

# ComputeCosts observatory report 003

## Workstation scale computing in contemporary scientific research

Mateo Mayer

22 February 2026

### **Abstract**

This report examines the role of local versus cloud computing in scientific computing as practiced in chemical engineering, environmental engineering, biotechnology and process technology, with specific attention to scientific machine learning, modelling, simulation and inference workloads typical of PhD level research. The analysis is based on a meta analysis of publications from 2025 and 2026 extracted from arXiv and associated GitHub repositories, complemented with peer reviewed literature in computational chemistry, engineering simulation and related fields.

Across this evidence base, a substantial body of papers explicitly reports experiments executed on single workstation systems equipped with consumer RTX GPUs and roughly 128 GB of host memory. The frequency of these disclosures is of the same order of magnitude as references to cloud based computation. This is notable because the broader technology narrative strongly promotes cloud infrastructure, largely through large scale commercial messaging, while the research literature shows that modern local compute remains an efficient, routine and accepted experimental platform.

Taken together, the evidence indicates that for a large majority of day to day scientific computing tasks in these domains, a modern local workstation, for example a system with an RTX 5090 class GPU, about 128 GB RAM, multi terabyte NVMe storage and a contemporary multi core desktop processor, is already sufficient to execute modelling, training, inference and simulation workflows.

Cloud and cluster resources remain important for genuinely large scale computations. However, the literature suggests a common development pattern: models are often built and stabilised on local machines first, with larger infrastructure used later when additional scale is needed. This reflects the iterative nature of research, where working locally helps ensure that methods behave correctly before larger resources are applied.

### **Data sources, scope, and interpretation rules**

This report combines two evidence streams.

First, a deterministic snapshot of a research database maintained by [ComputeCosts.nl](https://www.compute-costs.nl) was scanned for infrastructure vocabulary and explicit hardware disclosures. The archive contains 6690 unique arXiv papers collected between 1 January 2025 and 19 February 2026. The archive is

produced through an ingestion pipeline using predefined keyword queries related to scientific computing, machine learning, simulation and computing infrastructure.[1] As a result, the dataset represents a deterministic observatory sample designed to surface papers in which compute infrastructure is likely to be mentioned rather than the entire arXiv corpus. Within this archive, titles and abstracts were normalised and scanned using predefined phrase groups representing cloud computing, local execution terminology and explicit hardware disclosures.

In addition, the observatory maintains a deterministic Wikipedia infrastructure archive covering the same period.[2] Page and concept level activity around infrastructure related topics such as cloud computing and local compute environments was recorded to provide an independent signal about the relative visibility of these concepts in the surrounding technical knowledge ecosystem.

Second, a targeted inspection of domain literature was performed across chemical engineering, environmental engineering, biotechnology and analytical chemistry, including computational chemistry and engineering simulation work that is routinely used in these domains. The search focused on papers that explicitly disclose the compute hardware used in the experiments, in particular the presence of RTX class GPUs in workstation environments. The inclusion criterion was strict: the publication must explicitly report this compute environment used to execute the work.

The objective is not to estimate adoption rates. The objective is operational and falsifiable: to verify whether workstation class RTX based systems appear as explicitly reported experimental substrates in published research, and to connect those disclosures to a workflow pattern in which local iteration is dominant and external infrastructure is used selectively.

Interpretation is asymmetric. An explicit disclosure of a named RTX GPU and stated host memory is treated as direct evidence of a publishable and used workstation configuration. Aggregate abstract level counts are treated as conservative lower bound signals because most papers omit infrastructure details. Absence of a disclosure is not interpreted as absence of local compute.

The hardware envelope used in the argument is intentionally practical: an RTX class GPU, approximately 128 GB RAM, multi terabyte NVMe storage, and a contemporary multi core desktop processor. This envelope is treated as a rational default only to the extent that it matches published practice and supports the full daily research loop on one deterministic machine.

## **Evidence of RTX workstations in contemporary scientific computing**

The question addressed here is straightforward: are consumer RTX workstation systems actually used and reported in contemporary scientific research. The literature shows that they are.

