N/6/2026-NeGDN/6/2026-NeGD
National e-Governance Division
Digital India Corporation
Electronics Niketan, 6, CGO Complex Lodhi Road,
New Delhi – 110003
Website: www.negd.gov.in / www.dic.gov.in

Web Advertisement
02.06.2026

The National e-Governance Division (NeGD) is an independent business division under the Digital India Corporation, Ministry of Electronics and Information Technology. NeGD has been playing a pivotal role in supporting MeitY in Programme Management and implementation of e-Governance projects and initiatives undertaken by various Ministries/ Departments, both at the Central and State levels.

NeGD has been spearheading several innovative initiatives under the aegis of the Digital India Programme. Those have been developed keeping the vision areas of Digital India at the core- providing digital infrastructure as a core utility to every citizen, governance and services on demand and in particular, digital empowerment of the citizens of our country; some of these initiatives include DigiLocker, UMANG, Poshan Tracker, OpenForge Platform, API Setu, National Academic Depository, Academic Bank of Credits, Learning Management System.

It has myriad roles and responsibilities from supporting Central Line Ministries and State Government Departments on e-Governance projects, reviewing State Action Plans, offering support in technology management, strategy formulation & implementation of Emerging Technologies viz. AI, Blockchain, GIS etc., to facilitating digital diplomacy with focus on Indian startups and products

NeGD has been a leader in implementation and execution of a gamut of pilot/ infrastructure/ technical/ special projects and support components to framing core policies, project appraisals, R&D, and guiding /conducting assessments, undertaking activities for building capacities of both Government officials and] other stakeholders, and creating mass awareness about schemes and services under the Digital India Programme.

NeGD is currently inviting applications for the following positions purely on Contract basis initially for a period of 3 years which is further extendable as per the requirement of the project.

S. No	Position	Experience	Vacancy
1	Senior AI/ML Engineer	5-9 years	01
2	Senior MLOps Engineer	8-10 Years	01

* The maximum age limit shall be 55 years on the closing date of receipt of application.
** The place of posting shall be in New Delhi but transferable to project locations of NeGD as per existing policy of NeGD/DIC.
Screening of applications will be based on qualifications, age, and relevant experience. NeGD reserves the right to fix higher threshold of qualifications and experience for screening and limiting the number of candidates for interview. Only shortlisted candidates shall be invited for interviews. NeGD reserves the right to not to select any of the candidates without assigning any reason thereof.
The details can be downloaded from the official website of **DIC** and **NeGD**, viz. **www.dic.gov.in**, **www.negd.gov.in**.

Eligible candidates may apply ONLINE: https://ora.digitalindiacorporation.in/

Last date for submission of applications will be: 14.06.26

SENIOR AI/ML ENGINEER

Experience Required: 5-9 Years

JOB SUMMARY

We are seeking a Senior AI/ML Engineer to lead the research, design, and development of advanced machine learning models and algorithms. This role focuses exclusively on the model development lifecycle—including architectural design, training methodology, optimization techniques, and algorithmic innovation—while collaborating with MLOps teams for production deployment.

The incumbent will solve complex business problems through predictive modeling, deep learning, and generative AI research. Responsibilities span from mathematical formulation and prototype development to training large-scale models (1B+ parameters) and rigorous evaluation. This position requires deep expertise in machine learning theory, neural network architecture, and statistical modeling, with emphasis on creating novel solutions rather than infrastructure management.

