NexusForge AI Knowledge Base

Complete Documentation for AI Model Creation
15 Architectures | 7 Domains | 7 Frameworks

Version 2.0.0

AutomataNexus, LLC

Generated: December 27, 2025

Table of Contents

Introduction

NexusForge AI is an advanced artificial intelligence system designed for Building Management Systems (BMS) and HVAC control applications. Built on the LLaMA 3.1 foundation, NexusForge provides intelligent automation, fault detection and diagnostics (FDD), and optimized control strategies for commercial and industrial buildings.

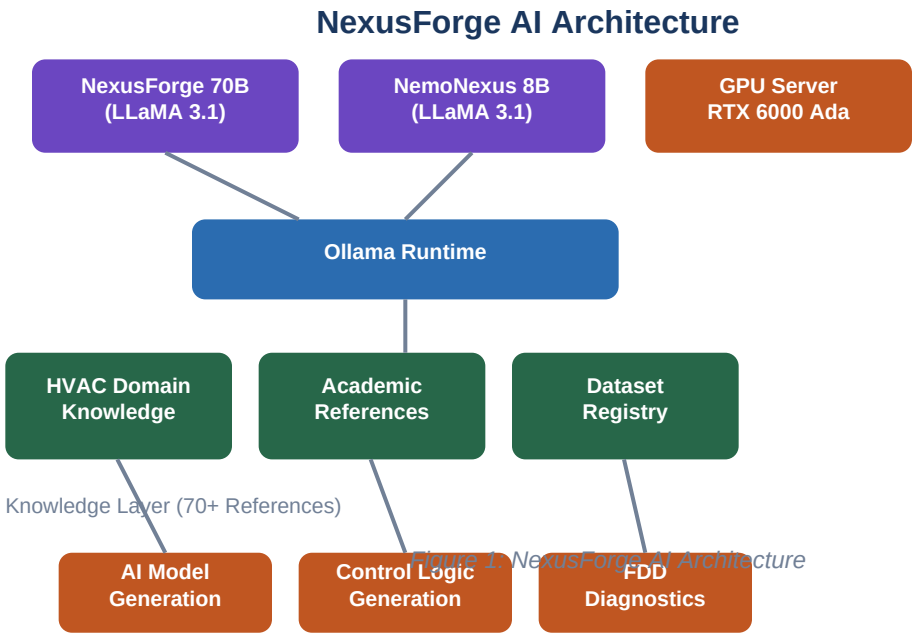

Figure 1: NexusForge AI Architecture

Key Components

NexusForge 70B - Primary reasoning model for complex HVAC analysis, control logic generation, and fault diagnosis.

NemoNexus 8B - Lightweight model for real-time monitoring, quick responses, and edge deployment.

HVAC Domain Knowledge - 70+ academic references and domain-specific context.

Dataset Registry - 26 curated datasets (13+ GB) for model training.

NexusForge 70B

NexusForge 70B is the primary AI model for complex reasoning tasks in building automation. Based on Meta's LLaMA 3.1 70B architecture, it has been fine-tuned with HVAC domain knowledge and optimized for BMS applications.

Model Specifications

Parameter	Value
Base Model	LLaMA 3.1 70B
Parameters	70 billion
Context Length	128K tokens
Inference Server	Ollama
GPU Requirement	RTX 6000 Ada (48GB VRAM)
Quantization	Q4_K_M
Response Time	~2-5 seconds

Capabilities

AI Model Generation - Creates custom neural network architectures for specific equipment types.
Control Logic Generation - Produces JavaScript control logic for equipment executors.
Fault Detection - Analyzes sensor data patterns to identify equipment faults.
Sequence of Operations - Generates ASHRAE Guideline 36 compliant sequences.
Training Notebooks - Creates Jupyter notebooks for model training with recommended datasets.

NemoNexus 8B

NemoNexus 8B is a lightweight model optimized for real-time operations and edge deployment. It provides quick responses for routine queries while maintaining HVAC domain expertise.

Model Specifications

Parameter	Value
Base Model	LLaMA 3.1 8B
Parameters	8 billion
Context Length	128K tokens
Inference Server	Ollama
GPU Requirement	Any CUDA GPU (8GB+ VRAM)
Quantization	Q4_K_M
Response Time	~0.5-1 second

Use Cases

Real-time Monitoring - Quick status checks and alarm interpretation.
Chat Interface - Responsive conversational AI for operators.
Edge Deployment - Runs on local controllers with limited resources.
Fallback Model - Handles requests when NexusForge 70B is busy.

HVAC Domain Knowledge Base

The knowledge base provides NexusForge AI with comprehensive domain expertise through 70+ curated academic references, equipment specifications, and industry standards. This knowledge is stored in hvac_domain.py and integrated into all AI responses.

