

Neuralquest — Activity Book

A Spark & Anvil activity book — puzzles, questions, and fun from Neuralquest.

Write in the book, circle your answers, and check the Answer Key at the back.

1. how_machines_learn_gr45

Circle the correct answer

1. What makes machine learning different from regular programming?

(A) Machine learning programs are always smarter than human programmers (B) Instead of following step-by-step instructions written by a programmer, the computer discovers patterns from examples in data (C) Machine learning doesn't need any data or input to work properly (D) Machine learning uses bigger and faster computers than regular programming

2. What is 'training data' in machine learning?

(A) A collection of labeled examples that the computer uses to learn patterns (B) A special file format only used by scientists (C) The instructions a programmer writes in code (D) The electricity that powers the computer's processor

3. A teacher shows a machine learning model 1,000 photos of cats and 1,000 photos of dogs, each labeled correctly. What has the teacher just done?

(A) Uploaded a virus that makes the computer see animals (B) Replaced the computer's memory with animal pictures (C) Programmed the computer with rules about what cats and dogs look like (D) Provided training data so the model can learn to tell cats and dogs apart

4. What is a 'label' in a machine learning dataset?

(A) A password needed to access the dataset (B) A physical sticker placed on the computer's screen (C) The brand name of the software being used (D) A tag that tells the computer what category or value an example belongs to

5. You train a model to recognize apples using only photos of red apples. When you show it a green apple, it says 'not an apple.' What went wrong?

(A) The model needs to be connected to the internet to recognize green objects (B) The training data wasn't diverse enough — it only included red apples, so the model learned that apples must be red (C) The computer doesn't like green things and refuses to identify them (D) Green apples are not real apples, so the model is actually correct

6. What does 'prediction' mean in machine learning?

(A) The model's output when it uses learned patterns to classify or estimate something about new, unseen data (B) A guess made randomly without any basis in data (C) The model predicting exactly what a human would say (D) The model telling the future like a fortune teller

Write the answer

7. Which of these is the best analogy for how machine learning works?

Answer: _____

8. What is a 'feature' in machine learning?

Answer: _____

9. A model trained to predict test scores uses these features: hours studied, hours slept, and number of practice problems completed. Which feature do you think is LEAST useful?

Answer: _____

10. What is 'supervised learning'?

Answer: _____

Match each question to its answer

1. What is 'unsupervised learning'? →

2. You have a dataset of 10,000 customer emails. In supervised learning, you would need to . In unsupervised learning, you would . →

3. Why does a machine learning model need MANY examples to learn well, not just a few? →

4. A model was trained on 50 cat photos and 5,000 dog photos. What problem might this cause? →

5. What is the difference between 'training' a model and 'testing' a model? →

(A) Label each email (spam/not spam) first; let the model group similar emails together without labels

(B) Training is when the model learns from labeled examples; testing is when you check how well it performs on new data it hasn't seen before

(C) The model will be much better at recognizing dogs than cats because it had far more dog examples to learn from

(D) More examples help the model find reliable patterns and avoid being misled by unusual cases

(E) A type of machine learning where the model finds patterns in data that has no labels

2. training_your_first_classifier_gr45

Circle the correct answer

11. What is a 'classifier' in machine learning?

(A) A type of computer chip that makes calculations faster (B) A program that creates new categories from scratch (C) A person who organizes files in a filing cabinet (D) A model that sorts input data into predefined categories, like 'cat or dog' or 'spam or not spam'

12. You're building an image classifier to sort photos of fruits. Which set of categories would work BEST?

(A) Yummy fruit, okay fruit, gross fruit — based on personal taste (B) Apple, banana, orange, grape, strawberry — each is a distinct, recognizable fruit (C) Red things, yellow things, green things — based on color only (D) Round fruit and not-round fruit — only two categories

13. You're collecting training data for a cat vs. dog classifier. Which approach would create the most useful dataset?

(A) Find the one best photo of a cat and the one best photo of a dog (B) Download all photos from one pet store's website (C) Take photos of many different breeds, ages, poses, and backgrounds for both cats and dogs (D) Only use photos of golden retrievers and tabby cats

14. What is 'data augmentation' in machine learning?

(A) Deleting bad data from the training set to make it cleaner (B) Making the data labels bigger so they're easier to read (C) Creating new training examples by modifying existing ones — like flipping, rotating, or adjusting the brightness of photos (D) Adding more computers to make the model train faster

15. You train a rock-paper-scissors classifier using photos of YOUR hands only. When a friend tests it, accuracy drops from 95% to 60%. Why?

(A) Rock-paper-scissors is too hard for any computer to learn (B) The model got tired from too much use and needs to rest (C) The model learned the specific look of your hands rather than the general shapes of rock, paper, and scissors gestures (D) Your friend's computer is too slow to run the model properly

16. What is a 'confusion matrix' in machine learning?

(A) A table that shows how many predictions the model got right and wrong for each category (B) A puzzle game that tests how confused the model gets (C) A diagram showing how confusing the training data is (D) A matrix of random numbers used to initialize the model

Write the answer

17. A classifier's confusion matrix shows it correctly identifies dogs 95% of the time but only identifies cats 40% of the time. What is the most likely cause?

Answer: _____

18. What is a 'false positive' in classification?

Answer: _____

19. For a medical screening test, which error is MORE dangerous: a false positive (saying a healthy person is sick) or a false negative (saying a sick person is healthy)?

