

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan ([http://energy.gov/downloads/doe-public](http://energy.gov/downloads/doe-public-access-plan) access-plan).

Supporting Information

SimuScan-Agent: language-model orchestration of synthetic-data-driven autonomous atomic force microscopy.

Ruben Millan-Solsona^{1*}, Marti Checa¹, Yongtao Liu¹, Bernadeta Srijanto¹, Rama K. Vasudevan¹,
Scott Retterer^{1,2}, Liam Collins¹

¹Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

²Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

*Correspondence to: solsonarm@ornl.gov

S1: Illustrative evaluation of a general-purpose prompt-based workflow

General-purpose prompt-based image generators can produce microscopy-like images directly from natural-language instructions. To scope their role in the present workflow, we evaluated one model and prompting strategy, GPT Image 2 (OpenAI), to determine whether it could generate AFM-like images together with candidate labels for machine-learning models. Figure S1.1a summarizes the conceptual workflow of this approach, from natural-language input to image generation and approximate masks.

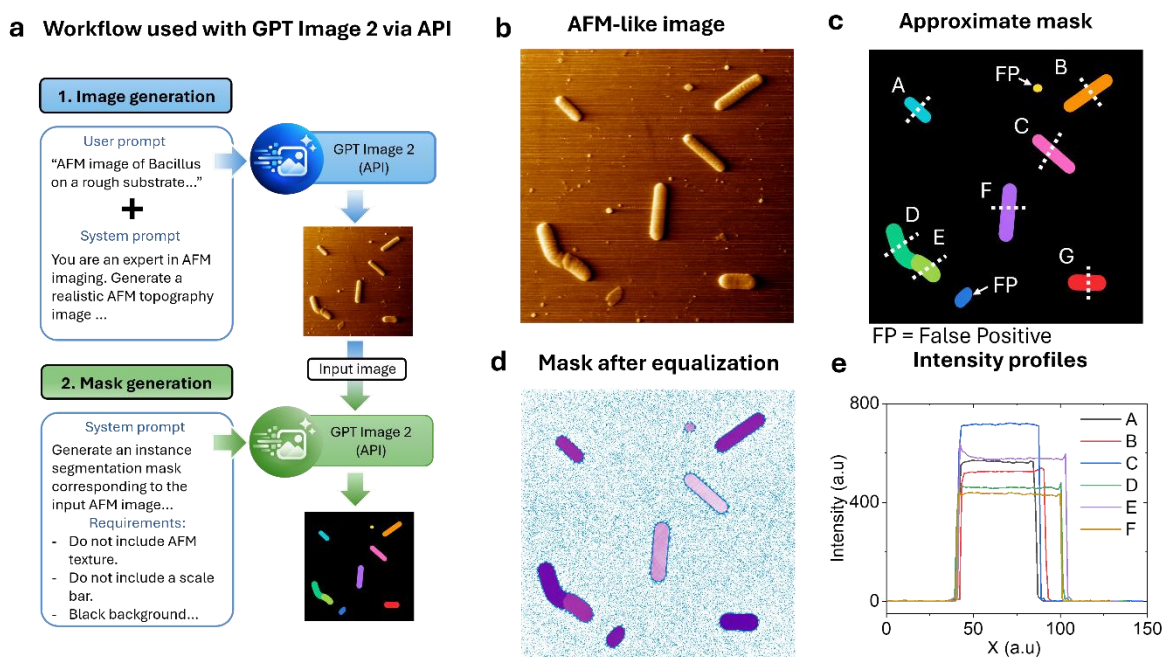


Figure S1.1: Illustrative assessment of one prompt-based workflow for AFM-like image and segmentation-mask synthesis. (a) Schematic of the workflow used with GPT Image 2 via API. In the first stage, a user *prompt* and a *system prompt* are combined to generate an AFM-like image. In a second stage, the generated image is used as input together with new *prompts* to request the corresponding segmentation mask. (b) Representative AFM-like image generated for objects with morphology similar to *Bacillus* on a substrate. (c) Segmentation mask generated from the image in (b); dashed lines A–G indicate the transects used to analyze the intensity values. (d) Representation of the same mask after contrast equalization, revealing intensity variations both within the objects and in the nominal background. (e) Intensity profiles corresponding to transects A–G in (c), showing variations among objects and a non-strictly discrete separation of the values assigned to the different regions.

Representative *Bacillus*-like bacterial cells and circular DNA structures were generated together with their corresponding segmentation masks (Fig. S1.1 and S1.2). The resulting images contained several visual cues associated with AFM-like renderings, including plausible

morphologies, topographic contrast, surface texture, and background variations. However, closer inspection revealed features inconsistent with a physically parameterized AFM image-formation process. The images frequently contained shading, shadows, and textures more characteristic of optical images, whereas experimentally relevant parameters such as object geometry, effective tip radius, tip-sample convolution, substrate roughness, instrumental noise, or scan distortions could not be independently defined or controlled. Thus, an AFM-like visual appearance does not necessarily imply physical fidelity or reproducible control of the image-formation process.