I. CORE RESPONSIBILITIES

1. Design and architect neural network models including Transformers, CNNs, RNNs, and hybrid architectures; make decisions on layer configurations, attention mechanisms, activation functions, and connectivity patterns for optimal performance.
2. Develop and implement training algorithms and optimization strategies including custom loss functions, learning rate schedules, gradient clipping, and regularization techniques to ensure stable convergence and generalization.
3. Fine-tune pre-trained foundation models (LLaMA, Mistral, BERT, GPT, T5) using Parameter-Efficient Fine-Tuning (PEFT) methods including LoRA, QLoRA, Prefix Tuning, and AdaLoRA for domain-specific applications.
4. Implement Reinforcement Learning from Human Feedback (RLHF) and Constitutional AI methodologies; design reward models, policy optimization algorithms (PPO, DPO), and human preference learning systems.
5. Engineer high-quality training datasets through data collection strategies, cleaning pipelines, augmentation techniques, and synthetic data generation; ensure data representativeness and bias mitigation.
6. Design and execute comprehensive model evaluation frameworks including statistical significance testing, cross-validation strategies, benchmark dataset evaluation (MMLU, HumanEval, GLUE, SuperGLUE), and custom metrics development.
7. Develop Retrieval-Augmented Generation (RAG) architectures including embedding model selection, retrieval algorithms, context integration strategies, and relevance scoring mechanisms to enhance model accuracy.
8. Optimize model architectures for efficiency through knowledge distillation, model pruning, quantization-aware training, and neural architecture search (NAS) without compromising accuracy.

9. Perform rigorous statistical analysis and hypothesis testing on model outputs; identify failure modes, error analysis, and edge cases requiring architectural improvements.
10. Collaborate with domain experts to translate business requirements into mathematical formulations and ML problem statements; define target variables, feature spaces, and success criteria.
11. Mentor junior researchers and engineers on machine learning theory, algorithmic best practices, experimental design, and research methodologies; conduct code reviews for model implementations.
12. Document research findings, model architectures, training methodologies, and experimental results in technical reports; publish papers in conferences or journals and present at technical forums.
13. Analyze model interpretability and explainability using attention visualization, SHAP values, LIME, and gradient-based attribution methods to ensure transparency in AI decision-making.

II. ESSENTIAL QUALIFICATIONS & EXPERIENCE

Educational Qualifications:

- Bachelor's degree (B.E./B.Tech) in Computer Science, Engineering, Mathematics, Statistics, Physics, or related quantitative field from a recognized university.
- Master's degree (M.Tech/MS) or PhD in Machine Learning, Artificial Intelligence, Computer Science, or related field highly **desirable**; exceptional candidates with Bachelor's degree and significant research experience may be considered.
- Strong foundation in linear algebra, calculus, probability theory, statistics, and optimization theory essential.

Experience Requirements:

- Minimum 5-9 years of research and development experience in applied machine learning, with demonstrable expertise in designing and training neural networks.
- Extensive experience in at least two domains: Natural Language Processing (NLP), Computer Vision, Speech Recognition, Recommendation Systems, or Reinforcement Learning.
- Research Publications: First-author publications in Tier-1 ML conferences (NeurIPS, ICML, ICLR, ACL, CVPR) or journals (JMLR, TPAMI, TACL) (are preferred).

III. TECHNICAL COMPETENCIES REQUIRED

Machine Learning Theory & Algorithms:

- Deep theoretical understanding of machine learning algorithms including supervised learning (SVMs, Random Forests, Gradient Boosting), unsupervised learning (clustering, dimensionality reduction, GMMs), and deep learning architectures.
- Expert-level proficiency in PyTorch (strongly preferred) or TensorFlow for implementing custom models, loss functions, and training loops from first principles.
- Advanced knowledge of transformer architectures (BERT, GPT, T5, LLaMA, Mistral) including self- attention mechanisms, positional encodings, layer normalization, and feed-forward networks.
- Mathematical optimization: Gradient descent variants (SGD, Adam, AdamW, LAMB), learning rate scheduling (cosine annealing, warm restarts), second-order methods, and convex/non-convex optimization theory.

- Statistical modeling: Bayesian methods, probabilistic graphical models, hypothesis testing, confidence intervals, and experimental design (A/B testing, factorial designs).

