School Connect Online

SCO INTERNATIONAL OLYMPIAD

CLASS 7 ARTIFICIAL INTELLIGENCE OLYMPIAD

Sample Question Paper for students, teachers, schools, and parents

Designed from Class 7 Artificial Intelligence learning pathways and aligned with SCO's guided preparation, practice, reporting, and future-ready academic growth.

- age-fit AI literacy for middle-school learners globally
- objective MCQ practice across machine learning, datasets, privacy, AI applications, and ethics
- answer key and explanations for guided revision and classroom discussion

Maths	English	Science	Mental Ability	Finance Knowledge
AI	Entrepreneurship	GK	Coding	Life Skills

School Connect Online

SCO International Artificial Intelligence Olympiad - Class 7

Sample Question Paper | Question Paper Set S

Total Questions	Time	Question Type	Sections
50	60 minutes	Objective MCQ	4

Guidelines for the Candidate

1. Before the exam begins, candidates may use the additional time given by the invigilator to complete OMR or identity information.
2. Clearly write name, school code, class, registration ID, and contact number wherever required.
3. This paper contains 50 objective-type questions. Each question has only one correct answer.
4. All questions should be attempted. There is no negative marking unless separately notified by the exam authority.
5. Calculators, mobile phones, smart watches, books, notes, and external help are not allowed.
6. Use only the permitted pen or pencil to mark answers on the OMR sheet or answer sheet.
7. At the conclusion of the test, hand over the OMR sheet or answer sheet to the invigilator.

Name:	Registration ID: .....
School:	Contact No.: .....

School Connect Online

Section 1 - Machine Learning Basics

Q.1

A machine learning algorithm is trained on labeled data to predict if an email is "spam" or "not spam." What type of learning is this?

A. Reinforcement Learning

B. Supervised Learning

C. Unsupervised Learning

D. Semi-supervised Learning

Answer: B

Explanation: The algorithm uses labeled data (spam or not spam) for training, which is the essence of supervised learning.

Q.2

Which of the following is an example of unsupervised learning?

A. Classifying images of cats and dogs

B. Grouping customers based on purchasing habits

C. Predicting the next word in a sentence

D. Determining whether a transaction is fraudulent

Answer: B

Explanation: Unsupervised learning identifies patterns or groups in unlabeled data, as in customer segmentation.

Q.3

What will the following Python code output?

```
from sklearn.linear_model import LinearRegression

model = LinearRegression()

X = [[1], [2], [3], [4]]

y = [2, 4, 6, 8]
```

model.fit(X, y)

```
print(model.predict([[5]]))
```

A. 10

B. 5

C. 12

D. 8

Answer: A

Explanation: The linear regression model learns the relationship y = 2x. For x = 5, the prediction is y = 2 × 5 = 10.

Q.4

In a supervised learning problem using a decision tree, what is the criterion typically used to split nodes?

A. Gradient Descent

B. Entropy

C. Clustering

D. Reinforcement

Answer: B

Explanation: Entropy measures the impurity in data splits, which is minimized to create the best split in decision trees.

Q.5

A student wants to separate similar shapes into groups without knowing their labels. What technique should they use?

A. Clustering

B. Regression

C. Reinforcement Learning

D. Classification

Answer: A

Explanation: Clustering is an unsupervised learning technique used to group data based on similarities.

Q.6

Which machine learning algorithm is best suited for supervised learning with continuous output?

A. K-Means

B. Support Vector Machines

C. Linear Regression

D. Principal Component Analysis

Answer: C

Explanation: Linear regression predicts continuous output, such as predicting house prices.

Q.7

What will the following code output?

```
from sklearn.neighbors import KNeighborsClassifier
X = [[1, 2], [2, 3], [3, 4]]
y = [0, 1, 0]
model = KNeighborsClassifier(n_neighbors=1)
```

model.fit(X, y)

```
print(model.predict([[2, 2]]))
```

A. 0

B. 1

C. 2

D. 3

Answer: A

Explanation: The KNN classifier predicts the class of the nearest neighbor. For point [2,2], the nearest neighbor is [1,2], which belongs to class 0.

Q.8

What is the purpose of normalizing data in machine learning?

A. To make data easier to read

B. To reduce overfitting

C. To scale features to a uniform range

D. To increase dataset size

Answer: C

Explanation: Normalizing data ensures all features contribute equally by scaling them to a common range.