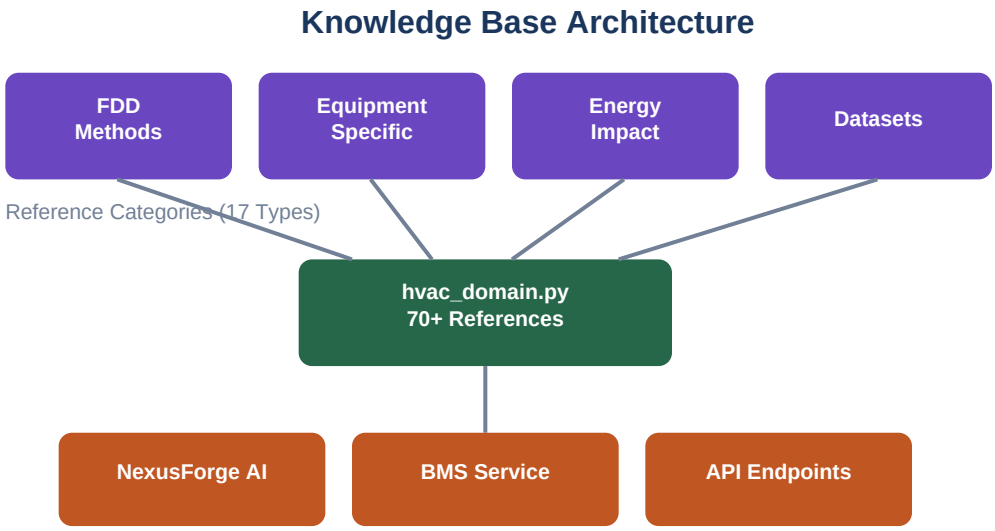

Figure 2: Knowledge Base Architecture

Figure 2: Knowledge Base Architecture

Reference Categories

FDD Methods - Fault detection algorithms, deep learning, transformers, CNNs
Equipment Specific - AHU, VAV, chiller, boiler, RTU documentation
Energy Impact - Energy efficiency, cost savings, performance metrics
Fault Prevalence - Empirical studies on common HVAC faults
Transfer Learning - Cross-building model adaptation techniques
Standards - ASHRAE Guidelines 36, Standard 55, Standard 90.1
Datasets - Academic dataset publications and documentation
Predictive Control - Model Predictive Control (MPC) research
Thermal Comfort - Occupant comfort and indoor air quality
Maintenance - Predictive maintenance and lifecycle management

Academic References

Key academic papers and research integrated into NexusForge AI:

FDD Methods

Automated Fault Diagnosis Detection of AHUs Using Transformer-based Methods at 24-hour Hospital

Wang, S. (2025). Building and Environment

Fault Diagnosis of AHUs in Auditorium Using Real Operational Labeled Data

Wang, S., Kim, J., Park, S., Kim, J. (2025). Computers in Civil Engineering

Data-driven FDD for AHUs Considering Undefined States

Yun, W.S., Hong, W.H., Seo, H. (2021). Journal of Building Engineering

Interpretation of CNN-based HVAC FDD Using Improved Layer-wise Relevance Propagation

Li, G. et al. (2023). Energy and Buildings

Review of Computing-based Automated FDD of HVAC Systems

Chen, J. et al. (2022). Renewable and Sustainable Energy Reviews

Transfer Learning

Transfer Learning for Cross-Building FDD

Chen, Y. et al. (2023). Applied Energy

Data-Centric Approach for Cross-Building Transfer Learning

Fan, C. et al. (2021). Energy and Buildings

Standards

ASHRAE Guideline 36: High-Performance Sequences of Operation

ASHRAE (2021)

ASHRAE Standard 55: Thermal Environmental Conditions

ASHRAE (2020)

Dataset Registry

NexusForge AI has access to 26 curated HVAC datasets totaling over 13 GB. These datasets are used for model training, validation, and benchmarking.

Dataset Summary

Category	Count	Total Size
FDD	8	~1.7 GB
Chiller	3	~9.4 GB
Boiler	2	~2.1 GB
VAV	1	32 MB
DOE Reference	6	~175 MB
Controls/Lab	4	~300 MB
Other	2	~560 MB

Top Datasets

LBNL Simulated Chiller Plant (8.14 GB)

7 faults, 1-year simulation, 77 parameters

LBNL Simulated Boiler Plant (2.08 GB)

5 faults, 1-year simulation, 22 parameters

LBNL Chiller Plant (Chen) (1.14 GB)

World's largest labeled HVAC FDD dataset

SHIFDR Michigan (500 MB)

14 commercial buildings, demand response data

AlphaBuilding (500 MB)

Synthetic building operation data

API Integration

NexusForge AI is accessible through the NexusConnectBridge REST API. All endpoints are authenticated via JWT tokens.

