NATIONAL SCHOOL OF SCIENCES

Roadmap to the International Olympiad in Artificial Intelligence

A Comprehensive Syllabus and Preparation Guide for NEB Grade 11–12 Students

---

PREPARED BY

Divya Darsheel Sharma

September 2026

Document Approval & Preface

This comprehensive roadmap has been meticulously prepared by Divya Darsheel Sharma and formally approved by the Department of Computer Science of the National School of Sciences. It is designed to serve as the definitive guide for students under the National Examination Board (NEB) curriculum in Nepal who aspire to transition from standard school computer science to the rigorous demands of the national AI selection contests and ultimately the International Olympiad in Artificial Intelligence (IOAI).

> *Approved by: The Department of Computer Science, National School of Sciences. Prepared by: Divya Darsheel Sharma. Date: September 2026.*

Unlike older academic Olympiads, the IOAI syllabus is explicitly designed around both theory and hands-on implementation. The official 2026 syllabus divides every topic into Theory, Practice, or Both, explicitly prioritizing breadth over excessive theoretical depth. It has also expanded significantly beyond earlier versions to include transformers, embeddings, audio models, diffusion models, and multimodal AI. In Nepal, the IOAI recognizes the Himalayan Computational Linguistics Olympiad as the accredited national organization, with teams selected through a nationwide contest pipeline.

Executive Summary

This document provides a structured, rigorous, and exhaustive syllabus breakdown for NEB Grade 11–12 students. The curriculum is divided into progressive levels, from Level 0 (Python & Computational Foundations) to Level 9 (AI Olympiad Problem Solving), followed by a strategic breakdown of the District, Province, National, and International competition tiers. A central tenet of this roadmap is that IOAI preparation requires building and evaluating AI solutions under constraints, rather than merely memorizing theoretical architectures or engaging in generic software engineering.

Level 0 — Python & Computational Foundations

Before machine learning, students must be able to work fluently with data and models. The official IOAI syllabus classifies Python basics, NumPy, Pandas, Matplotlib/Seaborn, Scikit-learn, PyTorch, tensors, and CPU/GPU training as essential practical skills.

Python & Data Manipulation (Priority: Maximum)

- Python: Variables, data types, loops, functions, exceptions, file handling, and basic object-oriented concepts.
- NumPy: Arrays, shapes, broadcasting, vectorized operations, and matrix manipulations. AI data is fundamentally tensors.
- Pandas: Series, DataFrames, loading data, handling missing values, grouping, merging, and basic feature processing.
- Visualization: Line plots, scatter plots, histograms, heatmaps, and distributions using Matplotlib and Seaborn.

PyTorch Fundamentals

Students do not need to become expert PyTorch engineers, but they must be able to take an unfamiliar dataset or model setup and work with it competently. Master tensor shapes, operations, CPU vs. GPU training, dataset concepts, and the standard training loop (Model, Loss, Optimizer).

Level 1 — Mathematics for AI

This level covers the minimum mathematical foundation required to understand what models are actually doing, without requiring a full university-level mathematics degree.

- Linear Algebra: Scalars, vectors, matrices, dot products, norms, linear transformations, and basic eigenvalue intuition. Understand how $x \rightarrow Wx + b$ represents a transformation.
- Probability & Statistics: Conditional probability, Bayes′ theorem, random variables, expected value, variance, covariance, sampling, and outlier detection.
- Calculus & Optimization: Derivatives, partial derivatives, gradients, the chain rule, objective/loss functions, local/global minima, and gradient-based optimization intuition.

Level 2 — Data Science & Machine Learning

This is the first true IOAI core. The 2026 syllabus includes supervised and unsupervised learning, evaluation, overfitting, tuning, cross-validation, feature engineering, and data processing.

Data Preparation & Feature Engineering (Priority: Maximum)

- Data Prep: Cleaning, missing values, normalization, standardization, train/validation/test splits, data leakage prevention, and basic augmentation.
- Feature Engineering: Transforming high-dimensional or raw data into compact, informative features via selection, one-hot encoding, statistical features, and dimensionality reduction.

Classical Machine Learning Algorithms

- Regression & Classification: Linear regression, logistic regression, K-Nearest Neighbors, and Support Vector Machines (SVM).
- Tree-Based Models: Decision trees, Random Forests (bagging), and Gradient Boosting.
- Regularization: L1 and L2 regularization, sparsity, and bias/variance implications.

Unsupervised Learning & Evaluation

- Clustering: K-Means, DBSCAN, hierarchical, and spectral clustering.
- Dimensionality Reduction: PCA (Theory + Practice), t-SNE, and UMAP.
- Model Evaluation: Accuracy, precision, recall, F1-score, confusion matrices, ROC/AUC, and threshold selection.