Q.9

What will the following code output?

```
from sklearn.preprocessing import LabelEncoder

labels = ['cat', 'dog', 'dog', 'cat']

encoder = LabelEncoder()

encoded = encoder.fit_transform(labels)

print(encoded)
```

A. [0, 1, 1, 0]

B. [1, 2, 2, 1]

C. [2, 0, 0, 2]

D. [0, 0, 1, 1]

Answer: A

Explanation: The LabelEncoder assigns a unique integer to each class in alphabetical order.

Q.10

What will this code output?

```
import numpy as np

from sklearn.linear_model import LogisticRegression

X = np.array([[1], [2], [3], [4]])

y = np.array([0, 0, 1, 1])

model = LogisticRegression()
```

model.fit(X, y)

```
print(model.predict([[2.5]]))
```

A. 0

B. 1

C. 2

D. Error

Answer: B

Explanation: Logistic regression predicts probabilities for classes. For X=2.5X = 2.5X=2.5, the model predicts class 1.

Q.11

Why is feature selection important in machine learning?

A. To add more features to the dataset

B. To reduce computational cost and improve model performance

C. To increase the number of data samples

D. To apply supervised learning

Answer: B

Explanation: Feature selection removes irrelevant data, improving efficiency and accuracy.

Q.12

What is the primary cause of overfitting in machine learning?

A. Too much data

B. A simple model

C. A complex model

D. Random data

Answer: C

Explanation: Complex models can memorize training data, failing to generalize to unseen data.

Q.13

What will this code output?

```
from sklearn.metrics import mean_squared_error

y_true = [3, -0.5, 2, 7]

y_pred = [2.5, 0.0, 2, 8]

mse = mean_squared_error(y_true, y_pred)

print(mse)
```

A. 0.25

B. 0.375

C. 0.5

D. 0.75

Answer: B

Explanation: The Mean Squared Error is calculated as (1/n) sum((y_true - y_pred)^2)\frac{1}{n} \sum (y_{\text{true}} - y_{\text{pred}})^2n1∑(ytrue−ypred)2.

Q.14

Which of the following is an example of supervised learning in real life?

A. Grouping images of galaxies based on similarity

B. Predicting whether a customer will buy a product

C. Creating playlists based on music listening habits

D. Segmenting customers by behavior

Answer: B

Explanation: Supervised learning uses labeled data for predictions.

Q.15

What will this code output?

```
from sklearn.cluster import KMeans

import numpy as np
```

```
X = np.array([[1, 2], [2, 3], [3, 4], [8, 8]])

kmeans = KMeans(n_clusters=2, random_state=0).fit(X)

print(kmeans.labels_)
```

A. [0, 0, 0, 1]

B. [1, 1, 1, 0]

C. [0, 0, 1, 1]

D. [1, 1, 0, 0]

Answer: A

Explanation: KMeans clusters the data into two groups based on proximity.

Section 2 - Data Collection and Privacy

Q.16

Which of the following is the most efficient method for collecting large amounts of data in real time?

A. Surveys

B. IoT Sensors

C. Interviews

D. Focus Groups

Answer: B

Explanation: IoT sensors can collect data automatically and in real time, unlike surveys or interviews, which are manual and time-consuming.

Q.17

What will the following SQL query return?

```
SELECT name, age

FROM students

WHERE age > 12 AND grade = '7';
```

A. Names and ages of all students

B. Names and ages of students in grade 7

C. Names and ages of students older than 12 in grade 7

D. Names and ages of students younger than 12 in grade 7

Answer: C

Explanation: The query filters students by age > 12 and grade = 7, returning their names and ages.

Q.18

What does the term "data anonymization" mean?

A. Encrypting data to prevent access

B. Removing personal identifiers from data

C. Restricting access to data

D. Collecting data without user consent

Answer: B

Explanation: Data anonymization removes information that can link data to specific individuals, protecting privacy.

Q.19

What will this Python code output?

```
database = {"Alice": 12, "Bob": 14, "Charlie": 13}

database["Alice"] = 15

print(database["Alice"])
```

A. 12

B. 13

C. 15

D. KeyError

Answer: C

Explanation: The code updates Alice's age in the dictionary to 15 and prints the updated value.

