



HUGGING FACE

Hugging Face Response to Request for Information Mobilizing Talent for the Genesis Mission and Developing an American Workforce to Advance Artificial Intelligence (AI) for Science and Engineering

Submitted to: Office of Science, Department of Energy
March 2026

About Hugging Face

Hugging Face is a U.S.-based company operating the world's largest open platform for machine learning, with over 2 million models and 500,000 datasets serving 11 million users, including thousands of scientific models spanning protein folding, materials science, climate, and biomedicine. Our Hugging Science initiative (700+ active researchers) demonstrates at scale how open AI infrastructure accelerates scientific participation across institutions of all sizes.

Executive Summary

The Genesis Mission can train 100,000 scientists and engineers, but **DOE must treat AI-for-science as shared national scientific infrastructure** rather than a collection of disconnected training grants. Open-source AI carries direct implications for U.S. competitiveness. The Genesis Mission is an opportunity to strengthen U.S. dominance in AI and scientific research by establishing open, reusable AI artifacts as the default substrate for federally funded science.

Our core recommendation is simple: **make open, reusable artifacts (datasets, models, benchmarks, documentation) the default output of every DOE-supported program, with tiered, risk-based exceptions**. Everything else follows from that. It reduces duplicative spending. It enables regional institutions to participate without bespoke infrastructure. It creates the measurement layer DOE needs to know whether the Genesis investment is actually compounding. And it positions the U.S. as the credible home of open scientific AI globally.

Three additional priorities underpin this:

- **DOE-backed distribution infrastructure** so artifacts are findable, versioned, and maintained over time
- **Evaluation and benchmarking as durable infrastructure**, with sustained funding rather than one-off contests



HUGGING FACE

- **Incentives that reward the work that makes everything else scale** (data curation, documentation, benchmark maintenance), currently the most undervalued and fragile part of the stack

Q1: Catalyzing Research Collaborations Between DOE, Universities, and Industry

The collaborations most likely to compound value are those organized around **open, shared outputs and durable evaluation infrastructure**, rather than individual applications or discipline-specific deliverables.

Require open, reusable artifacts by default with tiered risk exceptions. Every DOE-funded (and thereby paid for by the taxpayers) project should produce public curated datasets, documented code, reproducible pipelines, and benchmark suites as primary deliverables rather than optional byproducts. Where a security or IP rationale exists, delayed or conditional release is appropriate; the default should be open. Our analysis of over 850,000 ML models (2020–2025) shows training data disclosure dropped from nearly 80% in 2022 to 39% in 2025, worsening reproducibility and trust. This trend is driven by incentives, and federal programs are uniquely positioned to reverse it. DOE should require artifact release on a DOE-backed repository with standardized documentation (dataset cards, model cards, eval reports) as a condition of funding.

Establish DOE-backed scientific model and data distribution as national infrastructure. Many scientific domains lack openly available models trained on curated, representative data, so agencies, labs, and grantees recreate equivalent capabilities in parallel, wasting federal investment. DOE should publish high-quality lab datasets and host domain foundation models (biology, materials, climate, physics) on a versioned, FAIR-compliant repository. The Materials Project and the Hugging Face Hub are operating exemplars of how shared data and models scale community participation; the former serves 600,000+ users from a single NSF/DOE investment. DOE should treat this as infrastructure procurement rather than a research project.

Fund evaluation as durable infrastructure. The CASP benchmark, launched in 1994 and still running, ultimately enabled AlphaFold. That 20-year compounding return came from sustained investment in a community evaluation framework with standardized metrics, blind submissions, and post-assessment workshops. DOE should fund recurring challenge series and hosted evaluation services for Genesis-priority domains. Existing community workstreams such as Hugging Face-hosted EvalEval Coalition (400+ researchers building evaluation infrastructure) have direct relevance here; DOE programs could plug into rather than duplicate this work.

Score grants on artifact quality and reuse, not only publications. Explicitly require grant evaluation criteria to include dataset coverage and documentation quality, benchmark submissions, reproducibility of pipelines, and downstream reuse by other groups. This is the single most direct lever DOE has to shift incentives. Without it, artifact release remains a box-checking afterthought.



HUGGING FACE

Q2 and Q8–9: Incentivizing and Designing Dual Competency Programs

Dual-competency programs fail when they treat AI as a tool to be learned rather than a discipline to be practiced with scientific rigor. Domain scientists resist black-box AI systems that cannot be validated against known scientific constraints. Scientific AI requires interpretable outputs, transparent methodologies, and reproducibility standards that differ sharply from consumer AI norms. Training programs that do not embed these norms produce graduates who can run models but cannot evaluate whether the outputs are scientifically meaningful.

Tie training awards to open, reproducible outputs from student work. Programs should require that student research produces a tangible, reusable artifact (a benchmark submission, a validated dataset, or a documented model checkpoint) rather than a grade or a paper. This aligns training with demonstrated, transferable capability and builds the shared infrastructure the Genesis Mission needs as a byproduct of training itself.