- Generalization: Underfitting vs. overfitting, bias/variance tradeoff, cross-validation, and hyperparameter tuning.

Level 3 — Neural Networks & Deep Learning

This is where AI Olympiad preparation becomes substantially deeper, moving from classical ML to modern deep learning architectures.

Core Neural Network Theory

- Perceptrons & MLPs: Neurons, weights, biases, hidden layers, forward propagation, and decision boundaries.
- Activations & Loss: ReLU, Sigmoid, Tanh, MSE, MAE, and Cross-entropy. Understand why activations and specific loss functions are necessary.
- Training Mechanics: Gradient descent, learning rates, batch/epoch concepts, and the mathematical intuition behind backpropagation and computational graphs.

Modern Deep Learning Techniques

- Embeddings: Dense representations and semantic similarity for text, image, and audio data.
- Optimization: SGD, mini-batch training, Adam, and AdamW. Understand when and how to use them without needing to derive every internal detail.
- Regularization: Dropout, early stopping, weight decay, weight initialization, and batch normalization.
- Fine-Tuning: Pre-trained models, full fine-tuning, and parameter-efficient fine-tuning (PEFT).

Level 4 — Computer Vision

The 2026 IOAI syllabus explicitly expands into modern computer vision, requiring both theoretical understanding and practical implementation of visual models.

CNNs & Core Vision Tasks (Priority: Maximum)

- CNNs: Convolutions, kernels, feature maps, stride, padding, channels, and pooling. Understand local patterns leading to hierarchical features.
- Tasks: Image classification, object detection (YOLO, SSD, DETR), and semantic segmentation (U-Net).
- Pre-trained Encoders: Transfer learning and feature extraction using architectures like ResNet.
- Image Augmentation: Flipping, cropping, noise, and knowing when augmentation helps versus when it hurts.

Advanced Vision Topics

- Generative Models: GANs (Generator/Discriminator adversarial training) and Diffusion Models (forward noising, reverse denoising, conditioning).
- Self-Supervised Vision: Learning useful representations without conventional labels.
- Vision-Language Models: Shared embedding spaces, contrastive alignment, and CLIP-style models.

Levels 5, 6 & 7 — NLP, Audio & Multimodal AI

Level 5: Natural Language Processing

- Text Processing: Text classification, tokenization, vocabulary building, and handling variable-length sequences.
- Language Models: Autoregressive modeling, context, and probability distributions.
- Transformers & Attention: Queries, keys, values, self-attention, multi-head attention, and positional information. (Theory required for text/image).
- Pre-trained Models: BERT-style encoders, encoder-decoder architectures, and using open-source or API-based LLMs.

Level 6: Audio AI (2026 Expansion)

- Audio Processing: Audio signals, sampling, spectrograms, and time-frequency representations.
- Audio Models: Practical familiarity with HuBERT, Whisper, Qwen-Audio, and Voxtral. Focus on what problems they solve and how to process their outputs.

Level 7: Multimodal AI

- Handling systems that span text, image, audio, video, time-series, and tabular data.
- Multimodal representation, shared embeddings, cross-modal relationships, and multimodal inference.

Levels 8 & 9 — Practical AI Engineering & Problem Solving

This is the difference between knowing AI and actually solving an IOAI task. The hidden curriculum of the Olympiad.

Practical AI Engineering

- Dataset Handling: Loading, inspecting, cleaning, splitting, and transforming unknown datasets.
- Baseline Construction: Always build a simple baseline, evaluate it, inspect errors, and improve iteratively.
- Model Selection: Deciding between linear models, trees, SVMs, CNNs, or pre-trained transformers based on the specific data and task.
- Error Analysis: Identifying if failures are due to data imbalance, leakage, underfitting, weak representations, or incorrect preprocessing.

AI Olympiad Problem Solving

- Read the Dataset Before the Model: Inspect, visualize, identify data types, preprocess, and establish a baseline.
- Choose the Simplest Viable Model: Olympiad tasks reward results, not technological cosplay. Sometimes logistic regression or gradient boosting outperforms a massive neural network.
- Experimental Iteration: Hypothesis, experiment, metric evaluation, error analysis, change, and repeat.
- Time-Constraint Optimization: Balancing training time, memory limits, model quality, and implementation speed.

The Four Olympiad Levels

The progression from District to International requires a shift from AI literacy to competitive AI engineering.

District Level — AI Literacy & Foundations

Master Python, NumPy, Pandas, basic plotting, and classical ML (linear/logistic regression, KNN, trees, K-means). Objective: Understand what AI methods are and what problems they solve without drowning in transformers.

Province Level — Practical ML

Focus on data cleaning, feature engineering, train/validation/test splits, metrics, overfitting, cross-validation, and basic PyTorch. You should be able to receive a dataset, build a model, evaluate it, and improve it.