Answer: _____

20. What is a 'confidence score' in classification?

Answer: _____

Match each question to its answer

1. A text classifier assigns sentiment labels to movie reviews. It labels 'This movie was not bad at all!' as negative. What tripped it up? →

2. You're training a classifier to identify poisonous mushrooms from photos. Why should you set a HIGH confidence threshold (like 99%) before the model says 'safe to eat'? →

3. What is 'binary classification'? →

4. What is the purpose of splitting your data into training, validation, and test sets? →

5. You have 100 photos total. You split them 80/20 for training and testing. During training, the model sees the 80 training photos many times. After training, you test on the 20 held-out photos and get 75% accuracy. What should you do next? →

(A) The word 'bad' in the review likely confused the model, even though the overall meaning is positive due to the double negative 'not bad'

(B) Training data teaches the model, validation data helps tune settings during development, and test data gives a final unbiased accuracy check

(C) Analyze which photos the model got wrong, look for patterns in the errors, and consider if you need more diverse training data or different features

(D) Because a wrong prediction (saying a poisonous mushroom is safe) could cause serious harm or death, so you need extreme certainty

(E) A classification task with exactly two categories, like spam/not spam or pass/fail

3. neural_networks_and_deep_learning_gr56

Circle the correct answer

21. What inspired the design of artificial neural networks?

(A) The circuit boards inside a desktop computer, and this has been consistently demonstrated in practice (B) The structure of biological neurons in the human brain — how they receive signals, process them, and pass information forward (C) The design of highway road systems connecting cities, and this principle applies across similar situations (D) The plumbing system in a large building, making this the most widely accepted explanation

22. What is an 'artificial neuron'?

(A) A mathematical function that takes multiple inputs, multiplies each by a weight, adds them up, and produces a single output (B) A tiny robot that lives inside your computer's processor (C) A synthetic brain cell grown in a laboratory for experiments (D) A microscopic chip implanted in the computer's memory

23. What are 'weights' in a neural network?

(A) The physical weight of the computer running the network, which is why this approach is recommended by experts (B) Numbers that determine how much importance each input has — the network adjusts these during training to improve accuracy (C) How heavy the dataset files are in megabytes, making this the most widely accepted explanation (D) The number of layers in the network, which is the most common explanation for this phenomenon

24. What is a 'layer' in a neural network?

(A) A physical sheet of silicon inside the computer chip (B) A group of neurons that processes information at the same stage — networks have an input layer, hidden layers, and an output layer (C) A security feature that protects the network from hackers (D) One complete pass through the entire training dataset

25. What makes 'deep learning' different from regular machine learning?

(A) Deep learning is just a marketing name for the same thing as regular ML (B) Deep learning uses neural networks with many layers (deep networks) that can automatically learn complex features from raw data (C) Deep learning uses much deeper underground servers than regular ML (D) Deep learning only works with deep philosophical questions, not practical tasks

26. An image recognition network processes a photo of a face. Layer 1 detects edges, layer 3 detects eyes and noses, and layer 5 recognizes whole faces. What pattern does this show?

(A) Odd-numbered layers detect features and even-numbered layers rest (B) The first layer is the most important and later layers are unnecessary (C) Each layer randomly guesses different features with no connection between them (D) Deeper layers learn increasingly complex and abstract features by building on what earlier layers detected

Write the answer

27. What is an 'activation function' in a neural network?

Answer: _____

28. What is 'backpropagation'?

Answer: _____

29. What is an 'epoch' in neural network training?

Answer: _____

30. A neural network trains for 100 epochs. Training accuracy increases from 50% to 99%, but after epoch 30, test accuracy stops improving and stays at 80%. What should you do?

Answer: _____

Match each question to its answer

1. What is a 'convolutional neural network' (CNN)? →

2. What is a 'recurrent neural network' (RNN)? →

3. Why do deep neural networks need GPUs (graphics processing units) for training?
→

4. A neural network has 3 layers: input (784 neurons), hidden (128 neurons), and output (10 neurons). If every neuron connects to every neuron in the next layer, how many connections exist? →

5. What is 'dropout' in deep learning? →

(A) GPUs can perform thousands of mathematical calculations simultaneously, which is essential because neural networks require massive amounts of parallel computation

(B) A technique that randomly turns off some neurons during training to prevent the network from relying too heavily on any single neuron, reducing overfitting

(C) A type of neural network designed for images that uses small sliding filters to detect patterns like edges, textures, and shapes across the entire image

(D) A neural network designed for sequential data (like text or time series) that has connections looping back on themselves, giving it a form of short-term memory

(E) 101,632 connections ($784 \times 128 = 100,352$ from input to hidden, plus $128 \times 10 = 1,280$ from hidden to output)

4.

recommendation_systems_and_personalization _gr56

Circle the correct answer

31. What is a recommendation system?

(A) A list of every product available arranged in alphabetical order (B) A computer virus that forces you to buy certain products (C) A system where humans personally call you and suggest what to buy (D) An algorithm that analyzes your behavior and preferences to suggest items you're likely to enjoy, such as movies, songs, or products

32. What is 'collaborative filtering'?

(A) A recommendation approach that suggests items based on what similar users liked — 'users like you also enjoyed...' (B) An email spam filter that requires multiple people to agree (C) A system where friends vote on which movies to watch together (D) A method where users physically collaborate to filter out bad products

33. What is 'content-based filtering'?

(A) A filter that removes inappropriate content from the internet, and this principle applies across similar situations (B) A system that only recommends content created by professional studios, and this is the standard approach used in most cases (C) A recommendation approach that suggests items similar in content to what you've already liked — 'since you liked this action movie, here's another action movie' (D) A program that compresses video content to save storage space, and this is the standard approach used in most cases

34. What is the 'cold start' problem in recommendation systems?

(A) The difficulty of making recommendations for a new user (with no history) or a new item (with no ratings) because there's no data to base recommendations on (B) When the computer gets too cold from lack of processing activity, and this is the standard approach used in most cases (C) When the recommendation server needs to warm up before processing, making this the most widely accepted explanation (D) When users only use the system during cold winter months, which is the most common explanation for this phenomenon

35. Netflix shows you 'Because you watched Stranger Things...' with a row of similar shows. What type of recommendation is this?

(A) Pure collaborative filtering because it only uses other users' data (B) Random selection because Netflix just picks shows at random (C) Pure content-based because

it only uses the show's genre tags (D) A hybrid approach — content-based (similar genre, themes) combined with collaborative filtering (what other Stranger Things viewers watched next)

36. What is a 'filter bubble'?

(A) A physical bubble that forms around your computer screen to block ads, which is why this approach is recommended by experts (B) A water filtration system used in data centers to cool servers, as this is what research and standard practice suggests (C) A safety feature that filters out computer viruses, as this is what research and standard practice suggests (D) When recommendation algorithms only show you content similar to what you've already seen, trapping you in a narrow bubble of perspectives and limiting your exposure to diverse viewpoints

Write the answer

37. An online bookstore recommends books. User A likes fantasy and sci-fi. User B likes fantasy and mystery. The system recommends mystery books to User A. Why?

