ASA Response to Department of Energy's Request for Information on Mobilizing Talent for the Genesis Mission and Developing an American Workforce to Advance Artificial Intelligence for Science and Engineering

March 4, 2026

Prepared with the guidance of the ASA Scientific and Public Affairs Advisory Committee

The American Statistical Association appreciates the opportunity to comment on the Department of Energy: Request for Information on Mobilizing Talent for the Genesis Mission and Developing an American Workforce to Advance Artificial Intelligence for Science and Engineering. In particular, the RFI seeks input on strategies for meeting the technical challenges of the Genesis Mission and on developing a skilled American workforce to advance AI in science and engineering. Further, based on the input, the DOE wants to form highly collaborative research teams—partnerships of educational institutions, DOE, private industry, and philanthropic organizations—who would "establish new curricula and experiences for training bachelor's and master's level students, in AI for Science and Engineering and the future AI workforce, as well as research experiences as part of doctoral and post-doctoral associate training."

To achieve the full potential of these exciting aims, it is imperative that statisticians are fully engaged in all aspects of the Genesis Mission and that statistics is a fundamental part of the curricula, experiences, training, and professional development. As the science of learning from

data, statistics permeates nearly all aspects of AI, including in the following four ways: (i) ensuring data quality for reliable AI performance; (ii) accurate modeling; (iii) enhancing AI efficiency; and (iv) addressing AI accuracy and generalizability. Statisticians are regularly engaged in the objective evaluation of novel products including those with an AI component. They work in teams that can design studies to address concerns over novel AI products, establishing that outputs are reproducible as well as accurate. Statisticians are also experts in experimental design, which is critical to the testing of AI output and products. Research teams that include statisticians as well as a workforce fully versed in statistical concepts is therefore critical to the maximum success of the Genesis Mission.

We therefore urge that DOE require that

1. the research teams to be formed and other aspects of the work include statisticians and data scientists; and
2. statistics be a core component of the curricula and training experiences for the Genesis Mission, and that statisticians are engaged in the curriculum and training-experience development.

The prominence of AI is not solely dependent on algorithmic sophistication, but on a sound scientific foundation for dealing with data's inherent flaws and uncertainties. As the science of learning from data, statistics provides methods for quantifying, propagating, and communicating uncertainty, which is crucial for informed decision-making, especially in regulated environments. Furthermore, even the most advanced models will have uncertainty in its predictions. Statistical innovations are often model-agnostic, meaning they will be just as salient and necessary with tomorrow's AI as they are with today's. By requiring statistical leadership and expertise, DOE will ensure that the Genesis Mission's work is grounded in rigor and equipped for long-term trustworthiness.

Statistics supports a wide range of real-world AI applications: in healthcare, statistical modeling underpins AI systems that detect cancer and predict disease risk by distinguishing meaningful signals from noisy biomedical data; in autonomous vehicles and robotics, statistical methods help evaluate and validate sensor data and decision algorithms to ensure reliable performance under varied conditions; in education technology, statistical literacy is necessary for systems that adapt instruction or detect misinformation from AI-generated content; in economics and public policy, statistical design and bias mitigation inform predictive models used for labor market forecasting, resource allocation, and economic planning; and in national security and energy sectors, rigorous statistical evaluation frameworks are essential for assessing AI models that optimize manufacturing automation, energy distribution, or defense systems. Across these examples, statistical expertise contributes to data quality assessment, calibration of model

uncertainty, bias detection and correction, and the design of performance benchmarks, all of which are crucial components of trustworthy and effective AI deployments.

We now turn to answer select questions from the RFI:

1. How can DOE catalyze research collaborations between DOE National Laboratories, universities, and industry to meet the goals of the Genesis program?

   Statistics serves as a common language for data—one that translates naturally across scientific disciplines, institutional boundaries, and application domains. This positions statisticians as uniquely suited to serve as connectors and bridge-builders among the diverse partners that the Genesis program will need to convene. DOE should therefore make the intentional integration of statistical expertise a cornerstone of its strategy for catalyzing collaboration between National Laboratories, universities, and industry.

   Statisticians already have a long tradition of working across all three of these sectors, and they bring with them not just technical skills but the collaborative habits of mind that genuine interdisciplinary work demands. Because statistics sits at the interface of the underlying application domain and the technical challenges associated with deploying AI models, statisticians are especially well-positioned to help teams from different fields understand each other, learn from each other's approaches, and develop shared methodological frameworks. This kind of cross-pollination is essential for achieving the ambitious, multi-sector goals of the Genesis Mission.

   Beyond facilitating communication, the integration of statistical expertise at every stage of research collaboration is a methodological necessity. Working with the large datasets that AI for science and engineering requires introduces well-known risks—data integrity failures, selection bias, and overfitting—that can silently undermine even the most computationally sophisticated models. Statisticians are trained to identify and address these challenges rigorously. Without their involvement from the outset, collaborations risk producing results that are technically impressive but scientifically unreliable.