A targeted inspection of recent publications in chemical engineering[3][4][5][6][12][20][22][23], biotechnology[13][14][7][9][16][17][18][21] and analytical chemistry[15][8][10][11][19] reveals the structural appearance of papers that explicitly report experiments executed on single workstation systems equipped with NVIDIA RTX GPUs. In these papers the hardware platform is described as part of the experimental setup rather than as a subject of discussion. The disclosures therefore provide direct evidence of the infrastructure used to perform the work.

Across these domains the reported systems consistently fall within the same envelope: a single workstation equipped with an RTX class GPU and substantial host memory. These disclosures

appear in simulation studies, molecular modelling workflows, screening pipelines and machine learning based analysis used in engineering and chemical research.

The pattern is not limited to these domains. Independent machine learning and computational science papers also report experiments executed on RTX based workstations, including systems with roughly 128 GB of RAM.[24][25][26][27][28][29] [30][31][32][33][34] This indicates that such configurations are broadly accepted as practical research environments rather than exceptional setups.

An independent signal is provided by the ComputeCosts.nl arXiv observatory snapshot. Within 6690 papers collected between 1 January 2025 and 19 February 2026, references to cloud computing appear in 253 abstracts, while explicit references to RTX 4090 appear in 59 abstracts, with a similar number referring to RTX GPUs more generally.[1] Because abstracts rarely describe full hardware environments, these values are interpreted as conservative lower bound observations.

Taken together, the evidence shows that workstation systems equipped with RTX GPUs are routinely used and reported in contemporary research. The configuration used throughout this report, a workstation with an RTX class GPU and approximately 128 GB of host memory, therefore reflects hardware that is already generally applied in published scientific practice. This is notable because the broader technology narrative strongly promotes cloud infrastructure, largely through large scale commercial messaging, while the research literature shows that modern local compute remains an efficient, routine and accepted experimental platform.

## **Workstation scale computing in scientific and engineering research**

Scientific computing in chemical, environmental, biotechnological and process engineering research is dominated by iterative research workflows rather than by peak theoretical throughput. The relevant unit is the research loop repeated throughout a project: write or modify a model, prepare data, run the experiment, inspect results, adjust, and rerun.

Typical PhD level workloads in these domains include iterative modelling, numerical simulation, parameter estimation, optimisation, surrogate modelling and machine learning assisted analysis. These tasks can be computationally demanding, but during the dominant development and calibration phase they rarely require distributed infrastructure. The practical requirement is simply that the entire workflow fits into one stable environment with sufficient GPU capability and host memory to keep working sets resident.

The published disclosures in the previous section operationalise this requirement. Across chemistry computation, molecular simulation and engineering simulation, multiple papers report single node execution on consumer RTX GPUs with host memory around 128 GB. This corresponds closely to a workstation envelope consisting of an RTX class GPU, approximately 128 GB RAM, multi terabyte NVMe storage and a contemporary multi core desktop processor.

The reason this envelope matters is operational friction. Local execution starts immediately, the environment remains under direct control, data paths are local, and instrumentation and debugging are continuous. These properties increase the number of completed experimental cycles per day, which is the dominant driver of progress during method development.

Host memory around 128 GB forms a practical threshold because it accommodates larger working sets without constant pruning, swapping or staging. It also allows concurrent workloads typical for these domains, such as simulation with preprocessing and postprocessing, training with plotting, or caching with inference. Multi terabyte NVMe storage complements this by keeping checkpoints, datasets, intermediate artefacts and reproducibility assets locally available.

GPU capability matters even when a project is not framed as deep learning. Surrogate models, differentiable pipelines, inverse problems, optimisation loops and computational chemistry kernels frequently use GPU acceleration through existing toolchains. This is directly reflected in published RTX disclosures across chemistry and engineering. The references [3-34] mentioned in the previous section underline this.

This evaluation results in a limited but strong conclusion: A workstation with an RTX class GPU and roughly 128 GB of host memory is sufficient for the majority of day to day work because it supports the dominant iteration and calibration loop locally, consistent with repeatedly reported practice in the literature.