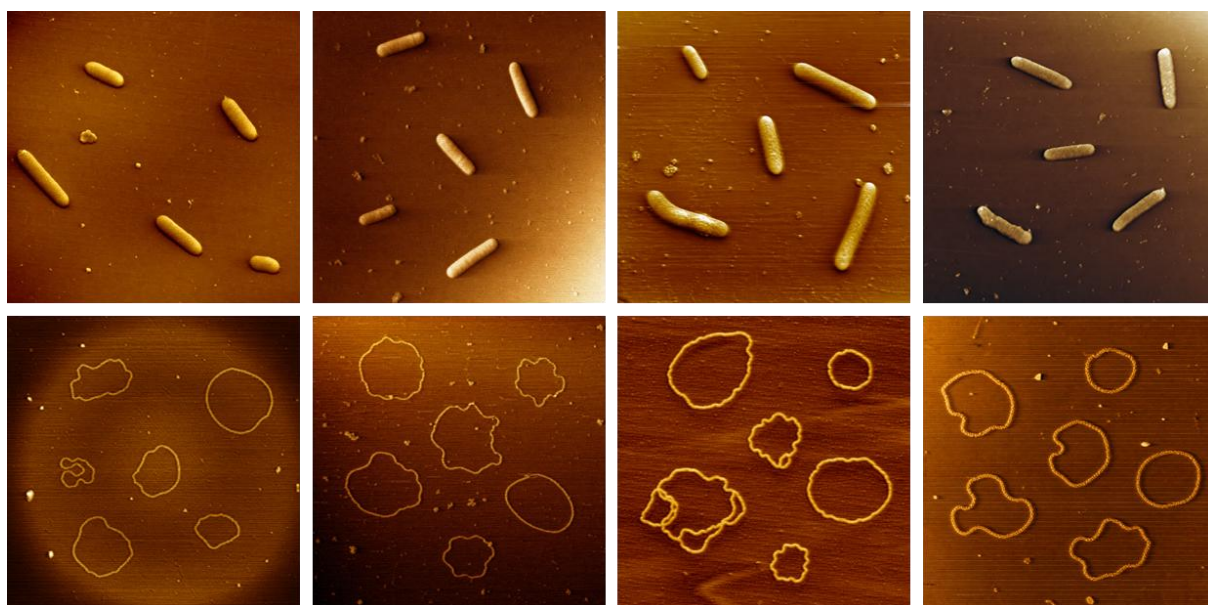


Figure S1.2 AI-generated AFM-like images of Bacillus cells and circular DNA. Representative images generated in May 2026 using OpenAI's GPT Image 2 model (gpt-image-2), with an image size of 1024×1024 pixels, medium quality (quality="medium"), and PNG output format (output_format="png"). The top row shows Bacillus cells, and the bottom row shows circular DNA molecules. Images were generated using the following prompt, replacing the target object with either "Bacillus cells" or "circular DNA molecules": "Generate a realistic AFM height-map image of [Bacillus cells/circular DNA molecules] adsorbed on a flat substrate. Each image should contain exactly six [Bacillus cells/circular DNA molecules] of varying shapes and sizes. Include distinct AFM artifacts such as scan lines, background tilt, line noise, tip convolution, drift, small debris particles, and mild substrate roughness. No labels, no scale bar, no mask, just the AFM image."

In addition, we found that, in the evaluated prompt-only workflow, the generated segmentation masks did not provide reliable reference labels without additional validation.

Although they visually appeared to be close to correct masks (Fig. S1.1c), pixel values varied within both the objects and the background (Fig. S1.1d), and intensity profiles showed poorly defined values across different regions (Fig. S1.1e). The difficulty of directly generating discrete labels has also been recognized in the diffusion-model literature, where task-specific diffusion formulations have been developed for discrete data, segmentation, and paired microscopy image-label synthesis^{1,2}. Subsequent binarization can produce a mask with defined class values, but it does not necessarily correct errors between the mask and the generated image. Additional examples showed false positives (FP), incomplete contours, and incorrectly segmented regions (Fig. S1.3). These outputs therefore require additional processing and independent validation before they can be used as reference labels for supervised training or quantitative evaluation.