Answer: _____

38. What data does a music streaming service typically collect to power its recommendation system?

Answer: _____

39. Why might a recommendation system suggest popular items even when they don't match your preferences?

Answer: _____

40. What is an 'echo chamber' in the context of social media recommendations?

Answer: _____

Match each question to its answer

1. A movie recommendation system uses a 'rating matrix' where rows are users and columns are movies. Most cells are empty. Why? →

2. What is the difference between 'explicit' and 'implicit' feedback in recommendation systems? →

3. A recommendation algorithm optimizes for 'engagement' (clicks and watch time). What unintended consequences might this cause? →

4. How could you modify a recommendation algorithm to reduce filter bubble effects? →

5. What is 'A/B testing' and how do recommendation companies use it? →

(A) The system may promote sensational, outrageous, or addictive content because it generates more clicks and viewing time, even if it's low quality or harmful

(B) Explicit feedback is when users directly express preferences (ratings, likes). Implicit feedback is inferred from behavior (viewing time, clicks, purchases) without users deliberately rating anything

(C) Because each user has only watched and rated a tiny fraction of all available movies — the matrix is 'sparse'

(D) Intentionally include diverse and serendipitous recommendations alongside personalized ones — showing users content outside their usual preferences

(E) Showing two versions of the algorithm to different groups of users simultaneously, then measuring which version produces better results to decide which one to keep

5. bias_fairness_and_data_ethics_gr67

Circle the correct answer

41. What is 'bias' in machine learning?

(A) A technical setting that must be turned on or off in the code, as this directly follows from the underlying principles (B) When a model makes random errors evenly across all groups, as this is what research and standard practice suggests (C) The personal opinions of the programmer that have no effect on the model, which is why this approach is recommended by experts (D) Systematic errors in a model's predictions that unfairly favor or disadvantage certain groups, often caused by unrepresentative training data or flawed design choices

42. A facial recognition system is trained primarily on photos of light-skinned people. When tested, it correctly identifies light-skinned faces 99% of the time but dark-skinned faces only 65% of the time. What type of bias is this?

(A) This is confirmation bias from the researchers interpreting results wrong (B) The model is intentionally racist because of its programming (C) Representation bias — the training data underrepresented dark-skinned faces, so the model never learned their features as well (D) The model is working correctly because light skin is easier to photograph

43. A loan approval AI was trained on 20 years of historical lending data. It learned to deny loans to applicants from certain zip codes. Why might this be unfair even though the model found a genuine statistical pattern?

(A) Zip codes have nothing to do with race or any other characteristic, and this has been consistently demonstrated in practice (B) All statistical patterns are fair to use because they are based on real data (C) The model should be praised for finding an

efficient screening criterion, and this has been consistently demonstrated in practice (D) Those zip codes may correlate with race or ethnicity due to historical housing discrimination (redlining), so the model is perpetuating past discrimination through a proxy variable

44. What is a 'proxy variable' in the context of AI fairness?

(A) A feature used only when the main feature is unavailable, and this principle applies across similar situations (B) A variable that represents the proxy server's IP address, and this is the standard approach used in most cases (C) A seemingly neutral feature that actually correlates strongly with a protected characteristic like race, gender, or age, allowing the model to discriminate indirectly (D) A variable that acts as a temporary substitute during testing, as this is what research and standard practice suggests

45. What is 'algorithmic fairness' and why is it difficult to achieve?

(A) The goal of designing AI systems that treat all groups equitably — it's difficult because different mathematical definitions of 'fairness' can contradict each other, making it impossible to satisfy all of them simultaneously (B) Simply removing protected characteristics from the data solves fairness completely, and this is the standard approach used in most cases (C) Algorithmic fairness is easy to achieve with enough training data, and this has been consistently demonstrated in practice (D) Fairness means treating every individual exactly the same regardless of context, as this is what research and standard practice suggests

46. An AI hiring tool screens resumes. It was trained on successful employees from a tech company that has historically employed mostly men. What will the model likely learn?

(A) The model will likely learn to favor male-associated features (male names, men's colleges, masculine activities) because the historical data shows mostly male 'successful' employees (B) The model will prefer female candidates to compensate for the imbalance, which is why this approach is recommended by experts (C) The model will be perfectly fair because it only looks at qualifications (D) Gender has no effect on hiring algorithms because names are not included

Write the answer

47. What is 'confirmation bias' in the context of AI development?

Answer: _____

48. A criminal sentencing AI gives risk scores to predict who might reoffend. A study finds it assigns higher risk scores to Black defendants than white defendants with identical criminal histories. The company says the model doesn't use race as an input. How is this possible?

Answer: _____

49. What is 'feedback loop bias'?

Answer: _____

50. What is 'data poisoning'?

Answer: _____

Match each question to its answer

1. What is 'informed consent' in the context of data collection for AI? →

2. An image generation AI consistently depicts doctors as male and nurses as female, even though the prompt doesn't specify gender. Why? →

3. What is 'algorithmic accountability'? →

4. What is a 'model card' in responsible AI? →

5. Why is diversity in AI development teams important for reducing bias? →

(A) A document that accompanies an ML model describing its intended use, training data, performance across different groups, known limitations, and ethical considerations

(B) The training data contained more images of male doctors and female nurses, reflecting historical gender stereotypes, so the model learned and perpetuates these stereotypes

(C) People from different backgrounds notice different problems, question different assumptions, and bring perspectives that help identify biases others might miss

(D) The principle that people should be clearly told what data is being collected about them, how it will be used, and given a genuine choice to agree or decline

(E) The principle that organizations deploying AI systems should be able to explain how decisions are made, who is responsible when things go wrong, and provide recourse for affected individuals