National Level — Competitive AI

Move aggressively into ensemble methods, SVMs, PCA, deep learning (CNNs, backpropagation, embeddings, optimization), and fine-tuning. Start seriously tackling computer vision and NLP pipelines with GPU acceleration and error analysis.

International (IOAI) Level — Complete Modern Toolkit

Requires the full breadth of the 2026 syllabus: advanced deep learning, transformers, object detection, diffusion models, audio models (Whisper, HuBERT), and multimodal inference. You must be able to reason about AI, implement unfamiliar solutions, and optimize them under strict competition constraints.

Final Priority Map

S-Tier — Master First (Non-Negotiable)

- Foundations: Python, NumPy, Pandas, Data visualization, PyTorch, Tensor manipulation.
- Mathematics: Linear algebra, Probability, Statistics, Gradients, Optimization.
- Classical ML: Linear/Logistic regression, Trees, Random Forests, Gradient Boosting, SVM, KNN, K-means, PCA.
- Data Science: Data cleaning, Feature engineering, Train/Val/Test splits, Metrics, Cross-validation.
- Deep Learning: Perceptrons, MLP, Activations, Loss functions, Gradient descent, Backpropagation, Embeddings, Fine-tuning.

A-Tier — Very Important

- CNNs, Computer Vision, ResNet, Object Detection, Image Segmentation.
- NLP, BERT, Attention, Transformers, Autoencoders.
- Model Regularization, Batch Normalization, Self-supervised learning, CLIP, GANs.

B-Tier — International Specialization

- Diffusion models, YOLO/SSD/DETR, U-Net, Vision-language models.
- Advanced embeddings, Language modeling, Encoder-decoder architectures.
- Audio models (HuBERT, Whisper, Qwen-Audio, Voxtral), Multimodal AI, Time-series data.

Exclusions — Do Not Waste Early Preparation On

The official syllabus explicitly states that topics not listed are excluded. Do not prioritize web development, databases, operating systems, cybersecurity, building huge LLMs from scratch, advanced CUDA programming, or research-level transformer theory. They are not the core of the IOAI.

The Clean Master Checklist

Domain	Core Topics
Programming	Python, NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch, Tensors, CPU/GPU
Mathematics	Linear algebra, Probability, Statistics, Derivatives, Gradients, Optimization
Data Science	Cleaning, Normalization, Train/Val/Test, Augmentation, Feature Eng, Tokenization
Classical ML	Regression, KNN, Trees, Random Forests, Gradient Boosting, SVM
Unsupervised	K-means, PCA, t-SNE, UMAP, DBSCAN, Hierarchical/Spectral clustering
Evaluation	Accuracy, Precision/Recall, F1, ROC/AUC, Overfitting, Cross-validation
Neural Nets	Perceptrons, Activations, Loss, Backpropagation, MLP, Embeddings, Pooling
Deep Learning	SGD, Adam, Dropout, Early stopping, Batch norm, Fine-tuning, Autoencoders
Vision	CNNs, Detection (YOLO/DETR), Segmentation, ResNet, GANs, CLIP, Diffusion
NLP & Audio	Tokenization, BERT, Transformers, LMs, HuBERT, Whisper, Qwen-Audio
Multimodal	Cross-modal embeddings, Vision-language, Text/Image/Audio/Video integration
Practical AI	Baselines, Model selection, Error analysis, Hyperparameter optimization

Conclusion & The Most Important Difference

The AI Olympiad differs fundamentally from other academic competitions. Mathematics asks you to prove something; Chemistry asks you to predict a system; Biology asks you to analyze a system; Informatics asks you to design an efficient algorithm. The IOAI asks you to: Given data and a problem, build or choose an AI method, implement it, evaluate it, and improve it.

This is why the IOAI deliberately divides its syllabus into Theory, Practice, and Both. You do not need to become an AI researcher before becoming competitive. The cleanest route from NEB Grade 12 to the IOAI is: Python → NumPy/Pandas → Math → Classical ML → Data Science → PyTorch → Neural Networks → CNNs → NLP/ Transformers → Fine-tuning → Vision/Audio → Multimodal → IOAI problems. By systematically mastering this roadmap, a dedicated student can develop the practical engineering and analytical depth required to excel on the global stage.

References and Recommended Literature

[1] International Olympiad in Artificial Intelligence (IOAI). Official IOAI 2026 Syllabus and Community Meeting Highlights. IOAI Secretariat, 2026.

[2] Himalayan Computational Linguistics Olympiad. National Selection Framework for IOAI. Nepal, 2026.

[3] Géron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2022.

[4] Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. Dive into Deep Learning. Cambridge University Press, 2023.

[5] Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press, 2016.

[6] Ministry of Education, Curriculum Development Centre (CDC). Secondary Level Computer Science Curriculum. Government of Nepal, 2078 (2021).