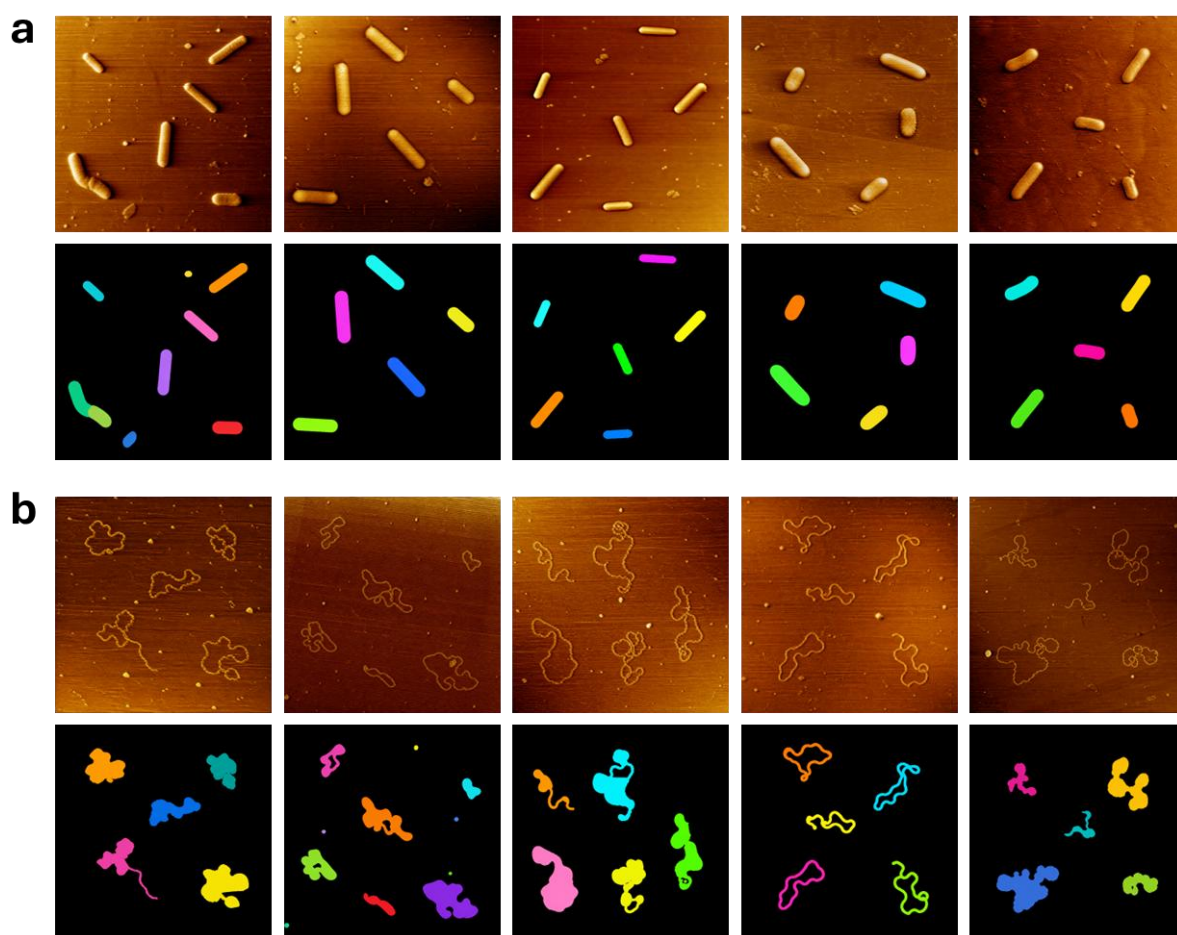
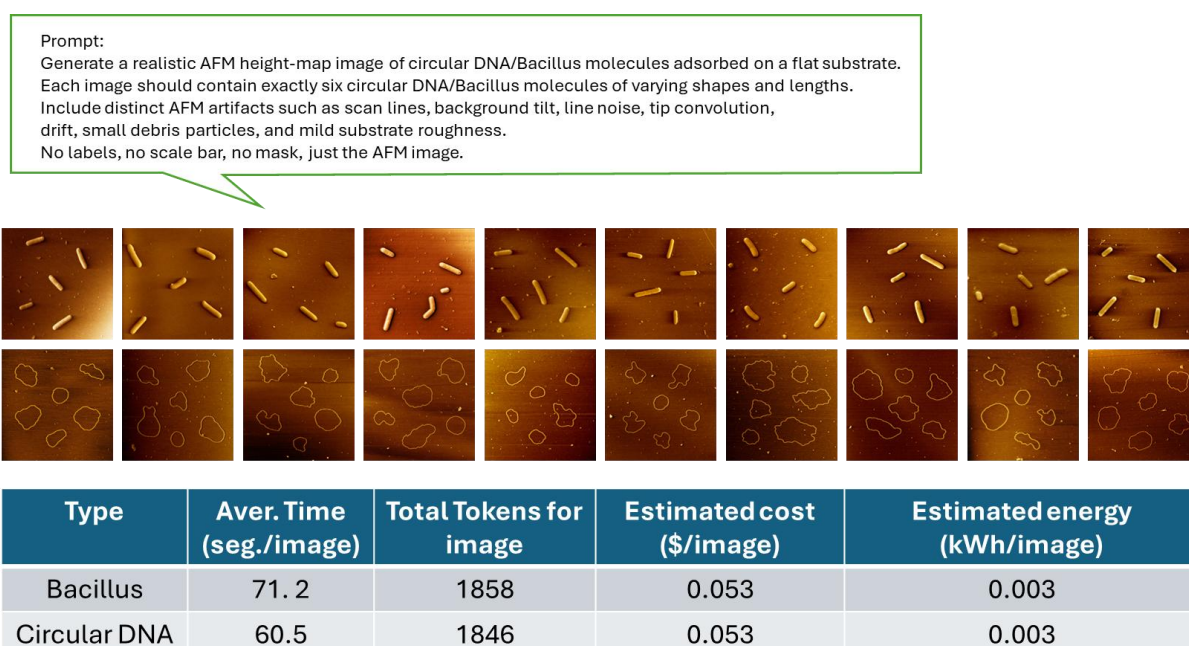


Figure S1.3: Instance masks generated from diffusion-model-produced AFM-like images of *Bacillus* cells and circular DNA molecules. Representative AFM-like images and their corresponding instance masks generated with GPT-Image-2. **(a)** *Bacillus* cells. The upper row shows the generated AFM-like images, whereas the lower row shows the corresponding instance masks generated from each image using the same model and a dedicated mask-generation prompt. **(b)** Circular DNA molecules. The upper row shows the generated AFM-

like images, and the lower row shows their corresponding instance masks. Each mask was intended to assign a distinct region to every detected object while separating the objects from the substrate background. The examples include false-positive regions, incomplete segmentation, and poorly defined object boundaries.

For the API-based settings evaluated here, scaling to large datasets also introduces practical costs^{3,4}. Training modern machine-learning models often benefits from large and diverse datasets, so individual generation with generative models introduces additional costs in time, computation, energy, and service usage. Figure S1.4 compares representative generation times, energy consumption, and monetary costs for the circular DNA and Bacillus-like datasets, while Table S1.1 summarizes the estimated resources required to generate 1,000 images using GPT Image 2 and SimuScan. Under the timing, pricing, and energy-estimation assumptions reported in the Supplementary Information, extrapolation to 1,000 images yielded approximately 12 h and an estimated API cost of \$20-100 for the evaluated prompt-based workflow; SimuScan required about 35 min of local computation and incurred no API-usage cost. Under these implementation-specific assumptions, this corresponds to a more than 20-fold difference in elapsed generation time; the energy comparison should be interpreted as an order-of-magnitude estimate rather than a platform-independent benchmark.



Medium quality by model gpt-image-2 (1024 x 1024) in May 2026

Figure S1.4: Additional AFM-like images of *Bacillus* cells and circular DNA molecules, with estimated generation metrics. Additional examples of AFM-like images generated with

GPT-Image-2 at medium quality and a resolution of 1024×1024 pixels. Ten images were generated for *Bacillus* cells and ten for circular DNA molecules using the same prompt structure, replacing only the target object type in each case. The table below the image set summarizes the token usage, average generation time, estimated monetary cost, and estimated energy consumption per image.

