



Model Curriculum

Fundamentals of Data Annotation using Python

SNOS Code: NIE/SSC/N1019

Version: 1.0

NSQF Level: 3.5

Training Parameters

Sector	IT-ITeS	
Sub-Sector	Artificial Intelligence	
Occupation	AI Development	
Country	India	
NSQF Level	3.5	
Aligned to NCO/ISCO/ISIC Code	NCO/2015-2512.0205	
Minimum Educational Qualification and Experience	a. Entry Qualification & Relevant Experience:	
	Academic/Skill Qualification (with Specialization – if applicable)	Required Experience (with Specialization - if applicable)
	Grade 10 pass and pursuing continuous schooling*	No Experience required
	Grade 8 with 2 years of NTC in the field of IT Sector*	No Experience required
	Grade 8th pass*	3-Year Relevant Experience in IT
	Previous NSQF Level 3 Required	1.5-year relevant experience in IT
Prerequisite: Fundamentals of Python *Should have a basic understanding of coding.		
Pre-Requisite License or Training	NA	
Minimum Job Entry Age	16 Year	
Last Reviewed On	18.02.2025	
Next Review Date	18.02.2028	
NSQC Approval Date	18.02.2025	
SNOS Version	1.0	
Model Curriculum Creation Date	18.02.2025	
Model Curriculum Valid Up to Date	18.02.2028	
Model Curriculum Version	1.0	
Duration of the Course	120 Hours	

Program Overview

This section summarises the end objectives of the program along with its duration

Training Outcomes of the course:

By completing this course, learners will gain proficiency in Python programming, Concepts of Data science and AI, and hands-on expertise in annotating text, images, video, and audio. They will master AI-assisted tools for annotation tasks and real-world projects implementation skills, while enhancing essential soft skills for success in the data annotation field.

Compulsory Modules

The table lists the modules and their duration corresponding to the Compulsory NOS of the Qualification file.

NOS/Module Name	Training Duration (Hours)		
	Theory	Practical	Total
Fundamentals of Data Annotation using Python NOS Code: NIE/SSC/N1019 NSQF Level: 3.5	38	82	120
Total	38	82	120

Module Details

Module1/NOS1: Fundamentals of Data Annotation using Python

NOS Code: NIE/SSC/N1019

Terminal Outcomes:

1. **Foundational in Python Programming:** Participants will achieve proficiency in Python, mastering programming concepts and libraries to independently develop applications.
2. **Basics of Artificial Intelligence & Data Science:** Learners will understand and apply key AI and data science concepts across multiple industries.
3. **Introduction to Data Annotation:** Students will master data annotation techniques, impacting machine learning models across various media.
4. **Text Annotation:** Participants will develop advanced text annotation skills, enhancing their data processing and analysis for AI.
5. **Image and Video Annotation:** Learners will master object detection and tracking, crucial for autonomous driving and surveillance applications.
6. **Audio Annotation:** Students will learn critical audio annotation techniques for developing effective voice-activated systems.
7. **Emerging Trends in AI-Assisted Annotation and Best Practices:** Participants will explore and prepare for future AI applications with advanced techniques like 3D annotation.
8. **Quality Control and Best Practices in AI:** Learners will implement AI and annotation best practices, focusing on accuracy and ethical standards.
9. **Application of Data Annotation:** Through real-world projects, students will demonstrate their annotation proficiency and teamwork in industry settings.

Detailed Syllabus:

Name of the Topics	Theory (Hours)	Practical (Hours)	Total (Hours)
Foundational in Python Programming	10	20	30
Basics of Artificial Intelligence & Data Science	5	10	15
Introduction to Data Annotation	6	9	15
Text Annotation	2.5	5	7.5
Image & Video Annotation	6	9	15
Audio Annotation	2.5	5	7.5
Emerging trends in AI-Assisted Annotation and best practices	6	9	15
Application of Data Annotation	-	15	15
Total Duration (in Hrs.)	38	82	120

Foundational in Python Programming (30 hours)

- **Theory (10 Hours)**
 - Understanding how to set up the Python environment.
 - Learning about variables and different data types in Python.
 - Understanding conditional statements (if, else, elif) and learning about loops (for, while).
 - Understanding the concepts of functions and learning to define and call functions.
 - Understanding the basics of file handling in Python.
 - Understanding Python lists (creation, indexing, slicing, etc.) and dictionaries (creation, accessing, etc.), their properties, and uses.
 - Understanding the basics of Python libraries - NumPy, Matplotlib, and Pandas.
- **Practical (20 Hours)**
 - Learning to install Python and an IDE.
 - Understanding how to write and execute simple Python scripts such as "Hello, World!"
 - Learning to perform basic operations with variables and data types.
 - Learning to write programs using conditional statements and loops.
 - Learning to create and use functions in Python programs.
 - Learning to read from and write to files using Python.
 - Learning to perform operations with lists and dictionaries in Python.
 - Learning to perform basic operations using various Python libraries.
 - Developing and implementing a project, integrating various concepts learned.