Q.20

What will this SQL query return?

```
SELECT AVG(salary)
FROM employees
WHERE department = 'AI';
```

A. The total salary of all employees

B. The average salary of all employees in the AI department

C. A list of salaries in the AI department

D. The highest salary in the AI department

Answer: B

Explanation: The AVG() function computes the average of salaries in the AI department.

Q.21

Which of these is a violation of data privacy?

A. Encrypting sensitive data before storage

B. Sharing user data without consent

C. Collecting anonymized user preferences

D. Using data for research after user consent

Answer: B

Explanation: Sharing user data without their consent is a direct violation of privacy laws and regulations.

Q.22

What is the purpose of hashing in protecting data privacy?

A. To encrypt data for secure transmission

B. To transform data into a unique fixed-size string

C. To store data in a relational database

D. To organize data for machine learning

Answer: B

Explanation: Hashing converts data into a unique string, ensuring sensitive information cannot be retrieved directly.

Q.23

A company wants to ensure GDPR compliance. Which of the following actions is NOT compliant?

A. Allowing users to delete their data on request

B. Collecting data without informing users

C. Encrypting sensitive user information

D. Asking for user consent before collecting data

Answer: B

Explanation: GDPR requires users to be informed about data collection and usage; collecting data secretly violates compliance.

Q.24

What does this Python code simulate?

```
import hashlib

data = "student_password"

hashed_data = hashlib.sha256(data.encode()).hexdigest()

print(hashed_data)
```

A. Encrypting a password

B. Hashing a password

C. Storing a password in plain text

D. Retrieving a password

Answer: B

Explanation: The code uses the sha256 algorithm to hash the password, converting it into a secure, irreversible string.

Q.25

SQL Injection Attack Prevention

Which method best prevents SQL injection attacks in a database?

A. Using simple string concatenation in queries

B. Storing passwords in plain text

C. Using parameterized queries

D. Allowing unrestricted user inputs

Answer: C

Explanation: Parameterized queries prevent malicious input from altering SQL statements, protecting the database from injection attacks.

Section 3 - AI Applications in the Real World

Q.26

A hospital is using AI to predict patient diseases based on their medical history. The AI model classifies patients into "High Risk," "Moderate Risk," and "Low Risk" categories. If the model fails to predict high-risk patients accurately, which type of error is most harmful in this case?

A. False Positive

B. False Negative

C. True Positive

D. True Negative

Answer: B

Explanation: A false negative means the model fails to identify high-risk patients, potentially delaying life-saving treatments.

Q.27

An AI learning platform analyzes student performance and suggests personalized learning paths. If a student keeps getting questions about advanced topics despite struggling with basics, what AI component might need improvement?

A. Natural Language Processing

B. Recommendation System

C. Image Recognition

D. Speech-to-Text Conversion

Answer: B

Explanation: The recommendation system is responsible for tailoring content to a student's needs. It should detect the student's struggles and adjust the difficulty accordingly.

Q.28

AI is used to monitor deforestation through satellite imagery. The model detects areas with significant tree loss by analyzing changes in color and patterns. What type of AI is being used here?

A. Rule-Based Systems

B. Computer Vision

C. Natural Language Processing

D. Predictive Analytics

Answer: B

Explanation: Computer vision allows AI to analyze and interpret visual data like satellite images to detect deforestation.

Q.29

A healthcare AI system identifies diseases based on X-ray images. The model wrongly predicts a healthy person as having a disease. What is this type of error called?

A. Overfitting

B. Underfitting

C. False Positive

D. False Negative

Answer: C

Explanation: A false positive occurs when the model incorrectly identifies a non-diseased person as diseased.

Q.30

An AI app helps teachers grade essays by detecting grammar mistakes and assessing writing style. If the app struggles with regional dialects and slang, which AI feature requires improvement?

A. Sentiment Analysis

B. Neural Networks

C. Contextual Understanding

D. Data Compression

Answer: C

Explanation: AI must understand context and regional variations to accurately assess language in diverse essays.

Q.31

AI models predict air pollution levels based on historical data, weather patterns, and traffic. If the model fails to consider sudden industrial emissions, which factor needs improvement?