Table S1.1: Estimated computational, economic, and energy costs.

Metric per 1,000 images	Diffusion generation	SimuScan
Estimated cost	\$20 - 100	No API cost
Estimated energy	8.4 kWh	0.3kWh
Estimated generation time	12 h	35 min

Table S1.1: Comparison of the estimated resources required to generate a dataset of 1,000 synthetic AFM-like images using GPT-Image-2 and SimuScan. The GPT-Image-2 estimates were extrapolated from the token usage, average generation time, monetary cost, and energy consumption estimated for the images reported in Figure S1.4. SimuScan estimates were calculated from the average execution time and computational power consumption of the hardware used for image generation. Reported values are estimates and may vary depending on model pricing, hardware configuration, computational load, and execution conditions.

Taken together, this evaluation indicates that the tested prompt-only workflow is useful for rapidly proposing conceptual images and morphologies but did not consistently provide the physical parameter control or label fidelity required by our training pipeline. Rather than implying a fundamental limitation of generative models, these results identify practical requirements for their integration into AFM machine-learning workflows, including reliable mask generation and post-processing, control over physically meaningful parameters, scalable generation costs, and management of heterogeneous training data. As these platforms continue to evolve, specialized models, code-based generation, or additional processing steps may overcome some of these limitations. Our evaluation therefore serves as a case study of the workflow requirements that must be addressed when generative models are used to produce labeled data for quantitative AFM analysis, rather than as a general assessment of their capabilities. At the same time, the ability to describe objects and morphologies through natural language provides a complementary advantage that can be combined with physical simulators such as SimuScan to generate controlled and automatically labeled synthetic data.

For the present workflow, SimuScan provides a complementary and cost-effective approach in which geometry, classes, positions, and labels are explicitly defined before image formation and experimental artifacts are simulated, enabling the generation of physically controlled, traceable, and reproducible datasets. The main differences and complementarities between the two approaches are summarized in Table S1.2.

Table S1.2: Comparison of strategies for generating synthetic AFM data

Feature	Prompt-based workflow	SimuScan
Generation method	Provider-hosted, prompt-based image synthesis	Local, physics-guided simulation
Dependency	Model/provider dependent	Locally controlled
Visual realism	Visually plausible in the examples	Experimentally benchmarked ⁵ in Ref. 5
AFM physics	Not explicitly parameterized	Explicitly parameterized components
Training labels	Post hoc outputs; validation required	Derived from simulator geometry
Metadata	Provider- and workflow-dependent	Configured object-level metadata
Reproducibility	Sensitive to model version and prompt	Deterministic conditional on code, configuration, and seed
Parameter control	Limited	User-specified simulator parameters
Marginal cost	Per-image API charges	Local compute only
Energy estimation	High	Low
Generation time	Long	Efficient batch generation

S2: Class-specific probabilistic active search for adaptive AFM imaging

Autonomous scanning probe microscopy (SPM) aims to replace fixed, exhaustive scan patterns with adaptive acquisition strategies that select the next measurement from the information already collected. Bayesian active learning and Bayesian optimization have been used in SPM to balance exploitation of promising regions with exploration of uncertain regions, enabling efficient selection of measurement locations or spectroscopy points. In related autonomous SPM studies, the reward is commonly defined from a functional response, spectral signature, or learned structure-property descriptor. Here, we formulate a complementary object-centric strategy for AFM imaging, where the acquisition target is not a scalar response but the discovery of a user-defined number of detected objects from one or more object classes.

The method combines object detection, class-specific probabilistic maps, spatial propagation of evidence, movement-aware path planning, and optional forced exploration of poorly known regions. After each AFM image is acquired, a detector such as YOLO returns object counts for each class. These detections update a probabilistic belief map for every class. The next scan is selected by maximizing a utility function that combines the current probability of finding each class, the uncertainty of the estimate, the remaining deficit relative to the target count, and the cost of moving the AFM stage. This formulation is intended for large-area AFM workflows where the goal is to find enough representative objects, rather than to raster scan the entire surface.

The simulations below evaluate five strategies: a raster scan baseline, a random sparse sampling, a baseline active search, an active search with optional unknown-region exploration, and an active search with a lower movement penalty. All active strategies start with one image at the center of the grid and five random images within a user-defined radius around the center before entering the adaptive loop.

Bayesian active learning for SPM is reviewed by Ziatdinov et al. ⁶. Autonomous SPM experiments using Bayesian optimization for spatial measurement selection were demonstrated by Vasudevan et al. ⁷. Closed-loop active-learning experimental design has also been demonstrated outside SPM, for example in CAMEO⁸, and explainability/human-intervention tools for autonomous SPM have been discussed by Liu et al. ⁹.

S2.1: Method rationale

The goal is to find a user-defined number of objects from one or more classes while minimizing the number of AFM images. The field of view is discretized into a grid of candidate scan positions. Each class has its own probabilistic map, allowing rare and abundant classes to be handled separately. At each iteration, the algorithm balances four terms: the probability of finding target objects, the remaining deficit for each class, the uncertainty of unmeasured or weakly constrained regions, and the cost of moving the scanner to a distant location.

In the simulations, the first scan was taken at the center of the grid, followed by five random scans within a user-defined radius around the center. Subsequent images were selected by the active search policy. A maximum step distance constrained how far the next scan position could be from the current one. The optional unknown-exploration parameter controls the probability of ignoring the main acquisition score and instead selecting the most uncertain reachable position. Setting this parameter to zero recovers the baseline active-search behavior.