Basics of Artificial Intelligence & Data Science (15 Hours)

- **Theory (5 Hours)**
 - **Introduction to Artificial Intelligence**
 - Understanding the basic concepts and evolution of Artificial Intelligence.
 - Understanding the key components of Artificial Intelligence: Machine Learning, Deep Learning, Computer Vision, and Natural Language Processing (NLP).
 - AI in Healthcare, Finance, Education, and Entertainment: Use cases & applications.
 - Introduction to Generative AI – tools and use cases.
 - **Basic Data Science and Statistics**
 - Introduction to Data Science concepts.
 - Learning to handle large & complex datasets.
 - Learning to implement data preprocessing techniques: Cleaning, Normalization, Handling Missing Data on sample datasets.
 - Understanding statistical concepts: Distributions, Variance, Data Sampling.
 - **Ethical Considerations in AI**
 - Introduction to AI Ethics and biasness in AI models.
 - Understanding Privacy Concerns with the use of AI.
 - Understanding safe and responsible use of AI.
- **Practical (10 Hours)**
 - Demonstration of popular AI tools.
 - Demonstration of AI use cases in Healthcare, Finance, Education, and Entertainment.
 - Hands-on experience with open-source Generative AI tools.
 - Introduction to low-code/no-code AI platforms.
 - Various test cases to use preprocessing techniques: data cleaning, normalization, and handling missing data.

Introduction to Data Annotation (15 Hours)

- **Theory (6 Hours)**
 - **Overview of Data Annotation**
 - Definition and scope of data annotation.
 - Understanding the difference between supervised and unsupervised learning.
 - Learning about the importance and impact of data annotation on the business.
 - Understanding the use cases and applications of data annotation across various industry verticals.
 - Introduction to various data annotation methods and understanding major differences between them.
 - Overview of Text, Image, Video, and Audio annotation.
 - Understanding how to handle various datasets: large-scale, complex datasets with limited labeled data.
- **Practical (9 Hours)**
 - Hands-on experience on the features and functionalities of various open-source Data Annotation tools.
 - Techniques to prepare datasets for annotation and machine learning tasks.

- Understanding how to identify, tag, and label distinct elements in huge datasets.
- Hands-on practice with different data types: Text, Audio, Video, Image.
- Learning to optimize the annotation process by focusing on the most informative samples.

Text Annotation (7.5 Hours)

- **Theory (2.5 Hours)**
 - **Understanding Basics of Text Annotation**
 - Definition and common use cases.
 - Understanding various tools used for Text Annotation.
 - Learning about various methods to classify text and Annotation schemas.
 - Techniques and applications of performing Named Entity Recognition (NER).
 - Learning to assign grammatical weighted categories to various words in a text format.
 - Understanding various methods of sentiment analysis.
- **Practical (5 Hours)**
 - Demonstrate the difference between various annotation methods and techniques, based on the project requirements and datasets.
 - Practical exercises using open-source text annotation tools.

Image and Video Annotation (15 Hours)

Image Annotation (7 Hours)

- **Theory (3 Hours)**
 - Understanding Basics of Image Annotation: definition and common use cases.
 - Learning the concepts of Image Annotation: Image segmentation, object detection, Bounding box, pixel-based polygon annotation.
- **Practical (4 Hours)**
 - Demonstrate the difference between various annotation methods and techniques, based on the project requirements and datasets.
 - Learning various techniques and tools used for Image Segmentation & Object Detection.
 - Learning various techniques of drawing bounding box in deep learning.
 - Understanding how to use precise pixel-based polygon annotation.
 - Practical exercises using open-source image annotation tools.

Video Annotation (8 Hours)

- **Theory (3 Hours)**
 - Understanding Basics of Video Annotation: definition and common use cases.
 - Learning about Object Tracking: Target initialization, appearance modeling, motion estimation, target positioning, Frame-by-Frame Annotation.
- **Practical (5 Hours)**
 - Practical exercises on various techniques for scale variations, annotating out-of-view objects, low resolution, occlusion object tracking, & Multi-object tracking.
 - Hands-on practical experiment on open-source tools.