Deep Learning & Generative AI:

- Fine-tuning methodologies: Full fine-tuning, freezing strategies, layer-wise learning rates, discriminative fine-tuning, and Parameter-Efficient Fine-Tuning (LoRA, QLoRA, Prefix Tuning, P- tuning, Adapter layers).
- Reinforcement Learning: Policy gradient methods (REINFORCE, A2C, PPO), Q-learning, actor- critic architectures, and RLHF implementation for language models.
- Generative modeling: Understanding of VAEs, GANs, diffusion models, and autoregressive generation techniques.
- Model compression: Knowledge distillation, pruning (structured/unstructured), quantization-aware training, and neural architecture search (NAS).
- Embedding techniques: Word2Vec, GloVe, FastText, contextual embeddings (ELMo, BERT), sentence embeddings (Sentence-BERT, SimCSE), and contrastive learning.

Research & Development Tools:

- Experiment tracking: MLflow, Weights & Biases, or Neptune for logging experiments, hyperparameters, and results.
- Data manipulation: Pandas, NumPy, SciPy, Scikit-Learn for data analysis and classical ML algorithms.
- NLP libraries: Hugging Face Transformers, Tokenizers, Datasets library, SpaCy, NLTK.
- Visualization: Matplotlib, Seaborn, Plotly, TensorBoard for model visualization and performance analysis.
- Version control: Git for code management; DVC for data and model versioning.
- Hardware Awareness: Understanding of GPU/TPU architecture implications for model design (memory constraints, mixed precision training, model parallelism strategies) without responsibility for infrastructure setup.

SENIOR MLOps ENGINEER

Experience Required: 5-9 Years

I. JOB SUMMARY

We are building the MLOps backbone for enterprise generative AI, supporting both air-gapped private GPU clusters (such as for sensitive financial/healthcare data) and hyperscale cloud platforms (such as AWS SageMaker, Google Vertex AI, Azure ML).

Expected to architect infrastructure that handles billions of inference tokens monthly across hybrid environments—such as optimizing Llama 3/Mistral deployments on private A100 clusters while orchestrating GPT-4/Claude fine-tuning pipelines on managed cloud services. This is a high-visibility role requiring deep expertise in LLM serving engines, distributed GPU systems, and cloud-agnostic MLOps.

II. CORE RESPONSIBILITIES

1. Architect self-hosted inference clusters using vLLM, TGI (Text Generation Inference), and TensorRT-LLM on on-premise NVIDIA DGX systems and GPU racks, ensuring sub-100ms latency for 70B+ parameter models.
2. Design parallel workflows on AWS SageMaker (Endpoints/Pipelines), Google Vertex AI (Prediction/Training), and Azure ML for elastic training workloads and managed foundation model APIs.
3. Implement cloud-agnostic model deployment using Kubernetes (EKS/GKE/AKS) with portability across private data centers and cloud VPCs, ensuring zero vendor lock-in.
4. Deploy multi-GPU inference parallelism (tensor + pipeline parallelism) for foundation models using Ray Serve, NVIDIA Triton, and custom FastAPI stacks.
5. Optimize inference economics through quantization (AWQ/GPTQ/FP8), KV-cache optimization, and continuous batching—reducing per-token costs by 40%+.
6. Build auto-scaling GPU node pools (Karpenter/Cluster Autoscaler) that respond to inference demand spikes within seconds.
7. Implement RLHF (Reinforcement Learning from Human Feedback) infrastructure using DeepSpeed, LoRA/QLoRA fine-tuning pipelines, and distributed training orchestration.
8. Design evaluation frameworks for LLMs: automated benchmarking (MMLU, HumanEval), A/B testing for model versions, and human-in-the-loop feedback systems.
9. Manage vector database infrastructure (Pinecone, Weaviate, Milvus, pgvector) for RAG systems spanning private and cloud environments.
10. Build CI/CD for ML using GitOps (ArgoCD/Flux) with model versioning (MLflow/DVC), automated testing for data drift, and canary deployments for model updates.
11. Implement feature stores (Feast/Tecton) and experiment tracking (Weights & Biases/MLflow) supporting both cloud and on-premise data lakes.
12. Create observability stacks for LLMs: token-level latency tracking, GPU memory saturation alerts, and cost-per-inference dashboards using Prometheus/Grafana/CloudWatch.