## **When cloud or clusters accelerate, and why they remain secondary**

External compute resources remain important in modern scientific computing. Certain workloads exceed the limits of a single workstation, including distributed memory multiphysics at very high resolution, extremely large ensemble studies, or training and sampling regimes that surpass the memory or compute envelope of a single GPU node.

However, in such cases the literature shows a consistent pattern in how research workflows evolve. Development, debugging and validation are typically carried out on a single local machine before larger resources are engaged. The dominant phase of research is iterative: models are written, modified, instrumented and repeatedly executed until the behaviour of the system is understood and the implementation is stable.

This pattern is consistent with the evidence base used in this report. Many papers that could have relied on clusters still report single machine RTX workstations for experiments and evaluation, indicating that local execution remains the default environment for substantial parts of modern research, with external resources introduced only when scale clearly benefits the investigation.[35][36][37]

## **Conclusion**

The literature shows that workstation scale computing is not exceptional but routine.

Across the literature examined in this report, numerous publications explicitly report experiments executed on single workstation systems equipped with consumer RTX GPUs and host memory around 128 GB. These disclosures occur across computational chemistry, molecular simulation and engineering simulation workloads relevant to chemical engineering, environmental engineering analytical chemistry and biotechnology.

Taken together with the observatory signals analysed earlier, this evidence indicates that workstation scale systems are not exceptional infrastructure but a routine experimental platform in contemporary scientific computing. This observation contrasts with the broader public narrative

around computing infrastructure, which frequently emphasises large scale cloud platforms, while published research practice continues to make extensive use of local compute environments.

For most research workflows, the dominant phase of work is iterative development, calibration and validation. A local workstation with an RTX class GPU, roughly 128 GB of RAM, multi terabyte NVMe storage and a modern multi core processor is the most efficient platform to support this loop on a single deterministic machine.

Cloud and cluster resources remain essential for genuinely large scale computations. The pattern visible in the literature, however, is consistent: methods are developed and stabilised locally, after which external infrastructure is used selectively and exclusively when additional scale materially advances the investigation.

## References

- [1] ComputeCosts Observatory. arXiv Observatory Epoch 1 Report 2. ComputeCosts.nl. [https://computecosts.nl/arxiv/arxiv\\_epoch1\\_report2.pdf](https://computecosts.nl/arxiv/arxiv_epoch1_report2.pdf).
- [2] ComputeCosts Observatory. Wikipedia Infrastructure Archive Epoch 1. ComputeCosts.nl. [https://computecosts.nl/wikipedia/wikipedia\\_epoch1.pdf](https://computecosts.nl/wikipedia/wikipedia_epoch1.pdf)
- [3] L. Allen, J. Cordiner. Knowledge-enhanced data-driven modeling of wastewater treatment processes for energy consumption prediction. *Computers and Chemical Engineering*, 194 (2025) 108982. <https://doi.org/10.1016/j.compchemeng.2024.108982>.
- [4] R. Stierle, G. Bauer, N. Thiele, B. Bursik, P. Rehner, J. Gross. Classical density functional theory in three dimensions with GPU-accelerated automatic differentiation: Computational performance analysis using the example of adsorption in covalent-organic frameworks. *Chemical Engineering Science*, 298 (2024) 120380. <https://doi.org/10.1016/j.ces.2024.120380>.
- [5] J. Meng, S. Tang, S. M. Wise, Z. Guo. A GPU-Accelerated Matrix-Free FAS Multigrid Solver for Navier–Stokes Equations with Memory-Efficient Implementations. arXiv:2510.11152, 2025. <https://arxiv.org/abs/2510.11152>.
- [6] J. Ock, R. S. Meda, S. Badrinarayanan, N. S. Aluru, A. Chandrasekhar, A. B. Farimani. Large Language Model Agent for Modular Task Execution in Drug Discovery. arXiv:2507.02925, 2025. <https://arxiv.org/abs/2507.02925>.
- [7] J. Hu, D. Hou, J. Jiang. Physical embedding machine learning force fields for organic systems. 2025. <https://arxiv.org/abs/2509.10270>.
- [8] Z. Chen, R. Zhou and P. Ren. Spectraformer: deep learning model for grain spectral qualitative analysis based on transformer structure. *RSC Advances*, 2024, 14, 8053–8066. <https://doi.org/10.1039/d3ra07708j>.
- [9] J. Ha, G. Hanrot, T. Noh, J. H. Cheon, J. W. Kim, D. Stehlé. Private Iris Recognition with High-Performance FHE. arXiv:2601.17561, 2026. <https://arxiv.org/abs/2601.17561>.
- [10] B. Petracchi, E. Torti, E. Marenzi, F. Leporati. Acceleration of Hyperspectral Skin Cancer Image Classification through Parallel Machine-Learning Methods. *Sensors*, 24, 1399, 2024. <https://doi.org/10.3390/s24051399>.
- [11] A. Muminov, A. Pham. CytoDINO: Risk-Aware and Biologically-Informed Adaptation of DINOv3 for Bone Marrow Cytomorphology. arXiv:2512.17930, 2025. <https://arxiv.org/abs/2512.17930>.
- [12] A. Sánchez-Rodríguez, R. D. Ullrich-Estrella, C. E. González-Gallardo, M. B. Jácome-Villacres, G. García-Vidal, R. Pérez-Campdesuñer. Deep Learning-Based Digital, Hyperspectral, and Near-Infrared (NIR) Imaging for Process-Level Quality Control in Ecuador’s Agri-Food Industry: An ISO-Aligned Framework. *Processes*, 13, 3544, 2025. <https://doi.org/10.3390/pr13113544>.