Audio Annotation (7.5 Hours)

- **Theory (2.5 Hours)**
 - Understanding basics of Audio Annotation: Definition and importance.
 - Learning about various types of Audio Data: Signals & formats used in annotation.
 - Understanding various techniques in Speech-to-Text conversion, multi-Speaker diarization, & Emotion Detection.
- **Practical (5 Hours)**
 - Practical exercises using open-source audio annotation tools.
 - Hands-on experience on various techniques in Speech-to-Text conversion, multi-Speaker diarization, & Emotion Detection.
 - Demonstrate the difference between various annotation methods and techniques, based on the project requirements and dataset.

Emerging Trends in AI-Assisted Annotation and Best Practices (15 Hours)

AI-Assisted Annotation and Emerging Trends (8 Hours)

- **Theory (3 Hours)**
 - Introduction to cutting-edge trends such as 3D annotation for autonomous vehicles, synthetic data generation.
 - Understanding the use of Augmented & Virtual Reality in annotation tasks.
 - Understanding how to review pre-annotated data from various AI tools and refining the outputs.
- **Practical (5 Hours)**
 - Hands-on practical exercises using open-source AI-assisted tools.
 - Learning about various techniques for ensuring high-quality annotations using automated tools—Text, Image, Video, & Audio.
 - Learning the use of augmented and virtual reality in annotation tasks.
 - Practical learning to cutting-edge annotation trends—3D annotation.

Quality Control and Best Practices in AI (7 Hours)

- **Theory (3 Hours)**
 - Understanding techniques for identification and mitigation of various common errors in annotation.
 - Learning to use best practices and the guidelines for effective annotation.
 - Understanding ethical considerations: Ethical guidelines and best practices in Data Annotation.
 - Understanding AI Ethics and Responsible Annotation Practices.
 - Learning techniques for Data Privacy Regulations and data anonymization.
- **Practical (4 Hours)**
 - Demonstrating the difference between various annotation methods and techniques, based on the project requirements and dataset.
 - Learning to implement quality assurance techniques on sample datasets.
 - Learning strategies for difficult or unclear annotation scenarios.
 - Learning to handle edge cases and ambiguity in data annotation.

Application of Data Annotation (Practical 15 Hours)

- Working on real-world data annotation problems sourced from industry partners.
- Learning to apply learned skills and knowledge in real-world scenarios.
- Learning to collaborate and work as a team for common objectives.
- Presenting the project findings to peers and instructors and learning to incorporate constructive feedback.

Key Learning Outcomes:

Duration: 38 hours	Duration: 82 hours
Theory	Practical
Learners will understand setting up a Python environment, and grasp basic programming concepts and Python libraries.	Learners will effectively install Python, execute scripts, and develop applications using loops, functions, and file handling techniques.
Learners will understand AI fundamentals, including machine learning, deep learning, and applications across various fields.	Learners will utilize AI tools in real-world applications, practicing data pre-processing and exploring generative AI platforms.
Learners will learn the scope of data annotation, differences between learning types, and its industry applications.	Learners will perform hands-on data annotation across text, images, video, and audio, optimizing annotation processes using industry tools.
Learners will grasp text annotation basics, including methods for text classification and named entity recognition.	Learners will use open-source tools for text annotation, applying methods like sentiment analysis and entity recognition in practical scenarios.
Learners will learn image and video annotation fundamentals, including segmentation, object detection, and tracking techniques.	Learners will practice advanced techniques in image and video annotation, applying object detection and tracking in practical exercises.
Learners will understand audio annotation basics, including speech-to-text conversion and speaker diarization techniques.	Learners will execute audio annotation projects using tools for speech-to-text conversion and emotion detection, refining data handling skills.
Learners will be introduced to advanced trends like 3D annotation, synthetic data, and augmented reality in annotations.	Learners will operate AI-assisted tools for tasks like 3D annotation and use augmented reality in practical annotation settings.
Learners will learn best practices for annotation, error mitigation techniques, and ethical considerations in AI.	Learners will implement quality control strategies in annotation projects, applying ethical practices and addressing ambiguous data scenarios.
	Learners will tackle real-world annotation problems from industry partners, developing projects that apply their theoretical knowledge and collaborative
Classroom Aid	
The aids required to conduct sessions in the classroom are: 1. LCD Projector 2. Pin-up Board 3. White Board	
Tools, Equipment and Other Requirements	

- Classroom, Students Chair, Students Table, Desktop computer with accessories
- IDEs like PyCharm, Jupyter, AI libraries (TensorFlow, Scikit-learn), Annotation tools like Labelbox, Prodigy, video/image/audio editors, Computer Vision Annotation Tool (CVAT) project management tools, and soft skills training resources.