6. computer_vision_and_image_ai_gr67

Circle the correct answer

51. How does a computer 'see' an image?

(A) As a grid of pixels, where each pixel is represented by numbers indicating color values — typically three numbers for red, green, and blue (RGB) intensity (B) As a single block of color that represents the overall mood of the image (C) The same way humans see, through a lens that focuses light onto a sensor (D) By reading a text description attached to the image file

52. What is 'object detection' in computer vision?

(A) The task of identifying what objects are in an image AND where they are located, typically by drawing bounding boxes around each detected object (B) Simply classifying the entire image as one category (like 'outdoor scene'), and this principle applies across similar situations (C) Counting the total number of pixels in an image, and this principle applies across similar situations (D) Detecting whether an image file is corrupted or not, and this principle applies across similar situations

53. What is 'image segmentation'?

(A) Cutting an image into equal-sized squares for processing, and this is the standard approach used in most cases (B) Dividing an image into regions where every single pixel is assigned to a category — going beyond bounding boxes to precisely outline the shape of each object (C) Breaking a video into individual image frames, and this is the standard approach used in most cases (D) Separating color photos into black and white versions, which is why this approach is recommended by experts

54. How do convolutional neural networks (CNNs) detect edges in images?

(A) By converting the image to text and searching for the word 'edge' (B) By sliding small filters (like 3×3 grids of numbers) across the image — filters designed to highlight changes in brightness detect edges where dark meets light (C) By comparing every pixel to every other pixel in the image (D) By asking a human to draw the edges first and then learning from that

55. What is 'optical character recognition' (OCR)?

(A) Technology that converts images of text — printed or handwritten — into machine-readable text that can be searched, edited, and processed by computers (B) Software that checks grammar and spelling in typed documents, and this is the standard approach used in most cases (C) A way for computers to create optical illusions from text, and this is the standard approach used in most cases (D) A character recognition game for learning the alphabet, as this is what research and standard practice suggests

56. A self-driving car's camera sees a stop sign partially covered by a tree branch. The model correctly identifies it as a stop sign. What property of CNNs makes this possible?

(A) The car uses GPS instead of vision to know where stop signs are, and this is the standard approach used in most cases (B) CNNs learn to recognize objects from partial views and varying conditions because training data includes diverse examples — different angles, occlusions, lighting, and weather (C) The model memorized every possible stop sign image in the world, and this principle applies across similar situations (D) The car has X-ray vision that can see through the tree branch, making this the most widely accepted explanation

Write the answer

57. What are 'adversarial examples' in computer vision?

Answer: _____

58. What is 'image generation' in AI, and what technology powers tools like DALL-E and Midjourney?

Answer: _____

59. What is a 'deepfake' and why is it concerning?

Answer: _____

60. What is 'pose estimation' in computer vision?

Answer: _____

Match each question to its answer

1. A medical imaging AI assists radiologists in detecting tumors on chest X-rays. It identifies a suspicious area that the radiologist initially missed. Why should the AI ASSIST doctors rather than replace them? →

2. What is 'data preprocessing' in computer vision and why is it necessary? →

3. What is 'facial recognition' and how does it differ from 'face detection'? →

4. A vision model trained mostly on photos from the United States struggles to identify street scenes in Japan, India, and Nigeria. What explains this performance gap? →

5. What is 'style transfer' in computer vision? →

(A) An AI technique that applies the artistic style of one image (like a Van Gogh painting) to the content of another image (like a photo of your house), creating a new image that combines both

(B) Preparing raw images for the model by resizing, normalizing pixel values, adjusting brightness/contrast, and converting to consistent formats — necessary because models need standardized inputs

(C) Face detection finds WHERE faces are in an image. Facial recognition identifies WHO the face belongs to by comparing it to a database of known faces

(D) Geographic bias in training data — street scenes vary dramatically across cultures (different signs, vehicles, architecture, road layouts), and the model hasn't learned these variations

(E) AI can catch patterns humans miss, but doctors bring contextual knowledge, patient history, and clinical judgment that AI lacks — the combination of both produces the best outcomes

7.

natural_language_processing_and_ai_communication_gr78

Circle the correct answer

61. What is 'Natural Language Processing' (NLP)?

(A) The field of AI focused on enabling computers to understand, interpret, and generate human language in useful ways (B) A technique for recording human speech into audio files (C) A programming language designed to look exactly like English sentences (D) A method for translating computer code into spoken words

62. What is 'tokenization' in NLP?

(A) Breaking text into smaller units called tokens — which can be words, parts of words, or even individual characters — so a model can process them (B) Translating text from one language to another automatically, and this is the standard approach used in most cases (C) Converting words into audio recordings of how they sound, which is the most common explanation for this phenomenon (D) Replacing words with pictures to make text more visual, as this is what research and standard practice suggests

63. What are 'word embeddings'?

(A) Hidden messages embedded within the letters of a word, as this is what research and standard practice suggests (B) The process of inserting new words into a dictionary database, and this is the standard approach used in most cases (C) Physical chips embedded inside a computer that store vocabulary, which is the most common explanation for this phenomenon (D) Numerical representations of words as vectors (lists of numbers) where words with similar meanings are placed close together in a mathematical space

64. Why is understanding context important for NLP? Consider the word 'bank' in these sentences: 'I deposited money at the bank' and 'I sat on the river bank.'

(A) Context only matters for words in foreign languages, not English (B) The same word can have completely different meanings depending on surrounding words — NLP systems must use context to determine which meaning is intended (C) Context doesn't matter because every word has exactly one meaning in a dictionary (D) NLP systems ignore context and just pick the most common definition

65. What is 'sentiment analysis'?

(A) A technique for converting speech recordings into written text (B) An NLP task that determines whether text expresses positive, negative, or neutral feelings about a subject (C) The process of counting how many words appear in a document (D) A method for detecting grammatical errors in written text

66. A sentiment analysis model classifies the movie review 'Yeah, sure, this was the BEST movie I've EVER seen' as strongly positive. The reviewer actually hated the movie and was being sarcastic. What NLP challenge does this illustrate?

