Charting a Path in a Shifting Technical and Geopolitical Landscape: Post-Exascale Computing for the National Nuclear Security Administration (2023)

Berkeley authors

- Kathy Yelick
- David Culler
- John Bell
- Conferees for faculty development: David Patterson, Michael Jordan, John DeNiro, Ani Adhikari, Brad DeLong, Kenichi Soga, Eric van Dusen, Bob Reich, Laura Tyson, Rich Lyons, Shankar Sastry, Josh Bloom, ...

“**The basic idea would be to replace some computationally expensive component of a simulation with a relatively inexpensive data-driven model** based on either experiment or data computed from a separate simulation. Although potentially expensive to train, **the resulting model could dramatically reduce simulation costs** as discussed in the artificial intelligence (AI) R&D section later in this chapter” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 46)

“In these situations, **AI methods may be used in concert with traditional simulations.** For example, neural networks can be trained on data from simulations to produce surrogates to computational functions (or even entire simulations), **achieving nonlinear improvements of multiple orders of magnitude in time-to-solution for HPC applications**.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 52)

“As this example illustrates, **AI methods can augment conventional computational simulation, enabling new approaches to old problems and providing paths to tackling previously intractable problems**. From a workload perspective, both high-fidelity simulations and AI methods must be supported.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 52)

“**we use the term ‘AI for Science’ to broadly represent the next generation of methods and scientific opportunities** in computing, including the development and application of AI methods (e.g., machine learning, deep learning, statistical methods, data analytics, automated control, and related areas) **to build models from data and to use these models alone or in conjunction with simulation and scalable computing to advance scientific research.**”2” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 53)

“**AI as surrogates for simulations of physical systems**, ranging from practical problems such as optimized representations of neutron group structures, as already demonstrated by the laboratories, to building emulators for the 3D evolution of engineered systems—a capability that is not realizable today and **is estimated to require hundreds of exaflop years to train with current methods.**” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 53)

“AI in the loop for **automated adaptive real-time control** of experimental facilities and manufacturing processes **that currently do not admit to real-time control** or that require increasingly scarce human expertise owing to an aging workforce.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 53)

“**AI to connect across the lifecycle** for end-to-end intelligent decisions, **such as designing to explicitly account for manufacturing and aging**, rather than suboptimal proxies for manufacturing/aging issues.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 53)

“AI to accelerate all stages of the complex physics cycle, encompassing hypothesis, design, execution/control, diagnosis, and analysis. For example, **AI-based surrogates may be used to screen large collections of candidates (e.g., materials, structures) that cannot reasonably be evaluated via conventional simulation**.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 54)

“**AI for managing surveillance via automated analysis of multimodal data, leveraging, for example, self-supervised learning methods to detect unusual events**.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 54)

“**AI for enabling digital twins that integrate heterogeneous models and data from multiple sources, while leveraging edge computing and integration across edge/HPC**” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 54)

“A crosscutting theme is the use of AI methods **to learn previously unknown relationships among entities (“serendipitous models”12) that can be evaluated far faster than by conventional means and/or without explicit programming—in the process, automating and accelerating previously manual steps to enable more rapid exploration of far larger design spaces**.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 54)

“RECOMMENDATION 2.3: NNSA should expand research in AI to explore the use of these methods both for predictive science and for emerging applications, such as manufacturing and control of experiments, and **develop machine learning techniques that provide the confidence in results** required for NNSA applications.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 55)

Demand for high performance computing, networking, algorithms, and mathematics is high across industries. **A strong entry level graduate in computer science will regularly receive a compensation package of salary, bonus, and stock from a hyperscale internet company exceeding $300,000 per year; a principal engineer or research scientist’s compensation can go to seven figures.** These compensation rates are many times the amount this talent can earn at the national labs. **The demand is driven by the impact that high performance computing has on the profitability and capabilities of these companies, in serving insatiable clickflows, data analytics, and particularly AI/ML computing demands.**

“In addition to normal hiring pathways, the laboratories should **work to find and attract exceptional minds in nontraditional venues (e.g., review committee members, professors, industrial experts, highschool students) and bring them in with the conscious intent to engage them in mission problems**—whether for a week, a summer, a sabbatical, an apprenticeship, an expert’s camp, or a career. In all of these areas, the committee encourages the NNSA laboratories not to work individually or in competition, but rather to formulate strategic cross-laboratory and interagency efforts to collect data, coordinate recruiting, and work with the dimension of diversity and inclusion as ways to improve the workforce.” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 63)

“RECOMMENDATION 3.2: NNSA should also **develop a deliberate strategy to attract an international workforce and to provide them with a welcoming environment while thoughtfully managing the attendant national security risks.**” (Committee on Post-Exascale Computing for the National Nuclear Security Administration et al., 2023, p. 64)