Build evaluation discipline into every degree level. At the bachelor's level: data stewardship, documentation practices, and hands-on work with community benchmarks. At the master's level: uncertainty quantification, physics-aware ML, and rigorous validation against domain baselines. At the doctoral and postdoctoral level: dual mentorship combining domain science and ML expertise (the Schmidt AI in Science Fellowship is the strongest existing model), with explicit credit toward degrees for contributions to shared benchmarks and datasets.

Use DOE modular curriculum and microcredentials as a distribution mechanism. DOE should publish validated course modules covering data curation, evaluation design, and reproducible pipelines that universities, community colleges, and regional institutions can adopt without rebuilding from scratch. Stackable microcredentials tied to demonstrated artifact production rather than seat time create a pathway into the bachelor's pipeline for institutions that cannot redesign entire curricula. Hugging Face already hosts educational resources at scale; this is a natural collaboration point.

Create incentives that recognize infrastructure contributions in career progression. Pilot explicit tenure and hiring evaluation language that counts dataset curation, benchmark maintenance, and documentation as research outputs. Without this, the work that makes training programs scalable remains fragile and volunteer-dependent, a structural fragility the Genesis Mission cannot afford at 100,000-person scale.

Q3: Attracting Undergraduates

Students are drawn to programs that offer real problems, real data, and a visible path to employment. Three things DOE controls that matter most:

- **Reserve paid research slots at national laboratories** tied to consortium programs. Unpaid research systematically filters out students from less-resourced backgrounds;



HUGGING FACE

paid, mission-relevant projects with actual DOE datasets are the strongest recruitment signal.

- **Require capstones to produce open artifacts** (a benchmark submission, a validated dataset, or a documented model result) so students graduate with a verifiable portfolio rather than coursework credit alone. Open portfolios are legible to employers in a way transcripts are not.
- **Connect programs to Hugging Science and similar communities** where students can see peers doing credible scientific AI work across institutions and disciplines. Cohort and community structures reduce the isolation that drives domain specialists away from cross-training.

Q4: Non-Funding Contributions DOE and National Laboratories Can Offer

DOE's strongest non-funding levers are data, evaluation infrastructure, and convening authority, none of which the private sector or universities can replicate at equivalent scale or credibility.

- **Publish curated DOE datasets in model-ready formats with provenance and documentation.** This enables reproducible fine-tuning by researchers nationwide and removes the largest single barrier to domain-specific foundation model development outside elite institutions. Distribute via a FAIR-compliant repository with standardized metadata.
- **Operate or sponsor hosted benchmark and evaluation services among experts** for Genesis-priority domains, including blind submission, automated scoring, and post-assessment workshops, replicating the compounding value of CASP and DREAM Challenges. HF's EvalEval Coalition is a direct potential partner here.
- **Reserve small validation allocations** (HPC cycles, user facility time) for student projects that produce reusable, documented artifacts. This ensures training produces deployable tools rather than classroom-only exercises and ties facility access to the open-output norms DOE wants to establish.

Q7: Disciplines Best Suited for Dual Competency Training

Priority should go to domains with strong existing data infrastructure, well-characterized evaluation metrics, and active computational bottlenecks, the conditions under which AI is genuinely accelerative:

- **Materials science and chemistry:** Strong community benchmarks (Materials Project), well-defined inverse problems, active DOE data assets
- **Biology and biomedicine:** Mature AI-for-science applications, active community infrastructure, direct alignment with NIH and DOE joint priorities



HUGGING FACE

- **Climate and earth science:** Strong NASA/DOE/NOAA data foundations, demonstrated AI gains in forecasting and modeling
- **Physics:** Well-defined simulation tasks (plasma fusion, high-energy physics) where AI surrogate models offer large and measurable efficiency gains

Q12 and Q13: Scaling and Including Community Colleges

The 100,000-person goal cannot be reached through research-intensive universities alone. Nearly 40% of recent ML model development already originates from nontraditional contributors. The capacity exists broadly; the infrastructure and on-ramps do not.

- **Develop articulation agreements** between community college AI-for-science programs and partner university bachelor's programs, with explicit credit for DOE microcredential completion
- **Deploy shared digital infrastructure** (compute access via NAIRR, curated FAIR datasets, open model repositories) to institutions without dedicated HPC resources. NAIRR expanded from 35 to 600+ projects in one year when barriers dropped; that demand signal should inform Genesis design
- **Align training with regional industry ecosystems** in biotechnology, agriculture, advanced manufacturing, and energy, sectors where community college graduates are already employed and where dual-competency skills create immediate, measurable labor market value

Conclusion

The structural barriers to AI-for-science are not primarily technical. They are incentive problems: data curation is unrewarded, documentation is skipped, evaluation infrastructure is underfunded, and the open artifacts that would let every institution build on prior work are never produced. The Genesis Mission has the scale and the mandate to change those incentives directly, by making open, reusable outputs the condition of funding rather than the aspiration. That one policy lever, consistently applied, does more to build durable national AI-for-science capacity than any curriculum redesign or fellowship program. It is also the lever that positions the U.S. as the credible home of open scientific AI infrastructure, ahead of China and proprietary platforms.

Submitted by:

Avijit Ghosh, Technical AI Policy Researcher, Hugging Face
Irene Solaiman, Chief Policy Officer, Hugging Face