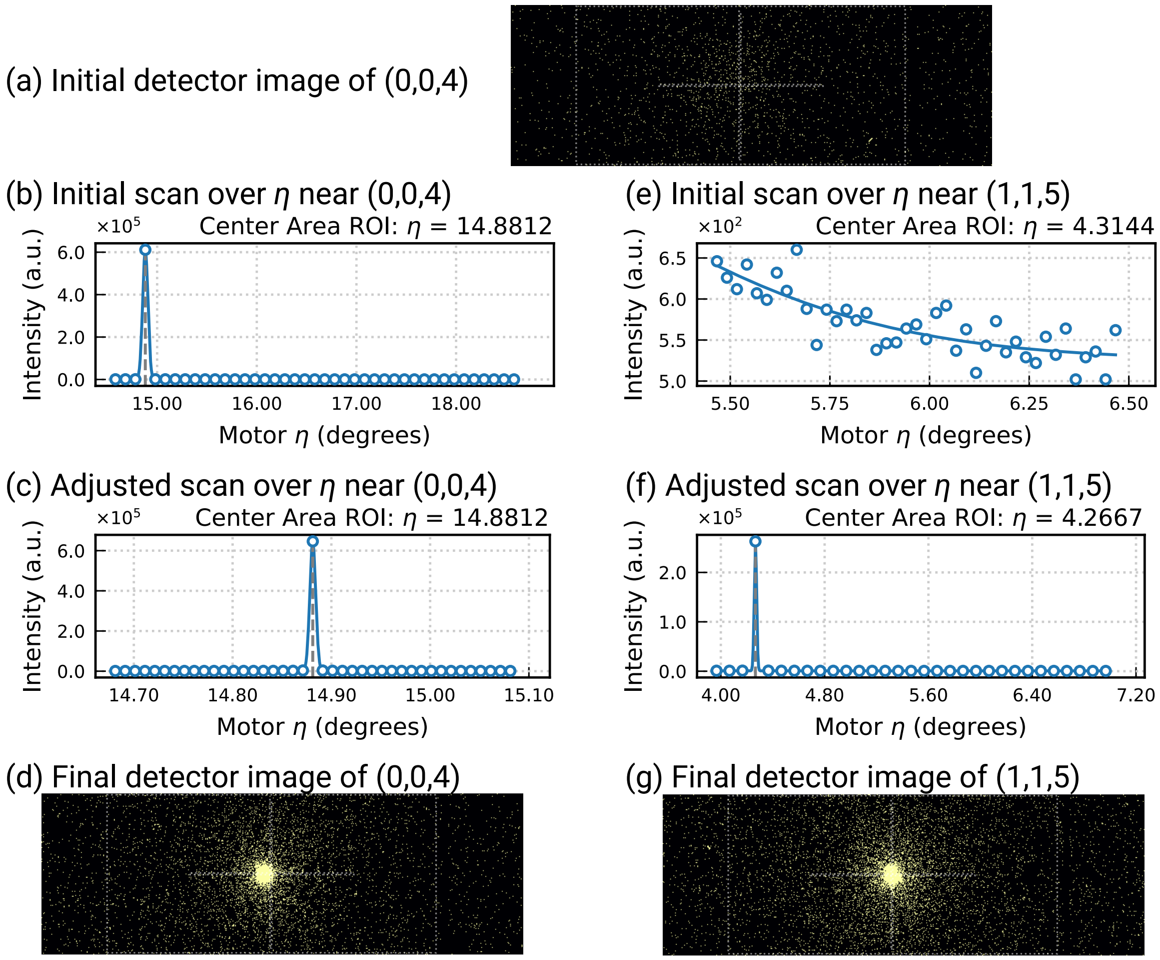


Supplementary Figure S7: Detector image of the verification peak $(2, 0, 14)$ from the real experiment demonstration. Detector images are displayed using a false-color intensity map normalized independently for visualization; color indicates relative detector intensity and is not used for quantitative comparison.