S2.2: Spatial Beta-distribution pseudo-count model for class-specific active search

For each object class c and candidate cell i , the adaptive exploration strategy maintains two evidence variables, $\alpha_{c,i}$ and $\beta_{c,i}$. These variables are treated as spatially distributed positive and negative pseudo-counts defining a Beta-distribution representation of the local occupancy belief. The corresponding mean and uncertainty are given by

$$\mu_{c,i} = \frac{\alpha_{c,i}}{\alpha_{c,i} + \beta_{c,i}} \quad (1)$$

where $(\alpha_{c,i})$ and $(\beta_{c,i})$ denote accumulated positive and negative pseudo-count evidence, respectively. In this formulation, $(\mu_{c,i})$ provides the current estimate that cell (i) contains an object of class (c) .

The uncertainty of this estimate is quantified from the standard deviation of the corresponding Beta distribution,

$$\sigma_{c,i} = \sqrt{\frac{\alpha_{c,i}\beta_{c,i}}{(\alpha_{c,i} + \beta_{c,i})^2(\alpha_{c,i} + \beta_{c,i} + 1)}} \quad (2)$$

Large values of $(\sigma_{c,i})$ indicate regions where the occupancy probability is still poorly constrained by the measurements collected so far.

To account for the user-defined search objective, we define a class-dependent deficit term,

$$D_c = \frac{\max(0, T_c - F_c)}{T_c} \quad (3)$$

where (T_c) is the target number of objects requested by the user, and (F_c) is the number of objects of class c found so far. Thus, (D_c) decreases as the target for class (c) is progressively satisfied and becomes zero once the target has been reached.

Information obtained at a measured cell (j) is propagated to neighboring candidate cells using a class-dependent spatial kernel,

$$K_c(i, j) = \exp\left(-\frac{d(i, j)^2}{2l_c^2}\right) \quad (4)$$

where $(d(i, j))$ is the distance between cells (i) and (j) on the exploration grid, and (l_c) is the correlation length associated with class (c) . This kernel encodes the assumption that nearby cells are more likely to share similar occupancy statistics than distant cells.

When the detector identifies $(n_{c,j} > 0)$ objects of class (c) in the measured cell (j) , the positive evidence is updated as

$$n_{c,j} > 0 \Rightarrow \alpha_{c,i} \leftarrow \alpha_{c,i} + n_{c,j} K_c(i,j) \quad (5)$$

This increases the local occupancy estimate not only at the measured location, but also in neighboring cells according to the spatial correlation kernel.

Conversely, when no object of class (c) is detected in cell (j), the absence of detection is incorporated as negative evidence,

$$n_{c,j} = 0 \Rightarrow \beta_{c,i} \leftarrow \beta_{c,i} + \gamma K_c(i,j) \quad (6)$$

Here, (γ) controls the strength with which an empty measurement updates the local belief. This update reduces the expected occupancy in the measured region and nearby cells while preserving a distance-dependent spatial weighting.

For the standard active-search step, each candidate cell is assigned a utility score

$$S_i = \sum_c D_c (\mu_{c,i} + \kappa_c \sigma_{c,i}) - \lambda d_i \quad (7)$$

where (κ_c) controls the relative weight of uncertainty for class (c), (d_i) is the distance from the current AFM position to candidate cell (i), and (λ) weights the cost associated with stage displacement. The score therefore balances three contributions: the expected presence of target objects, the value of exploring uncertain regions, and the practical preference for shorter movements.

During forced exploration, the algorithm instead uses an uncertainty-driven score,

$$U_i = \sum_c D_c \sigma_{c,i} \quad (8)$$

This score prioritizes cells that remain uncertain for classes whose requested target counts have not yet been reached. It therefore encourages exploration of poorly characterized regions of the sample, especially when the current search has not yet collected sufficient examples of one or more classes.

The next cell is selected according to

$$i^* = \begin{cases} \arg \max_i U_i, & r < \varepsilon_{unknown} \\ \arg \max_i S_i, & r \geq \varepsilon_{unknown} \end{cases} \quad (9)$$

where ($r \sim \mathcal{U}(0,1)$), and ($\varepsilon_{unknown}$) is the probability of selecting the uncertainty-driven exploration mode. Setting ($\varepsilon_{unknown} = 0$) recovers the baseline active-search strategy based only on (S_i).

Candidate cells are also subject to a reachability constraint,

$$d_i \leq d_{max} \quad (10)$$

so that cells farther than the maximum allowed step distance from the current AFM position are excluded from the next-scan decision.

The autonomous search terminates once all user-defined class targets have been reached,

$$F_c \geq T_c \quad \forall c \quad (11)$$

At this point, the acquisition has collected the requested number of objects for every class and no additional scan locations are selected.

If a newly acquired image contains no objects of a given class, the algorithm increases beta for that class in the measured cell and in nearby cells according to the spatial kernel. This decreases the estimated probability μ in that region. As more empty measurements accumulate, the local score S_i decreases and the algorithm is pushed away from unproductive regions.

The algorithm still explores because uncertainty remains high in unmeasured cells. In the baseline strategy, uncertainty enters the normal score through $\kappa_c \sigma_{c,i}$. In the unknown-exploration strategy, a random draw with probability $\varepsilon_{unknown}$ selects the most uncertain reachable cell according to U_i . Therefore, if the initial central region is empty, the search expands outward through reachable cells while respecting $d_i \leq d_{max}$.

S2.3: Simulation framework and evaluation

We next evaluated the adaptive exploration strategy using simulated two-class spatial distributions. The simulations were performed on a (20×20) candidate grid, where each grid cell represents a possible AFM scan location.