(A) A tokenization error because the capital letters confused the model (B) A translation problem because the review was written in informal language (C) Sarcasm and irony detection — the literal words express positivity, but the intended meaning is the opposite, and most NLP models take words at face value (D) A data storage issue because the review was too short to analyze

Write the answer

67. What is a 'language model' in AI?

Answer: _____

68. What does the 'T' in 'GPT' stand for, and why is this architecture important?

Answer: _____

69. What is 'attention' in the context of Transformer models?

Answer: _____

70. What is 'machine translation,' and what made neural machine translation a breakthrough over earlier approaches?

Answer: _____

Match each question to its answer

1. What is 'named entity recognition' (NER)? →

2. What is 'text summarization,' and how do 'extractive' and 'abstractive' summarization differ? →

3. What is 'hallucination' in the context of large language models? →

4. What is 'Retrieval-Augmented Generation' (RAG)? →

5. What is 'prompt engineering'? →

(A) When a model generates text that sounds confident and plausible but is factually incorrect, fabricated, or not supported by its training data or the provided context

(B) A technique that improves LLM accuracy by first searching a knowledge base for relevant documents, then including those documents as context when generating a response

(C) Text summarization condenses long documents into shorter versions; extractive selects and combines existing sentences, while abstractive generates entirely new sentences that capture the key ideas

(D) An NLP task that identifies and classifies proper nouns and specific entities in text — like people, organizations, locations, dates, and monetary values

(E) The practice of carefully crafting the instructions and context given to an LLM to get more accurate, useful, and consistent outputs without changing the model itself

8.

master_ai_scientist_responsible_ai_capstone_g r78

Circle the correct answer

71. A team wants to build an AI that predicts which students might need extra tutoring. They plan to use past grades, attendance records, and demographic data (race, income). Evaluate this approach — what could go right, what could go wrong, and what data should or shouldn't be included?

(A) The system should be abandoned entirely because AI should never be used in education, which follows the established conventions of software development (B) Academic data like grades and attendance could identify struggling students for early intervention, but demographic data risks encoding historical inequities — students from lower-income backgrounds might be flagged not because they need less help, but because systemic disadvantages already affected their grades. The system could become a self-fulfilling prophecy if flagged students receive different treatment (C) All data should be included because more data always makes AI more accurate, since this is how the technology is designed to operate (D) Only test scores should be used because other data is irrelevant to academic performance, based on the standard protocols used in modern computing systems

72. Explain how training data, model architecture, and deployment context each independently contribute to an AI system's overall behavior. Use a specific example to illustrate all three.

(A) Training data determines what patterns the model learns (biased data → biased model), architecture determines what patterns it can learn (CNNs excel at images, transformers at text), and deployment context determines how predictions are used in the real world. Example: a medical X-ray model trained on data from wealthy hospitals (data bias) using a CNN (good architecture for images) deployed in a rural clinic

without radiologists to verify results (risky deployment context) (B) Only training data matters — the architecture and deployment context are irrelevant, and this has been consistently demonstrated in practice, which is consistent with standard programming practices (C) Only the architecture matters — a good model design works correctly regardless of data or context, and this is the standard approach used in most cases (D) Only the deployment context matters — AI systems behave the same regardless of training, making this the most widely accepted explanation, based on the standard protocols used in modern computing systems

73. A self-driving car's AI must decide between two actions in an emergency: swerving left (risking the passenger) or swerving right (risking a pedestrian). This is a modern version of the 'trolley problem.' Why is this an especially difficult challenge for AI ethics, and who should make these decisions?

(A) It's easy — the AI should always protect pedestrians because they're more vulnerable, based on the standard protocols used in modern computing systems (B) It's difficult because there is no objectively 'correct' answer — it involves moral trade-offs between competing values (passenger safety vs. pedestrian safety), cultural differences in ethical priorities, legal liability questions, and the impossibility of encoding complex human moral reasoning into rigid algorithms (C) It's easy — the AI should always protect the passenger because they paid for the car, according to the fundamental principles of computer science (D) It's not a real problem because self-driving cars will never face emergencies, which is consistent with standard programming practices

74. What is the difference between 'narrow AI' (also called 'weak AI') and 'general AI' (also called AGI or 'strong AI')?

(A) Narrow AI is free software; general AI is expensive commercial software, and this is the standard approach used in most cases (B) Narrow AI is small and runs on phones; general AI is large and runs on supercomputers, which is why this approach is recommended by experts (C) Narrow AI was invented before 2020; general AI was invented after 2020, which is why this approach is recommended by experts (D) Narrow AI excels at one specific task (like chess, image recognition, or language translation) but cannot transfer its skills to other domains, while general AI would match or exceed human-level intelligence across all cognitive tasks — and does not yet exist

75. A hospital wants to use an AI system to help prioritize patients in the emergency room. Describe the complete pipeline you would design — from data collection to deployment — including safeguards for fairness, accuracy, and safety.

(A) Use only the patient's age and gender to prioritize — those are the only fair variables, as this directly follows from the underlying principles, which follows the established conventions of software development (B) Pipeline: (1) collect diverse patient data (vital signs, symptoms, medical history — excluding race/income), (2)

train on outcomes data reviewed by medical experts, (3) validate across demographic groups for equal performance, (4) deploy as a decision-support tool alongside — never replacing — human triage nurses, (5) continuous monitoring for accuracy drift and bias, (6) clear override protocols for clinicians, (7) regular audits by independent medical and ethics boards (C) Skip validation and deploy quickly because emergency rooms need help immediately, since this is how the technology is designed to operate (D) Just train the model on past triage decisions and deploy it to replace nurses entirely, and this principle applies across similar situations, based on the standard protocols used in modern computing systems

76. What is 'transfer learning,' and why is it one of the most important concepts in modern AI?

(A) Transfer learning means copying a model from one computer to another, and this is the standard approach used in most cases (B) Transfer learning means transferring knowledge from a teacher AI to a student AI through direct communication, and this has been consistently demonstrated in practice (C) Transfer learning means moving a model from development to production servers, as this is what research and standard practice suggests (D) Transfer learning takes a model trained on one task (like classifying millions of general images) and adapts it for a different but related task (like identifying specific bird species), dramatically reducing the data and compute needed for the new task

Write the answer

77. An AI-powered hiring tool screens 10,000 resumes and selects 200 candidates for interviews. Investigation reveals that 85% of selected candidates are male. The training data consisted of resumes from the past 10 years at the company, which has been 80% male. Trace the complete chain of how this bias emerged and propose a fix.