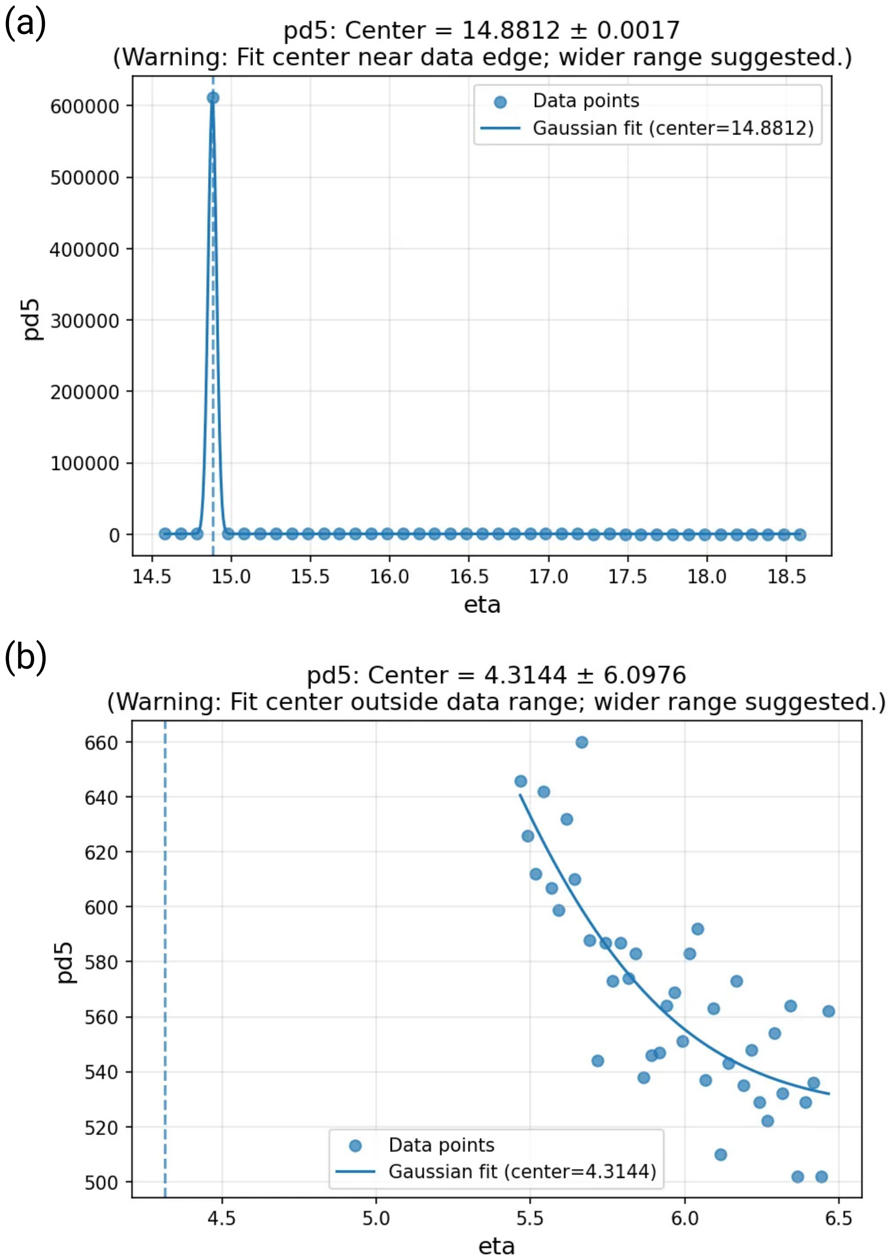
Supplementary Note 8 Real experimental demonstration with single-crystal Si

We also carried out a real experimental demonstration using a single-crystal Si sample. The sample was a fragment broken from a Si wafer. Motor offsets were applied as $\Delta\eta = 1.5^\circ$ and $\Delta\phi = 8^\circ$. Since the Si sample had already been manually aligned during mounting such that $\phi = 0^\circ$ approximately corresponded to the (H, H, L) scattering plane, the $\Delta\phi$ offset did not significantly affect the AI X-ray Scientist. By contrast, because diffraction peaks can be very sharp, the $\Delta\eta$ offset posed a greater challenge.

Key steps taken by the AI X-ray Scientist to resolve the η offset are summarized in Figure S8. In Figure S8(a-d), the AI X-ray Scientist successfully aligns to the $(0, 0, 4)$ peak using wide and adaptive η scans. The true scan image presented to the AI X-ray Scientist included additional instructions from the MCP tool; the corresponding direct MCP output is shown in the left panel of Figure S9. For the secondary reference reflection $(1, 1, 5)$, the initial η scan failed to capture the peak center, see Figure S8(e) and the right panel of Figure S9. The AI X-ray Scientist subsequently adjusted the scan parameters and successfully captured the peak, as shown in Figure S8(f). The final aligned image of the $(1, 1, 5)$ reflection is displayed in Figure S8(g).



Supplementary Figure S8: Key steps of the AI X-ray Scientist resolving the η offset and aligning the $(0, 0, 4)$ and $(1, 1, 5)$ reflections of the Si sample. Detector images are displayed using a false-color intensity map normalized independently for visualization; color indicates relative detector intensity and is not used for quantitative comparison.



Supplementary Figure S9: Initial scans over η near (0,0,4) and (1,1,5) including embedded MCP-tool warnings shown to the AI agent.

Supplementary Note 9 AI conversation history

We have included two files as Supplementary Data:

- `Co3Sn2S2_Claude_Conversation_History.pdf`
- `Si_Claude_Conversation_History.pdf`

which contain the full conversation histories corresponding to the results reported in the main text Deploying the AI X-ray Scientist on a real synchrotron experiment.