Figure S2.1 shows a representative single-case simulation. In this example, two object classes were generated from spatially clustered distributions. Class A was assigned a higher overall density than class B, reproducing a situation in which one target population is more abundant than the other. All active-search strategies were initialized at the center of the grid and first sampled five additional cells near the center before entering the adaptive decision phase. The raster strategy, in contrast, followed a serpentine trajectory across the grid until the requested target count for both classes was reached.

To quantify the robustness of the approach beyond a single example, we then performed a Monte Carlo comparison across 100 randomly generated two-class spatial distributions, summarized in Figure S2.2. For each simulated distribution, both classes contained at least the requested number of objects, ensuring that the targets were attainable for all methods without

imposing an artificial maximum number of scans. The requested targets were ($T_A = 95$) for class A and ($T_B = 24$) for class B.

The different exploration strategies were compared using four complementary metrics: the number of scans required to reach the target, the fraction of the grid that had to be sampled, the total movement of the stage expressed in grid-cell units, and the composition of the search focus, defined by the relative contribution of each class to the selected scan locations.

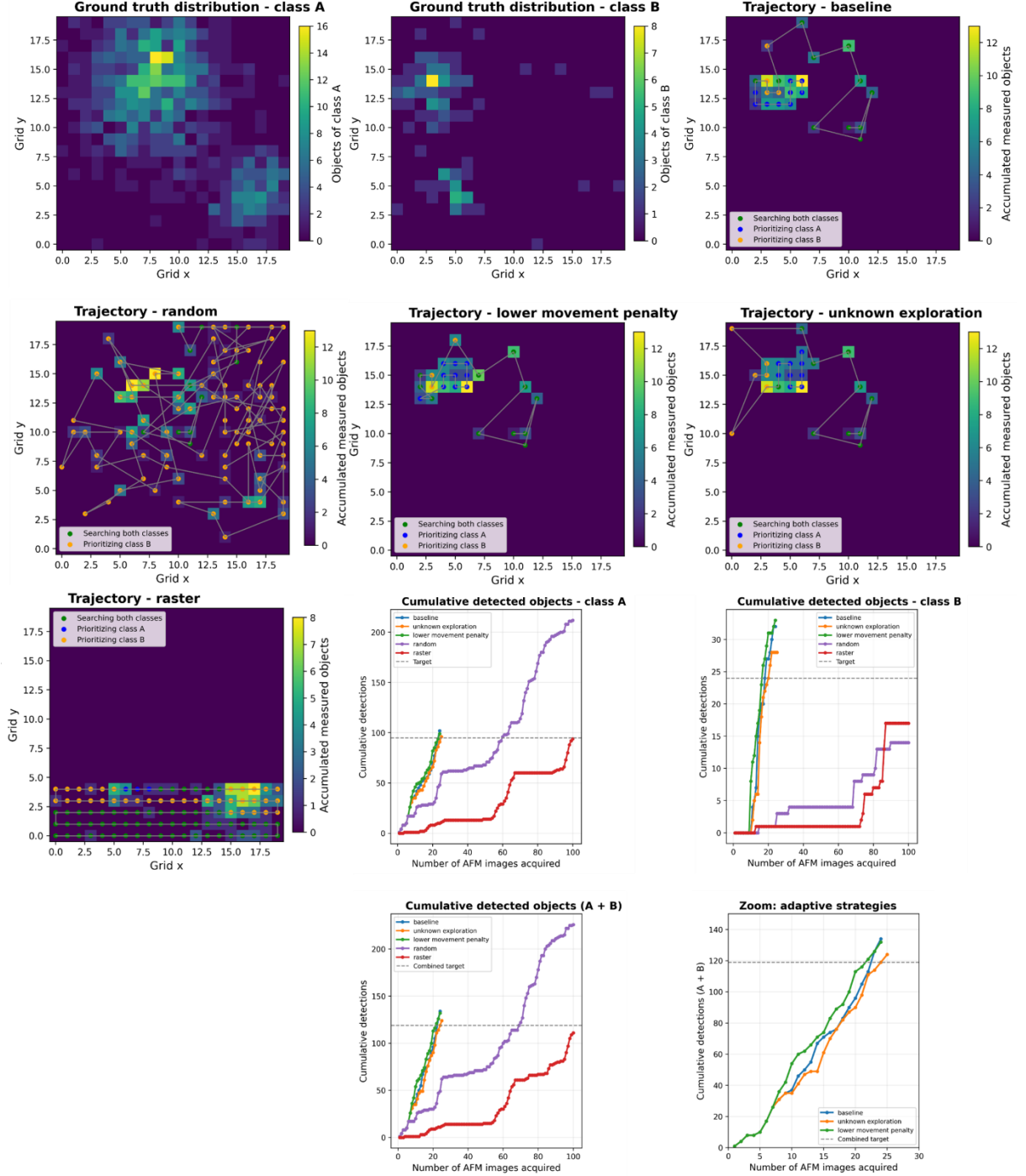


Figure S2.1. Single-distribution example of the class-aware active-search workflow.

Ground-truth object distributions for classes A and B are shown alongside the acquisition trajectories for baseline active search, random sampling, active search with a lower movement penalty, active search with unknown-region exploration, and raster scanning. The trajectory panels display the accumulated measured object map as the background. Point color indicates the search priority assigned to each scan: green indicates that both classes contributed similarly to the utility, blue indicates prioritization of class A, and orange indicates prioritization of class B. Random and raster sampling are included as reference strategies. The cumulative detection plots show the number of objects detected for class A, class B, and both classes combined as a function of the number of AFM images acquired, with an additional magnified view comparing the three adaptive strategies. Dashed horizontal lines indicate the class-specific target counts (95 for class A and 24 for class B) and their combined total (119). Acquisition is terminated when both class-specific targets are reached.