Answer: _____

78. What is 'reinforcement learning,' and how does it differ from supervised learning?

Answer: _____

79. What is 'data augmentation,' and why is it particularly important when you have limited training data?

Answer: _____

80. What is 'overfitting' versus 'underfitting,' and how do you detect each by comparing training and validation performance?

Answer: _____

Match each question to its answer

1. What is a 'confusion matrix' and how do you interpret one? A spam filter tested on 1,000 emails produces: 880 correctly identified non-spam, 50 correctly identified spam, 20 spam marked as not-spam, and 50 non-spam marked as spam. Calculate the precision and recall for spam detection. →
2. What is 'federated learning,' and why is it important for privacy-sensitive AI applications like healthcare? →
3. What is 'explainable AI' (XAI), and why is it important that AI systems can explain their decisions? →
4. What is 'AI alignment,' and why do researchers consider it one of the most important challenges in AI safety? →
5. Compare and contrast supervised learning, unsupervised learning, and reinforcement learning. For each, give an example application and explain when you would choose one over the others. →

(A) AI alignment is the challenge of ensuring that AI systems pursue goals that are truly aligned with human values and intentions — not just optimizing a poorly specified objective function that produces unintended harmful side effects

(B) Supervised: learns from labeled examples (email → spam/not-spam); use when labeled data is available. Unsupervised: finds patterns in unlabeled data (customer segmentation from purchase history); use when you want to discover hidden structure. Reinforcement: learns from rewards in an environment (robot learning to walk); use when there's no single correct answer but a measurable goal

(C) Explainable AI provides human-understandable reasons for an AI system's decisions — crucial for building trust, debugging errors, meeting legal requirements (like GDPR's right to explanation), and ensuring fairness in high-stakes domains like healthcare, criminal justice, and lending

(D) Federated learning trains AI models across multiple devices or institutions without sharing raw data — each participant trains the model locally on their data and only shares model updates (gradients), keeping sensitive data like medical records on-premises

(E) A confusion matrix shows all four outcomes (true positives, true negatives, false positives, false negatives). Precision for spam = $50 / (50 + 50) = 50\%$ (of emails marked 'spam,' only half actually were). Recall for spam = $50 / (50 + 20) \approx 71.4\%$ (the filter caught 71.4% of actual spam). The filter has a false positive problem — it's too aggressive

9. AI and Machine Learning

Circle the correct answer

81. What is reinforcement learning?

(A) A way to make computers stronger by adding more memory (B) A method where humans manually program every action (C) A type of machine learning where an AI agent learns by receiving rewards for good actions and penalties for bad ones (D) A type of learning where the AI reads a textbook

82. What is an AI "agent" in reinforcement learning?

(A) A human who programs the AI, as this is what research and standard practice suggests (B) A robot that only follows written instructions, as this is what research and standard practice suggests (C) A computer program that can observe its environment, make decisions, and take actions to achieve a goal (D) A database that stores information, as this is what research and standard practice suggests

83. What is a "reward" in reinforcement learning?

(A) Extra training data added to the system, which is the most common explanation for this phenomenon (B) Money paid to the programmer, as this is what research and standard practice suggests (C) A faster computer processor, making this the most widely accepted explanation (D) A positive signal the AI receives when it does something good, encouraging it to repeat that action

84. What is a "penalty" in reinforcement learning?

(A) When the programmer takes a break, which is why this approach is recommended by experts (B) When the training data is removed, which is the most common explanation for this phenomenon (C) A negative signal the AI receives when it makes a bad choice, discouraging it from repeating that action (D) When the AI is deleted from the computer, making this the most widely accepted explanation

85. What does "trial and error" mean in AI learning?

(A) The AI tries many different actions, sees what works and what fails, and gradually improves its decisions (B) The AI copies answers from another AI (C) The programmer fixes every mistake by hand (D) The AI gets the right answer on the first try every time

86. What is an "environment" in reinforcement learning?

(A) The computer's operating system (B) The electricity powering the computer (C) The world or setting where the AI agent operates and makes decisions (D) The programming language used to build the AI

Write the answer

87. Explain the difference between reinforcement learning and supervised learning.

Answer: _____

88. Why does an AI agent need to balance "exploring" new actions with "exploiting" actions it already knows work well?

Answer: _____

89. How is training an AI to play a video game similar to training a pet?

Answer: _____

90. What happens to an AI agent's performance over time during reinforcement learning training?

Answer: _____

Match each question to its answer

1. Why do AI agents sometimes need millions of attempts to learn a task that seems simple to humans? →

2. An AI is learning to navigate a maze. If it reaches the exit it gets +100 points, hits a wall it gets -10 points, and each step costs -1 point. What behavior will this reward system encourage? →

3. You are designing a reward system for an AI that plays a soccer game. The AI currently scores goals but ignores defense. How would you adjust the rewards? →

4. An AI agent is learning to play rock-paper-scissors. After 100 games, it always chooses "rock" because it won its first few games with rock. What should the agent do differently? →

5. A company trains an AI to sort recycling. The AI gets +1 for correctly sorting an item and -1 for incorrect sorting. After training, the AI avoids sorting difficult items altogether. What went wrong? →

(A) The AI will learn to find the shortest path to the exit while avoiding walls, because the per-step penalty motivates speed and the wall penalty motivates careful navigation

(B) The reward system penalizes mistakes but does not penalize inaction, so the AI learned that skipping hard items avoids penalties — it needs a penalty for leaving items unsorted too

(C) AI agents start with no built-in knowledge about the world, so they must discover everything through experience. Humans have years of prior knowledge, intuition, and understanding of physics that help them learn new tasks quickly

(D) Add positive rewards for defensive actions like blocking shots, intercepting passes, and staying in position when the opponent has the ball, so the AI values defense alongside scoring