- [13] S. P. Srivastava, R. Gorantla, S. K. Chundru, C. J. R. Winkelman, A. S. J. S. Mey, R. K. Singh. Explainable Active Learning Framework for Ligand Binding Affinity Prediction. *bioRxiv*, 2025. <https://doi.org/10.64898/2025.12.17.694851>.
- [14] Z. Xiao, H. B. Pakrasi, Y. Chen, Y. J. Tang. Network for Knowledge Organization (NEKO): An AI knowledge mining workflow for synthetic biology research. *Metabolic Engineering*, 87 (2025) 60–67. <https://doi.org/10.1016/j.ymben.2024.11.006>.
- [15] D. E. Kleiman, J. Feng, Z. Xue, D. Shukla. ESMDynamic: Fast and Accurate Prediction of Protein Dynamic Contact Maps from Single Sequences. *bioRxiv*, 2025. <https://doi.org/10.1101/2025.08.20.671365>.
- [16] C. Lu, K. C. Weber, S. K. McBride, A. Reckers, A. Glasgow. A machine learning method for calculating highly localized protein stabilities. *bioRxiv*, 2025. <https://doi.org/10.1101/2025.10.21.683809>.
- [17] D. Kim, J. Kim, GPU-accelerated non-dominated sorting genetic algorithm III for maximizing protein production, *Electronic Research Archive*, 32(4): 2514–2540, 2024, <https://doi.org/10.3934/era.2024116>.
- [18] J. Zhao, Q. Zhou, T. Wang, Y. Fan, Q. Yang, L. Jiao, C. Liu, Z. Guo, Q. Lu, F. Zhou, R. Zhang, MolProphecy: Bridging medicinal chemists' knowledge and molecular pre-trained models via a multi-modal framework, *Journal of Advanced Research*, 2025, <https://doi.org/10.1016/j.jare.2025.10.035>.
- [19] S. Huang, Y. Jin, W. Jin, Y. Mu, Analytical-Chemistry-Informed Transformer for Infrared Spectra Modeling, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-25)*, 2025, <https://ojs.aaai.org/index.php/AAAI/article/view/33917>.
- [20] D. Quevedo, K. Do, G. Delic, J. Rodríguez-Borbón, B. M. Wong, C. E. Ivey, GPU Implementation of a Gas-Phase Chemistry Solver in the CMAQ Chemical Transport Model, *ACS ES&T Air*, 2, 226–235, 2025, <https://doi.org/10.1021/acsestair.4c00181>.
- [21] R. S. M. Freitas, Z. Xing, F. A. Rochinha, R. F. Cracknell, D. Mira, N. Karimi, X. Jiang, Pathways to sustainable fuel design from a probabilistic deep learning perspective, *Advances in Applied Energy*, 19, 100226, 2025, <https://doi.org/10.1016/j.adapen.2025.100226>.
- [22] F. J. W. A. Martins, C. T. Foo, A. Unterberger, S. Karaminejad, T. Endres, K. Mohri, Analyzing 3D fields of refractive index, emission and temperature in spray-flame nanoparticle synthesis via tomographic imaging using multi-simultaneous measurements (TIMes), *Applications in Energy and Combustion Science*, 16, 100213, 2023, <https://doi.org/10.1016/j.jaecs.2023.100213>.
- [23] H. Zhu, C. He, G. Li, X. Liu, Y. Jiao, CFD-DEM simulation on flow behavior and heat transfer characteristics in a zigzag moving bed for soil disinfection, *Case Studies in Thermal Engineering*, 72, 106248, 2025, <https://doi.org/10.1016/j.csite.2025.106248>.