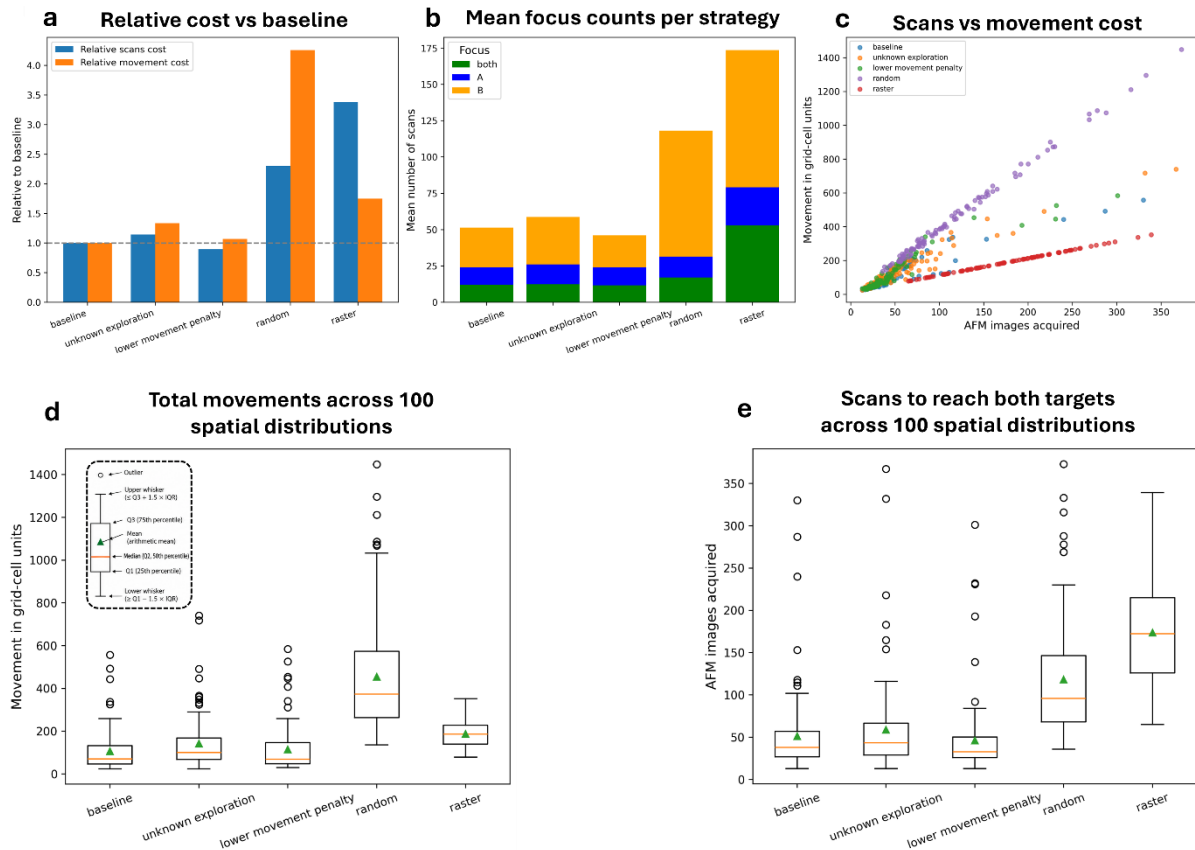


Figure S2.2. Monte Carlo comparison across 100 randomly generated two-class spatial distributions. All methods were run until both class-specific targets were reached. **(a)** Relative scan and movement costs normalized to the baseline active-search strategy; values below one

indicate lower costs than baseline, whereas values above one indicate higher costs. **(b)** Mean search-focus composition, showing the contributions of class A, class B, and both classes to acquisition decisions. **(c)** Number of scans versus total stage movement for each method and simulated distribution. **(d)** Distribution of total stage movement for each acquisition strategy. **(e)** Number of AFM images required to reach both class-specific targets.

The simulations were performed on a (20×20) grid of candidate AFM image locations. Each simulated sample contained two object classes, denoted as class A and class B. The requested target counts were ($T_A = 95$) objects for class A and ($T_B = 24$) objects for class B.

The acquisition was initialized by first sampling the center cell of the grid, followed by five additional randomly selected cells within a near-center radius of (4.5) grid-cell units. After this initialization step, the active strategies selected new scan locations according to their corresponding utility functions.

For the active-search strategies, the maximum allowed displacement between consecutive scan locations was set to ($d_{\max} = 6.0$) grid-cell units. Two movement-weighting conditions were evaluated: a baseline condition with ($\lambda = 0.018$), and a lower movement-weight condition with ($\lambda = 0.005$). The unknown-exploration strategy used an exploration probability of ($\epsilon_{\text{unknown}} = 0.10$), whereas the baseline active-search strategy was obtained by setting ($\epsilon_{\text{unknown}} = 0$).

The class-dependent spatial correlation lengths used in the Beta occupancy model were ($l_A = 2.0$) cells for class A and ($l_B = 1.6$) cells for class B. The uncertainty weights were set to ($\kappa_A = 0.55$) and ($\kappa_B = 0.70$), giving a slightly larger weight to uncertainty in the less abundant class. The negative update strength was set to ($\gamma = 1.4$).

Each simulation was run until both class targets were reached, that is, until ($F_A \geq T_A$) and ($F_B \geq T_B$). No artificial maximum number of scans was imposed in the Monte Carlo analysis.

Table S2: Performance summary across 100 randomly generated distributions.

Lower values are better for scans to target, grid sampled, and movement. All methods reached both class targets because no maximum-scan stopping criterion was imposed in this comparison.