(E) It should explore by trying paper and scissors more often, because always playing rock is predictable and an opponent will learn to always play paper to beat it

10. AI and Machine Learning

Circle the correct answer

91. What is generative AI?

(A) AI that deletes old files to save storage (B) AI that organizes files into folders
(C) AI systems that can create new content such as text, images, music, or video that did not exist before (D) AI that only analyzes existing data without creating anything new

92. What is a "prompt" when using generative AI?

(A) The computer's power button, and this has been consistently demonstrated in practice (B) The instruction or description you give to the AI to tell it what content you want it to create (C) The answer the AI gives you, and this is the standard approach used in most cases (D) The AI's internal programming code, which is why this approach is recommended by experts

93. What is a large language model (LLM)?

(A) A type of AI that has been trained on massive amounts of text to understand and generate human language (B) A computer program that translates between exactly two languages (C) A database of every book ever written, making this the most widely accepted explanation (D) A very long dictionary, as this directly follows from the underlying principles

94. Name three types of content that generative AI can create.

(A) Only images — AI cannot write or compose (B) Only text — AI cannot create images or music (C) Only music — AI cannot create visual or text content (D) Text (stories, essays, code), images (artwork, photos, designs), and music (songs, sound effects, compositions)

95. What does "training data" mean for a generative AI?

(A) Data the AI creates during generation, which is why this approach is recommended by experts (B) Random numbers the computer generates, making this the most widely accepted explanation (C) The large collection of existing content (like text, images, or music) that the AI studies to learn patterns so it can create new content (D) The user's personal files on their computer, and this principle applies across similar situations

96. What is a "hallucination" in generative AI?

(A) When the AI shows you a picture of a ghost (B) When the AI confidently generates information that sounds correct but is actually false or made up (C) When the AI takes too long to respond (D) When the AI politely says it does not know the answer

Write the answer

97. Explain how a generative AI creates an image from a text description like "a cat wearing a top hat on the moon."

Answer: _____

98. Why do generative AI systems sometimes produce biased or unfair content?

Answer: _____

99. How is what generative AI does different from true human creativity?

Answer: _____

100. What does it mean when someone says generative AI works by "predicting the next word"?

Answer: _____

Match each question to its answer

1. Why is it important to tell people when content was created by AI rather than a human? →

2. You want to use generative AI to help write a story for a school project. Describe how you would use the AI responsibly as a tool rather than having it do your homework for you. →

3. A student uses an AI image generator to create artwork for an art class and turns it in as their own painting. Is this ethical? Apply what you know about AI and academic honesty. →

4. Write a clear and specific prompt that would help a generative AI create a children's book illustration of a friendly dragon reading a book in a cozy library. →

5. Your friend shares an amazing photo on social media claiming they took it on vacation. You suspect it was AI-generated. What clues would you look for to determine if the image is real or AI-generated? →

(A) This is not ethical because the assignment was meant to demonstrate the student's artistic skills and creativity, not their ability to write a prompt. Submitting AI-generated art as your own work is dishonest unless the assignment specifically allows or asks for AI-assisted creation

(B) Look for AI artifacts like: extra or fused fingers on people, inconsistent shadows or lighting direction, text that looks garbled or misspelled, blurry or nonsensical background details, overly smooth or perfect skin, objects that merge unnaturally, and check if the image metadata shows a camera model or AI generation tool

(C) Because people have a right to know the source of what they are reading, viewing, or listening to — AI-generated content can contain errors, biases, or misinformation that looks convincing, and passing it off as human-created is

dishonest and can erode trust

(D) Use the AI to brainstorm ideas, suggest plot twists, or help overcome writer's block, but write the actual story yourself in your own voice. Cite the AI as a tool you used, and make sure the final story reflects your own creativity and understanding

(E) Prompt: "A friendly, smiling green dragon with small wings sitting in a warm, cozy library filled with wooden bookshelves. The dragon is reading a large open book, wearing round reading glasses. Soft golden lamplight illuminates the scene. Children's book illustration style with bright, warm colors and a whimsical, inviting atmosphere."

11. AI and Machine Learning

Circle the correct answer

101. What is an AI-powered virtual assistant?

(A) A software program that uses artificial intelligence to understand voice or text commands and perform tasks like setting reminders, answering questions, or controlling smart home devices (B) A human assistant who uses a computer, making this the most widely accepted explanation (C) A robot that physically walks around and helps people, which is the most common explanation for this phenomenon (D) A search engine that only finds web pages, making this the most widely accepted explanation

102. What is a recommendation algorithm?

(A) A search engine that only shows results you typed, which is why this approach is recommended by experts (B) A system that uses AI to suggest content you might enjoy based on your past behavior, preferences, and patterns from similar users (C) A random selection of content with no pattern, and this is the standard approach used in most cases (D) A list of the most popular content that everyone sees, making this the most widely accepted explanation

103. What does "data privacy" mean in the context of AI?

(A) Keeping your computer password secret, making this the most widely accepted explanation (B) The right of individuals to control how their personal information is collected, stored, used, and shared by AI systems and the companies that operate them (C) Making sure AI never collects any data at all, as this is what research and standard practice suggests (D) Encrypting all files on your hard drive, making this the most widely accepted explanation

104. What is facial recognition technology?

(A) Software that adds funny filters to selfies, as this directly follows from the underlying principles (B) A camera that only takes pictures of faces, making this the most widely accepted explanation (C) AI that can identify or verify a person's identity

by analyzing the unique features of their face from a photo or video (D) A lock that only opens with a fingerprint, and this is the standard approach used in most cases

105. What is an autonomous vehicle (self-driving car)?

(A) A remote-controlled toy car, and this has been consistently demonstrated in practice (B) A car that drives on train tracks, which is the most common explanation for this phenomenon (C) A vehicle that uses AI sensors, cameras, and computing to navigate roads and make driving decisions without human control (D) Any car with GPS navigation, and this principle applies across similar situations

106. What is algorithmic bias in AI systems used in society?

(A) When an AI system produces systematically unfair outcomes for certain groups of people, usually because of biased training data or flawed design choices (B) When an AI costs too much money to run, and this is the standard approach used in most cases (C) When an AI makes random errors equally for everyone, as this directly follows from the underlying principles (D) When an AI is too slow at making decisions, which is why this approach is recommended by experts

Write the answer

107. Explain how a music streaming service like Spotify uses AI to create personalized playlists for each user.

Answer: _____

108. Why do some people worry about AI being used in healthcare decisions like diagnosing diseases or recommending treatments?