13. Manage secrets, model encryption at rest (HashiCorp Vault), and network policies (Istio/Linkerd) for multi-tenant model serving.

III. ESSENTIAL QUALIFICATIONS & EXPERIENCE

Educational Qualifications:

- Bachelor's degree (B.E./B.Tech) in Computer Science, Engineering, Mathematics, or related technical field from a recognized university.
- Master's degree (M.Tech/MS) in Machine Learning, Computer Science, Artificial Intelligence, or related field **desirable**.
- Relevant professional certifications in cloud platforms (AWS/Azure/GCP) and Kubernetes (CKA/CKAD) highly desirable.

Experience Requirements:

- Minimum 5-9 years of hands-on experience in production ML infrastructure engineering, with at least 2 years dedicated to large-scale model deployment and MLOps.
- Demonstrable track record of deploying and maintaining 70B+ parameter models in production environments (are preferred).
- Proven experience managing both on-premise GPU clusters (NVIDIA DGX, A100/H100) and cloud-based ML platforms (AWS SageMaker, Google Vertex AI, or Azure ML).

IV. TECHNICAL COMPETENCIES REQUIRED

Infrastructure & Systems:

- Expert-level proficiency in Kubernetes (GPU operators, taints/tolerations, multi-tenancy) across both on-premise (Rancher/OpenShift) and cloud (EKS/GKE/AKS) environments.
- Deep expertise in LLM serving engines: Proven hands-on experience with vLLM, TGI (Text Generation Inference), or TensorRT-LLM in production settings.
- Professional-level certification or equivalent experience in AWS SageMaker, Google Vertex AI, or Azure ML—including model registry, endpoints, and pipeline orchestration.
- Strong understanding of NVIDIA Hopper/Ampere architectures, NVLink/InfiniBand networking, and CUDA optimization.
- CUDA kernel optimization, custom inference kernels, or TritonML server extensions.
- Infrastructure as Code: Terraform, Helm, Kustomize for reproducible GPU cluster provisioning.

Machine Learning & Distributed Systems:

- Exper-level Python programming with PyTorch/TensorFlow.
- Distributed training frameworks: DeepSpeed, Horovod, PyTorch DDP/FSDP.
- LLM Stack: LangChain, LlamaIndex, Hugging Face Transformers, and agentic workflow

orchestration.

- Data Engineering: Apache Spark, Airflow, and feature engineering at scale (terabyte+ datasets).
- Database Systems: Vector databases (Pinecone, Weaviate, Milvus, pgvector) and feature stores (Feast/Tecton).

DevOps & Observability:

- CI/CD for ML: GitOps (ArgoCD/Flux), model versioning (MLflow/DVC), and automated testing.
- Observability: Prometheus, Grafana, ELK stack, and cloud-native monitoring (CloudWatch/Stackdriver/Azure Monitor).
- Security: HashiCorp Vault, Istio/Linkerd service mesh, network policies, and secrets management.

General Conditions applicable to all applicants covered under this advertisement

1. Those candidates, who are already in regular or contractual employment under Central / State Government, Public Sector Undertakings or Autonomous Bodies, are expected to apply through proper channel or attach a 'No Objection Certificate' from the employer concerned with the application OR produce No Objection Certificate at the time of interview.
2. NeGD reserves the right to fill all or some or none of the positions advertised without assigning any reason as it deems fit.
3. The positions are purely temporary in nature for the project of NeGD/DIC and the appointees shall not derive any right or claim for permanent appointment at NeGD/DIC or on any vacancies existing or that shall be advertised for recruitment by NeGD in future.
4. NeGD reserves the right to terminate the appointments of all positions with a notice of one month or without any notice by paying one month's salary in lieu of the notice period.
5. The designation of the selected candidates shall be mapped as per the existing designation policy of NeGD.
6. In case of a query, the following officer may be contacted:

HR Team
National e Governance Division, 4th Floor, Electronics Niketan,6-CGO, Complex
Lodhi Road, New Delhi – 110003
Email: Negdhr@digitalindia.gov.in