   To realize these benefits, DOE should actively encourage and expand existing university-government data-sharing partnerships while establishing new mechanisms to promote national-level research collaboration. This includes supporting joint appointments, shared research infrastructure, and collaborative grant mechanisms that bring statisticians into Genesis-related projects from the ground up—not as consultants after the fact, but as core members of interdisciplinary teams. Done well, this approach will not only strengthen individual projects but help build the durable, cross-sector research community that the Genesis Mission will depend on for long-term success.

2. How can DOE incentivize partnerships between universities, DOE National Laboratories, industry, and philanthropic organizations to 1) establish new training paths for bachelor's and master's degrees focused on dual competencies in AI and scientific/engineering disciplines and 2) provide innovative experiences to prepare doctoral students and post-doctoral associates for careers in AI for science?

   DOE can create new internship opportunities for master's, doctoral students, and new postdoctoral programs. Internship opportunities can be integrated into the graduate training (rather than relegated to a summer-only experience). New post-graduate training programs may partner with academia to provide for a shared DOE-academia mentoring team or to create new positions in which junior faculty participate in DOE programs/projects.

3. What attributes might attract undergraduates to programs that offer dual competency degrees?

   Programs offering dual competency degrees would attract undergraduates by minimizing additional academic burden — since many statistics students already double major with math, CS, or molecular biology, AI competency could be integrated through shared courses satisfying both departments simultaneously. To maximize the domestic talent pipeline, program design should ensure accessibility across all institution types, including community colleges, regional universities, and historically underserved institutions, where a large untapped reservoir of STEM-capable students exists. Embedding AI literacy into existing degree pathways rather than requiring separate credentials ensures the US captures the full breadth of its domestic talent base — a strategic necessity for maintaining global competitiveness in AI development.

4. Beyond funding, what other opportunities could DOE, including its National Laboratories and user facilities, bring to these partnerships?

   Beyond funding, DOE and its National Laboratories offer unique scientific assets that no private sector partner can replicate. Decades of federally funded research have generated large, complex, high-quality datasets in nuclear simulation, energy grid behavior, climate modeling, and materials science — data that is otherwise siloed and underleveraged. National Laboratories such as Oak Ridge, Lawrence Berkeley, and Los Alamos can serve as testbeds where AI and statistical models are trained and validated against real-world scientific problems of national consequence. Their user facilities produce experimental data critical for AI-driven advances in materials discovery, battery

technology, and energy security. DOE's supercomputing infrastructure provides the computational scale necessary for training and validating large statistical and AI models — a resource that complements academic and industry partnerships and strengthens the national AI research ecosystem.

5. In addition to classroom and research training, what other contributions could universities make to support these partnerships?

6. Beyond funding, what types of contributions could industry and philanthropic organizations make to enhance these partnerships and enrich the experiences of the students?

   Industry and philanthropic organizations could provide long term mentorship programs, study and work programs, and federal and state government opportunities in AI policy.

7. In addition to AI, which scientific and engineering disciplines are well-suited for this dual competency training approach?

   Statistics is particularly well-suited as a scientific discipline for dual-competency training with AI because it provides the theoretical and methodological foundations needed to design, implement, evaluate, and monitor AI models. Statistical training equips practitioners to understand data quality, sampling mechanisms, bias, uncertainty, and variability, all of which critically shape how AI models learn and perform in real-world settings. Statistics departments, specifically, are well-positioned to teach core AI concepts to students in other scientific and engineering disciplines, thereby facilitating true interdisciplinary competency. Dual competency in AI and statistics enables researchers and engineers to move beyond algorithm development toward principled model development, rigorous validation, and interpretable results. This combination is essential for guaranteeing reproducibility and trustworthiness in high-stakes domains such as healthcare, energy systems, economics, and national security. This combination also supports reproducibility, fairness assessment, and uncertainty quantification, ensuring that AI systems are not only powerful but scientifically grounded and trustworthy. As AI increasingly influences scientific discovery and decision-making, integrating statistical expertise into AI training strengthens both innovation and responsible use across disciplines.

8. What components are needed for an effective dual competency degree program to best prepare students for successful careers at the bachelor’s level? At the master’s level?

An effective dual competency degree program—one that prepares students to work at the intersection of AI and scientific or engineering domains—must be built on a rigorous quantitative foundation. At both the bachelor's and master's levels, this means ensuring that data analysis, programming, and probability and statistical literacy form the core of the curriculum. These are not simply introductory prerequisites to be dispensed with early; they are the conceptual bedrock that allows students to deploy AI tools responsibly and interpret their outputs correctly. Upon this foundation, degree programs should layer targeted coursework in AI methods, including large language models, so that students understand both the capabilities and the limitations of the tools they will be expected to use in practice.