Annexure

Trainer Requirements

Trainer Prerequisites						
Minimum Educational Qualification	Specialization	Relevant Industry Experience		Training/Assessment Experience		Remarks
		Years	Specialization	Years	Specialization	
B. Tech in Computer Science/ IT/ Electronics and Communication/ allied branches Or MCA Or BCA Or A-Level (IT)	Computer Science/ IT/ Electronics and Communication/ allied branches	1	-	-	-	-

Trainer Certification	
Domain Certification	Platform Certification
Certified to TOT for the Standalone NOS: NIE/SSC/N1019 "Fundamentals of Data Annotation using Python" or equivalent as per NCrF. Minimum accepted score is 80%	Recommended that the Trainer is certified for the Job Role: "Trainer (VET and skills)", mapped to the Qualification Pack: "MEP/Q2601, v2.0" or equivalent as per NCrF. Minimum accepted score is 80%

Assessor Requirements

Assessor Prerequisites						
Minimum Educational Qualification	Specialization	Relevant Industry Experience		Training/Assessment Experience		Remarks
		Years	Specialization	Years	Specialization	
B. Tech in Computer	Computer Science/ IT/	2	-	-	-	-

Science/ IT/ Electronics and Communication / allied branches Or MCA Or BCA Or A- Level (IT)	Electronics and Communica tion/ allied branches					
---	---	--	--	--	--	--

Assessor Certification	
Domain Certification	Platform Certification
<p>Certified to TOA for the Standalone NOS: NIE/SSC/N1019 “Fundamentals of Data Annotation using Python” or equivalent as per NCrF.</p> <p>Minimum accepted score is 80%</p>	<p>Recommended that the assessor is certified for the Job Role: “Trainer (VET and skills)”, mapped to the Qualification Pack: “MEP/Q2701, v2.0” or equivalent as per NCrF.</p> <p>Minimum accepted score is 80%</p>

Assessment Strategy

This section includes the processes involved in identifying, gathering, and interpreting information to evaluate the Candidate on the required competencies of the program.

Assessment of the qualification evaluates candidates to ascertain that they can integrate knowledge, skills and values for carrying out relevant tasks as per the defined learning outcomes and assessment criteria.

The underlying principle of assessment is fairness and transparency. The evidence of the outcomes and assessment criteria. competence acquired by the candidate can be obtained by conducting Theory (Online) examination.

About Examination Pattern:

1. The question papers for the theory exams are set by the Examination wing (assessor) of NIELIT HQS.
2. The assessor assigns roll number.
3. The assessor carries out theory online assessments. Theory examination would be conducted online and the paper comprise of MCQ
4. Pass percentage would be 50% marks.
5. The examination will be conducted in English language only.

Quality assurance activities: A pool of questions is created by a subject matter expert and moderated by other SME. Test rules are set beforehand. Random set of questions which are according to syllabus appears which may differ from candidate to candidate. Confidentiality and impartiality are maintained during all the examination and evaluation processes.

References

Glossary

Term	Description
Declarative Knowledge	Declarative knowledge refers to facts, concepts and principles that need to be known and/or understood in order to accomplish a task or to solve a problem.
Key Learning Outcome	Key learning outcome is the statement of what a learner needs to know, understand and be able to do in order to achieve the terminal outcomes. A set of key learning outcomes will make up the training outcomes. Training outcome is specified in terms of knowledge, understanding (theory) and skills (practical application).
OJT (M)	On-the-job training (Mandatory); trainees are mandated to complete specified hours of training on site
OJT (R)	On-the-job training (Recommended); trainees are recommended the specified hours of training on site
Procedural Knowledge	Procedural knowledge addresses how to do something, or how to perform a task. It is the ability to work, or produce a tangible work output by applying cognitive, affective or psychomotor skills.
Training Outcome	Training outcome is a statement of what a learner will know, understand and be able to do upon the completion of the training.
Terminal Outcome	Terminal outcome is a statement of what a learner will know, understand and be able to do upon the completion of a module. A set of terminal outcomes help to achieve the training outcome.

Acronyms and Abbreviations

Term	Description
QP	Qualification Pack
NSQF	National Skills Qualification Framework
NSQC	National Skills Qualification Committee
NOS	National Occupational Standards