- [24] VDB: A Deep-Learning Framework for Sparse, Large-Scale Volumetric Data. arXiv:2407.01781, 2025. <https://arxiv.org/html/2407.01781v2>.
- [25] Scaling Behaviors of Evolutionary Algorithms on GPUs, RTX 3090 disclosure. arXiv:2601.18446, 2026. <https://arxiv.org/html/2601.18446v1>.
- [26] C. C. Wang et al. QueenV2: Future of Quantum Circuit Simulation. arXiv:2409.14697, 2024. <https://www.arxiv.org/pdf/2409.14697>.
- [27] Craftax: A Lightning-Fast Benchmark for Open-Ended Reinforcement Learning. arXiv:2402.16801, 2024. <https://arxiv.org/html/2402.16801v2>.
- [28] NoiseHGNN: Synthesized Similarity Graph-Based Neural Network. arXiv:2412.18267, 2024. <https://arxiv.org/html/2412.18267v1>.
- [29] E. Adabag et al. Differentiable Model Predictive Control on the GPU, RTX 4090 disclosure. arXiv:2510.06179, 2025. <https://arxiv.org/html/2510.06179v1>.
- [30] A GPU Implementation of Multi-Guiding Spark Fireworks Algorithm, RTX 3090 disclosure. arXiv:2501.03944, 2025. <https://arxiv.org/html/2501.03944v1>.
- [31] S. Li et al. Zero-Shot Artifact2Artifact: Self-incentive artifact removal, RTX 3090 Ti disclosure. *Photoacoustics*, 2025. <https://www.sciencedirect.com/science/article/pii/S2213597925000461>.
- [32] Improving Policy Exploitation in Online Reinforcement Learning, RTX 3090 Ti disclosure. arXiv:2601.19720, 2026. <https://arxiv.org/html/2601.19720v2>.
- [33] W. Kang et al. A Digital Twin of River Experiment Infrastructure, RTX 3090 disclosure. *Applied Sciences*, 2025. <https://www.mdpi.com/2076-3417/15/23/12507>.
- [34] E. Amaro et al. Mechanical Response of FeNiCrCoAl High-Entropy Alloys, MD simulations on RTX GPUs including RTX 4090 disclosure. *Nanomaterials*, 15(9), 652, 2025. <https://www.mdpi.com/2079-4991/15/9/652>.
- [35] W. Brewer, A. Gainaru, F. Suter, F. Wang, M. Emani, S. Jha. AI-coupled HPC Workflow Applications, Middleware and Performance. arXiv:2406.14315, 2025. <https://arxiv.org/abs/2406.14315>.
- [36] D. Medeiros, G. Schieffer, J. Wahlgren, I. Peng. A GPU-accelerated Molecular Docking Workflow with Kubernetes and Apache Airflow. arXiv:2410.10634, 2024. <https://arxiv.org/abs/2410.10634>.
- [37] F. Le Piane, M. Voza, M. Baldoni, F. Mercuri. Integrating high-performance computing, machine learning, data management workflows, and infrastructures for multiscale simulations and nanomaterials technologies. *Beilstein Journal of Nanotechnology*, 15, 1498–1521, 2024. <https://doi.org/10.3762/bjnano.15.119>.