A critical and often underappreciated component of this foundation is the development of genuine quantitative intuition. Students should graduate not just knowing how to run analyses, but knowing what to look for—how to recognize when a result seems implausible, how to spot anomalies in data, and when to interrogate a model's outputs rather than accept them at face value. This kind of statistical sensibility is distinct from formal technical training, and degree programs should design experiences—through case studies, research projects, and applied problems—that cultivate it deliberately. In developing curricula, programs can and should draw on existing frameworks such as the CSAB recommendations for data science master's programs, which offer a well-developed model for balancing technical rigor with applied competency.

At the bachelor's level, the emphasis should be on building this quantitative and computational core alongside broad exposure to an application domain, equipping graduates to contribute meaningfully to data-intensive teams even as they continue to develop expertise. At the master's level, programs can go further—deepening methodological sophistication, introducing more advanced AI coursework, and placing students in substantive research or applied experiences that require them to integrate statistical and domain knowledge in realistic, complex settings.

Across both levels, programs must also attend seriously to professional skills that are specific to working in AI-augmented environments. As students enter workplaces where large language models and other AI tools are available around the clock, new challenges around time management, task prioritization, and cognitive sustainability become salient. Unlike traditional workflows, AI tools can generate outputs continuously and encourage rapid task-switching, which research suggests can reduce productivity and contribute to burnout.

9. What components are needed to prepare graduate students and post-doctoral students for successful careers at the nexus of AI and science?

   A curriculum that focuses solely on the mechanical explanation of algorithms without this statistical grounding risks producing graduates who can operate AI tools but cannot judge their accuracy, detect statistical bias, or innovate responsibly.

   The core principles of ethical AI—accountability, transparency, and fairness—are fundamentally statistically-informed endeavors, requiring practitioners to possess the knowledge of sampling, measurement error, and confounding necessary to disclose known biases in data collection and assess system limitations.

   Uncertainty Management Training: Incorporate training that directly addresses the challenge of designing education for unknown future job requirements, focusing on statistical tools and concepts for managing uncertainty, risk, and variability in emerging professional roles.

10. How could partnerships between DOE, universities, and industry provide new training opportunities for students to best prepare them for private and public sector jobs?

    Increasing the number of internships at DOE labs and opportunities for faculty to work at a DOE lab for sabbatical leave could increase the number of individuals with the right expertise in the marketplace. Faculty are often encouraged to take sabbatical leave in order to retool their thinking and update skills. While many of these questions are focused on training students in college, there is a compelling need to retool faculty and those that have already graduated.

11. What type of experiential opportunities are needed to complement classroom instruction to best prepare students for future jobs?

    At the center of any experiential curriculum should be sustained engagement with real data. Unlike the clean, well-structured datasets that tend to appear in coursework, real data is messy: it contains missing values, inconsistencies, and artifacts of how it was collected. Learning to navigate this messiness—knowing how to assess data quality, make defensible cleaning decisions, and document those choices transparently—is a skill that can only be developed through practice. Capstone projects and internships are particularly valuable vehicles for this kind of learning, especially when they are designed around the large, complex datasets generated at DOE National Laboratories and user facilities. These settings offer students exposure to data at a scale and complexity that is

genuinely representative of what they will encounter in careers at the frontiers of AI for science and engineering.

Equally important is experience working on interdisciplinary teams. The Genesis Mission will depend on sustained collaboration among people with different disciplinary backgrounds, institutional affiliations, and technical vocabularies. Students who have only worked within their own field arrive at these collaborations underprepared. Structured team experiences—whether in capstone courses, lab placements, or collaborative research projects—give students the opportunity to practice the negotiation, coordination, and mutual translation that interdisciplinary work requires, and to develop the professional habits that make such collaboration productive.

Finally, degree programs should create significantly more opportunities for students to communicate their work to diverse audiences through both oral and written presentations. The ability to present quantitative findings clearly to an interdisciplinary group—one that may include domain scientists, engineers, policymakers, and AI practitioners—is a distinct and learnable skill. Yet it is one that many statistics and data science programs treat as peripheral. Regular presentation experiences, with feedback from audiences that genuinely span disciplines, would go a long way toward closing this gap and producing graduates who can not only do rigorous work but make that work legible and useful to the broader teams they will join.

12. How many American bachelor's and master's degree students could be trained in a steady state program at your institution? How could the program be scaled across the nation?
13. How could community colleges and other educational institutions contribute and be included in this pipeline?

In summary, the United States can best secure its position as a global AI leader by embracing a holistic, statistically informed approach to AI research and development. Statisticians offer critical expertise in data collection, research design, model evaluation, uncertainty quantification, and ethical considerations—elements essential to responsible AI innovation. Statistics has been, and will continue to be, the foundation of AI technologies. With the support and advancement of statistics, the progress of AI at DOE can better achieve its full potential.

The ASA stands ready to collaborate with the DOE in realizing a successful Genesis Mission. Questions or comments may be directed to ASA Director of Science Policy Steve Pierson: spierson@amstat.org.