A. Dataset Size

B. Real-Time Data Integration

C. Labeling Accuracy

D. Image Preprocessing

Answer: B

Explanation: For accurate predictions, the model must integrate real-time data like sudden industrial emissions.

Q.32

A robotic surgery AI learns by observing hundreds of surgeries. What type of machine learning is this?

A. Supervised Learning

B. Reinforcement Learning

C. Unsupervised Learning

D. Deep Learning

Answer: A

Explanation: The AI learns using labeled data (surgery outcomes) to make accurate decisions.

Q.33

An AI tutor creates quizzes for students based on their mistakes in previous tests. Which approach is the AI most likely using?

A. Decision Trees

B. Natural Language Processing

C. Personalized Learning Algorithms

D. Image Recognition

Answer: C

Explanation: Personalized learning algorithms adapt quizzes based on individual student needs and progress.

Q.34

AI-powered drones monitor wildlife populations by counting animals in videos. Which AI domain is applied here?

A. Data Analytics

B. Computer Vision

C. Natural Language Processing

D. Predictive Modeling

Answer: B

Explanation: Computer vision processes and analyzes video frames to identify and count animals.

Q.35

An AI chatbot answers patient questions about medications. If it misunderstands the meaning of “Can I take this at night?” and provides a wrong response, what is the likely issue?

A. Lack of Data

B. Poor Natural Language Processing

C. Weak Encryption

D. Overfitting

Answer: B

Explanation: NLP helps the chatbot understand and respond correctly to human language.

Q.36

AI predicts floods using weather data and river levels. If the system works well in one region but not another, what might be the issue?

A. Model Overfitting

B. Model Generalization

C. Insufficient Training Data

D. All of the above

Answer: D

Explanation: Overfitting, poor generalization, and insufficient training data can all lead to region-specific inaccuracies.

Q.37

AI detects diseases by analyzing DNA. If the model uses a dataset from adults but applies it to children, what is the risk?

A. Underfitting

B. Bias in Data

C. Data Redundancy

D. Real-Time Errors

Answer: B

Explanation: The model may be biased as the training data (adults) doesn't represent the new population (children).

Q.38

An AI grading system gives low scores to essays with creative ideas but grammatical mistakes. How can this issue be resolved?

A. Train the model with creative essays

B. Improve grammar detection algorithms

C. Use a multi-factor grading system

D. All of the above

Answer: D

Explanation: Training on diverse data, enhancing algorithms, and using multi-factor evaluation can address the issue.

Q.39

AI helps farmers optimize water usage by predicting rainfall. If farmers face crop loss due to inaccurate predictions, what should the AI improve?

A. Real-Time Weather Data

B. Image Processing

C. Sensor Calibration

D. Natural Language Understanding

Answer: A

Explanation: Incorporating real-time data ensures better prediction accuracy for farming decisions.

Q.40

A hospital uses AI for drug discovery. The AI identifies a compound that works well in simulations but fails in real-world trials. What is the likely cause?

A. Overfitting in Simulations

B. Poor Data Labeling

C. Insufficient Training Data

D. Incorrect Model Selection

Answer: A

Explanation: Overfitting means the model performs well on training data (simulations) but fails to generalize in real-world conditions.

Achievers Section - Case Studies and Ethical Data Usage

Q.41

A student builds an AI chatbot to collect user preferences for recommending books. The chatbot also stores user names and email addresses without permission. Which ethical principle is violated?

A. Transparency

B. Data Minimization

C. Fairness

D. Bias Detection

Answer: B

Explanation: Data minimization ensures that only necessary data is collected. Collecting user names and email addresses without consent violates this principle.

Q.42

A teacher creates a Python program to predict student grades using AI. The dataset includes grades from only one school. What issue is likely to occur?

```
import numpy as np

from sklearn.linear_model import LinearRegression

# Dataset

X = np.array([[10, 8], [9, 7], [10, 10]])  # Study hours, attendance

y = np.array([90, 85, 95])  # Grades

# Train model

model = LinearRegression().fit(X, y)
```

A. Algorithm bias

B. Overfitting

C. Data Redundancy

D. Underfitting

Answer: A

Explanation: The dataset is limited to one school, so the model may not generalize to other schools, leading to biased predictions.

Q.43

An AI company uses personal medical data to predict diseases. The company shares the data with third parties without informing users. What should the company implement to ensure ethical data sharing?