Method	Scans mean $\pm$ SD (#)	Scans median [Q1–Q3] (#)	Grid sampled (%)	Movement mean $\pm$ SD (Grid-cell)	Focus both / A / B (#)
<i>Baseline active search</i>	51 $\pm$ 49	38.0 [27.0–57.0]	12.8	107 $\pm$ 97	12.1 / 12.0 / 27.3
<i>Active search + 10% unknown exploration</i>	59 $\pm$ 55	43.5 [29.0–66.5]	14.7	143 $\pm$ 124	12.3 / 13.7 / 32.7
<i>Active search + lower movement penalty</i>	46 $\pm$ 45	33.0 [26.0–50.2]	11.5	114 $\pm$ 109	11.7 / 12.3 / 22.1
<i>Random sparse sampling</i>	118 $\pm$ 70	96.0 [68.0–146.5]	29.5	455 $\pm$ 273	17.1 / 14.5 / 86.5
<i>Raster scan</i>	174 $\pm$ 63	172.5 [126.0–215.0]	43.4	187 $\pm$ 63	53.0 / 26.1 / 94.5

Table S2. Performance summary across 100 randomly generated spatial distributions.

Lower values indicate better performance for scans to target, grid sampled, and total movement. Scan counts represent the number of acquired AFM images, while movement is expressed as the cumulative Euclidean distance between scan positions in grid-cell units. Focus both / A / B reports the mean number of acquisitions per simulation classified according to the relative contribution of both classes, class A, or class B. The six initialization scans in the active-search strategies are assigned to the “both” category. For random and raster sampling, focus categories are descriptive rather than acquisition-selection criteria. Small differences between the sum of focus counts and the reported mean scan count arise from rounding. All methods reached both class-specific targets because no maximum-scan stopping criterion was imposed.

S3. Layout and class-specific spatial distributions of the nanostructure test sample.

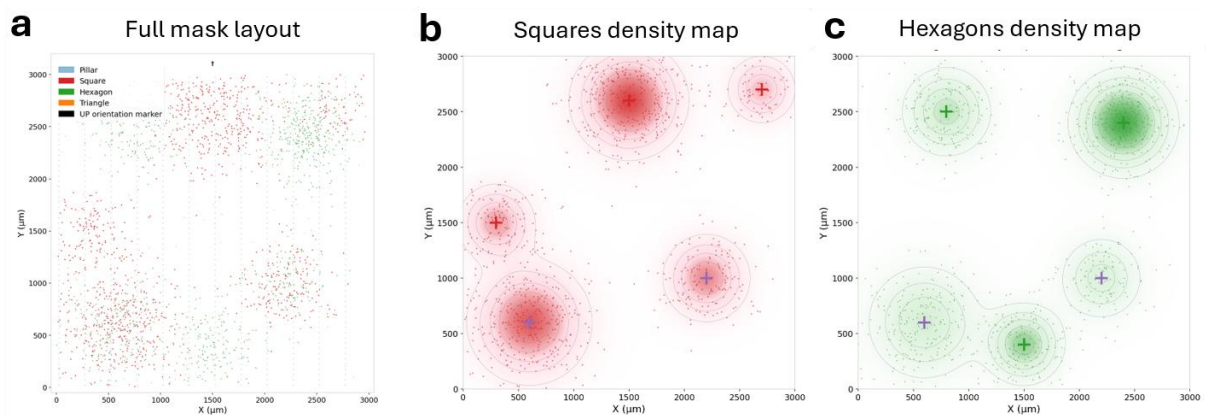


Figure S3. Layout and class-specific spatial distributions of the nanostructure test sample.

a, Full $3 \times 3\text{mm}^2$ mask layout used to evaluate the class-specific probabilistic active-search method. The layout contains 3,600 periodically spaced ellipse nanopillars arranged on a $50\text{ }\mu\text{m}$ -pitch square lattice, together with 1,000 squares, 1,000 hexagons, and 100 triangles. All nanostructures have characteristic lateral dimensions of approximately $4\text{ }\mu\text{m}$. Squares and hexagons were distributed according to predefined class-specific spatial density maps, whereas triangles were placed with a uniform random distribution. **b**, Spatial density map used to place the square nanostructures. **c**, Spatial density map used to place the hexagonal nanostructures. Both class-specific maps consist of multiple two-dimensional Gaussian regions, including shared high-probability areas and regions preferentially enriched in only one of the two classes.

Bibliography

- 1 Austin, J., Johnson, D. D., Ho, J., Tarlow, D. & Van Den Berg, R. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems* **34**, 17981-17993 (2021).
- 2 Chen, T., Wang, C. & Shan, H. in *International conference on medical image computing and computer-assisted intervention*. 491-501 (Springer).
- 3 Luccioni, S., Jernite, Y. & Strubell, E. in *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*. 85-99.
- 4 Bertazzini, G., Albisani, C., Baracchi, D., Shullani, D. & Verdecchia, R. The hidden cost of an image: quantifying the energy consumption of AI image generation. *arXiv preprint arXiv:2506.17016* (2025).
- 5 Chowdhury, J. *et al.* Quality-controlled active learning via Gaussian processes for robust structure–property learning in autonomous microscopy. *npj Computational Materials* (2026).
- 6 Ziatdinov, M., Liu, Y., Kelley, K., Vasudevan, R. & Kalinin, S. V. Bayesian active learning for scanning probe microscopy: From Gaussian processes to hypothesis learning. *ACS nano* **16**, 13492-13512 (2022).
- 7 Vasudevan, R. K. *et al.* Autonomous experiments in scanning probe microscopy and spectroscopy: choosing where to explore polarization dynamics in ferroelectrics. *ACS nano* **15**, 11253-11262 (2021).
- 8 Kusne, A. G. *et al.* On-the-fly closed-loop materials discovery via Bayesian active learning. *Nature communications* **11**, 5966 (2020).
- 9 Liu, Y., Ziatdinov, M. A., Vasudevan, R. K. & Kalinin, S. V. Explainability and human intervention in autonomous scanning probe microscopy. *Patterns* **4** (2023).