Key Endpoints

POST /api/bms/ai/chat
Chat with NexusForge AI

POST /api/bms/ai/assist
Get AI assistance with domain context

POST /api/bms/models/generate
Generate custom AI models

POST /api/bms/logic/generate-file
Generate equipment control logic

GET /api/bms/datasets
List available datasets

GET /api/bms/domain/context
Get HVAC domain knowledge

GPU Infrastructure

NexusForge models run on dedicated GPU infrastructure with NVIDIA RTX 6000 Ada (48GB VRAM). Access is provided through the NexusConnectBridge API at nexusconnectbridge.automatanexus.com.

ML Architecture Patterns

NexusForge AI includes comprehensive knowledge of 15 neural network architecture patterns for generating models across any domain, not just HVAC.

Core Architectures

MLP (Multi-Layer Perceptron)

Fully connected feedforward network. Best for tabular data and simple tasks.

CNN (Convolutional Neural Network)

Extracts spatial features using convolutions. Best for images and spectrograms.

LSTM (Long Short-Term Memory)

Recurrent architecture with gates for long-range dependencies. Best for time series.

Transformer

Attention-based architecture. Foundation of modern LLMs. Best for NLP and sequences.

ResNet (Residual Network)

CNN with skip connections enabling 100+ layer networks. Best for image classification.

Specialized Architectures

Autoencoder - Encoder-decoder for dimensionality reduction and anomaly detection.

VAE (Variational Autoencoder) - Probabilistic autoencoder with structured latent space for generation.

GAN (Generative Adversarial Network) - Generator vs discriminator for high-quality image generation.

Diffusion Model - Iterative denoising for state-of-the-art generation (DALL-E, Stable Diffusion).

U-Net - Encoder-decoder with skip connections for segmentation tasks.

Vision Transformer (ViT) - Transformer applied to image patches. Scales better than CNNs.

BERT - Bidirectional transformer encoder for NLP understanding tasks.

GPT - Autoregressive transformer decoder for text generation.

Siamese Network - Twin networks with shared weights for similarity learning.

Graph Neural Network - Message passing on graph-structured data.

Training Strategies

NexusForge includes 7 training strategy templates optimized for different scenarios.

Strategy	Key Configuration
Standard Classification	Adam optimizer, LR 1e-3, ReduceOnPlateau scheduler, Dropout regularization
Transfer Learning	AdamW optimizer, differential LR (backbone vs head), gradual unfreezing
Large Scale Training	AdamW with cosine warmup, mixed precision, distributed training
Few-Shot Learning	Meta-learning algorithms, heavy augmentation, episodic training
Time Series Forecasting	Adam optimizer, proper time-based splits, recurrent dropout
Generative Training	Separate G/D learning rates, spectral normalization, FID monitoring
Self-Supervised	Large batch sizes, contrastive loss, momentum updates

Domain Templates

NexusForge supports model creation across 7 application domains with domain-specific preprocessing, architectures, metrics, and best practices.

Domain	Common Tasks	Architectures
Computer Vision	Image classification, object detection, segmentation	CNN, ResNet, ViT, U-Net, YOLO
Natural Language Processing	Text classification, NER, generation, translation	Transformer, BERT, GPT, LSTM
Time Series	Forecasting, classification, anomaly detection	LSTM, GRU, Transformer, CNN
Tabular Data	Classification, regression on structured data	MLP, TabTransformer, XGBoost
Audio Processing	Speech recognition, music generation, classification	CNN, Transformer, RNN
Generative AI	Image/text generation, style transfer	GAN, VAE, Diffusion, GPT
Reinforcement Learning	Game playing, robotics, optimization	MLP, CNN, Transformer

Framework Knowledge

NexusForge can generate code for 7 deep learning frameworks with cross-framework layer mappings, best practices, and ready-to-use templates.

Supported Frameworks

Framework	Strengths	Best For
PyTorch	Dynamic computation graph, research-focused	Research, custom architectures
TensorFlow	Production-focused with TF Serving/Lite	Production, mobile/edge deployment
Keras	High-level API, multi-backend (TF/PyTorch/JAX)	Beginners, quick prototyping
JAX/Flax	Functional, composable transformations	TPU training, custom autodiff
PyTorch Lightning	Organized PyTorch with built-in distributed	Production PyTorch projects
Hugging Face	100k+ pretrained models, unified API	NLP, transfer learning
scikit-learn	Classical ML with consistent API	Baselines, preprocessing

Code Templates Available

- PyTorch Image Classifier (CNN)
- Keras Image Classifier (CNN)
- Hugging Face Text Classifier (Transformer)
- PyTorch LSTM Time Series Forecaster
- PyTorch Autoencoder (Anomaly Detection)

Cross-Framework Layer Mappings

NexusForge includes mappings for common layers across frameworks: Dense/Linear, Conv2D, LSTM, BatchNorm, Dropout, Attention, LayerNorm, and Embedding. This enables generating equivalent code for any supported framework.