A. Anonymization of Data

B. Increasing Model Accuracy

C. More Training Data

D. Data Normalization

Answer: A

Explanation: Anonymizing data ensures that personal details are not identifiable, protecting user privacy while using the data ethically.

Q.44

A student writes a program to clean data for an AI model but forgets to handle missing values. What problem could occur in the following code?

```
import pandas as pd

# Load data

data = pd.read_csv('student_scores.csv')

# Train the model

model = LinearRegression().fit(data[['Hours_Studied']], data['Score'])
```

A. Model bias

B. Data leakage

C. Model crash

D. Overfitting

Answer: C

Explanation: Missing values in the dataset will cause the model training to crash since the AI algorithm requires complete numerical inputs.

Q.45

An AI program for hiring employees discards applications with poor grammar. Many candidates from non-native English-speaking countries are rejected. What ethical issue does this scenario highlight?

A. Data Privacy

B. Unfair Bias

C. Model Complexity

D. Transparency

Answer: B

Explanation: The algorithm unfairly discriminates against non-native English speakers by equating grammar with competency.

Q.46

A student trains an AI model for predicting exam scores but accidentally includes the final exam score in the training data. What issue arises?

```
import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.tree import DecisionTreeClassifier

# Data includes final scores

data = pd.read_csv('student_data.csv')

X = data[['Hours_Studied', 'Attendance', 'Final_Score']]

y = data['Pass/Fail']

# Split and train
```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

```
model = DecisionTreeClassifier().fit(X_train, y_train)
```

A. Overfitting

B. Data Leakage

C. Model Complexity

D. Insufficient Data

Answer: B

Explanation: Including the target variable (Final_Score) in the training features gives the model access to answers, leading to data leakage.

Q.47

An AI-powered app tracks user activities to recommend fitness tips. If the app fails to inform users how their data will be used, which fundamental data ethic is violated?

A. Transparency

B. Data Accessibility

C. Diversity

D. Fairness

Answer: A

Explanation: Transparency requires AI systems to inform users about how their data is collected, used, and shared.

Q.48

A student trains an AI model to classify plant species but uses only a small dataset. The following code consistently misclassifies plants. What is the issue?

```
from sklearn.linear_model import LogisticRegression

# Small dataset

X = [[5.1, 3.5], [4.9, 3.0], [4.7, 3.2]]  # Sepal length, Sepal width

y = ['Species_A', 'Species_B', 'Species_A']

model = LogisticRegression().fit(X, y)
```

A. Overfitting

B. Underfitting

C. Bias in Data

D. Model Complexity

Answer: B

Explanation: A small dataset results in a poorly trained model that cannot capture the patterns necessary to classify plant species accurately.

Q.49

An AI program predicts crop yields using soil data. The model performs well on training data but fails in new regions. What is the primary cause?

A. Overfitting

B. Underfitting

C. Poor Generalization

D. Insufficient Features

Answer: C

Explanation: The model has not learned patterns that generalize well to unseen data, limiting its usefulness in new regions.

Q.50

A developer trains a face recognition AI using publicly available images without user consent. The AI is later sold to security companies. What ethical and legal principle is most violated?

A. Fairness

B. Data Privacy

C. Model Accuracy

D. Model Transparency

Answer: B

Explanation: Using personal images without consent violates user privacy and ethical AI principles.

School Connect Online

Answer Key

Q.No.	Ans.	Q.No.	Ans.	Q.No.	Ans.	Q.No.	Ans.	Q.No.	Ans.
1	**B**	2	**B**	3	**A**	4	**B**	5	**A**
6	**C**	7	**A**	8	**C**	9	**A**	10	**B**
11	**B**	12	**C**	13	**B**	14	**B**	15	**A**
16	**B**	17	**C**	18	**B**	19	**C**	20	**B**
21	**B**	22	**B**	23	**B**	24	**B**	25	**C**
26	**B**	27	**B**	28	**B**	29	**C**	30	**C**
31	**B**	32	**A**	33	**C**	34	**B**	35	**B**
36	**D**	37	**B**	38	**D**	39	**A**	40	**A**
41	**B**	42	**A**	43	**A**	44	**C**	45	**B**
46	**B**	47	**A**	48	**B**	49	**C**	50	**B**

School Connect Online

Space for Rough Work