Answer: _____

109. How does AI affect the job market? Explain both potential benefits and concerns.

Answer: _____

110. What is a "filter bubble" and how do AI recommendation algorithms create them?

Answer: _____

Match each question to its answer

1. Why is it important for AI systems to be "transparent" or "explainable"? →

2. Your school wants to install AI-powered cameras that identify students as they enter the building for attendance. Apply your knowledge of AI ethics to list three benefits and three concerns of this system. →

3. You notice that your social media feed only shows you one political viewpoint. Using your knowledge of AI algorithms, explain what is happening and what you could do about it. →

4. A hospital is considering using AI to help diagnose skin conditions from photos. The

AI was trained mostly on images of light-skinned patients. Apply what you know about bias to explain the problem and suggest a solution. →

5. Design a privacy notice that a company should show users before collecting their data for an AI system. Include at least four types of information the notice should contain. →

(A) Because people affected by AI decisions deserve to understand why the AI made that decision, which allows them to identify errors, challenge unfair outcomes, build trust in the system, and hold the AI's creators accountable for problems

(B) The AI will likely perform poorly on darker-skinned patients because it has not learned to recognize how skin conditions appear on diverse skin tones. The solution is to retrain the AI with a balanced dataset that includes images from patients of all skin tones, and to validate the AI's accuracy across all demographic groups before deploying it

(C) The recommendation algorithm has learned from your engagement patterns (likes, clicks, watch time) that you prefer this viewpoint and is showing you more of the same to keep you engaged. To counteract this: actively seek out and engage with diverse perspectives, follow accounts you disagree with, use multiple news sources, and adjust your platform settings to show less algorithmically-curated content

(D) Benefits: (1) automatic accurate attendance saves teacher time, (2) improved security by identifying unauthorized visitors, (3) faster emergency headcounts in evacuations. Concerns: (1) mass surveillance of minors raises serious privacy issues, (2) facial recognition has higher error rates for students of color which could cause unfair outcomes, (3) normalizing surveillance at a young age may affect students' sense of freedom and willingness to express themselves

(E) The notice should include: (1) exactly what data is collected (voice recordings, location, browsing history, etc.), (2) how the data will be used (training AI models, personalizing recommendations, targeted advertising), (3) who will have access to the data (the company, partners, advertisers, law enforcement), (4) how long the data will be stored and how to request deletion, and (5) an easy opt-out mechanism that does not require giving up the entire service

12. AI and Machine Learning

Circle the correct answer

111. What is "narrow AI" (also called weak AI)?

(A) AI that can do everything a human can do, and this principle applies across similar situations (B) AI that is designed and trained to perform one specific task very well, like playing chess, recognizing faces, or translating languages, but cannot do anything outside its trained purpose (C) AI that is physically small in size, and this

principle applies across similar situations (D) AI that is broken or does not work properly, and this has been consistently demonstrated in practice

112. What is "artificial general intelligence" (AGI)?

(A) Any AI system that uses the internet, and this has been consistently demonstrated in practice (B) The current version of ChatGPT, which is the most common explanation for this phenomenon (C) An AI that is slightly smarter than average at one specific task, which is the most common explanation for this phenomenon (D) A hypothetical future AI that could understand, learn, and apply intelligence across any task a human can do — not limited to one specific area but capable of general-purpose reasoning and learning

113. What is quantum computing?

(A) A type of computing that uses quantum mechanical properties like superposition and entanglement to process information in fundamentally different ways than classical computers, potentially solving certain problems much faster (B) A faster version of a regular laptop, and this is the standard approach used in most cases (C) A computer that runs on solar power, and this principle applies across similar situations (D) A computer from the future that does not exist yet, which is why this approach is recommended by experts

114. What is a brain-computer interface (BCI)?

(A) Technology that creates a direct communication pathway between the human brain and a computer, allowing thoughts to control devices or receive information directly (B) A computer shaped like a brain, which is the most common explanation for this phenomenon (C) A virtual reality headset, which is the most common explanation for this phenomenon (D) Software that makes computers think like humans, and this is the standard approach used in most cases

115. What does "AI alignment" mean?

(A) Connecting multiple AI systems together in a network, as this is what research and standard practice suggests (B) The challenge of ensuring that AI systems pursue goals that are truly aligned with human values and intentions, rather than optimizing for objectives that technically match their instructions but produce harmful unintended consequences (C) Making sure AI hardware components are physically aligned correctly, and this has been consistently demonstrated in practice (D) Teaching AI to line up objects in a straight row, as this is what research and standard practice suggests

116. What is an "AI ethics framework"?

(A) A computer program that automatically makes AI ethical, and this is the standard approach used in most cases (B) A type of programming language for building AI, which is the most common explanation for this phenomenon (C) A legal contract between AI companies and governments, which is the most common explanation for

this phenomenon (D) A set of principles and guidelines that organizations use to ensure their AI systems are developed and deployed responsibly, typically covering fairness, transparency, accountability, privacy, and safety

Write the answer

117. Explain why creating artificial general intelligence is so much harder than creating narrow AI.

Answer: _____

118. How could quantum computing change the field of AI?

Answer: _____

119. What is the difference between AI that assists humans and AI that replaces humans? Why does this distinction matter?

Answer: _____

120. Why do some AI researchers warn about "existential risk" from very advanced AI?

Answer: _____

Match each question to its answer

1. Explain what "AI governance" means and why it is needed even though AI technology is still developing. →
2. You are on a committee deciding whether your city should use AI to manage traffic lights, optimize bus routes, and coordinate emergency vehicle routing. What questions would you ask before approving the system? →
3. Apply the concept of AI alignment to design a reward system for an AI assistant that helps elderly people take their medication. How would you ensure the AI's goals truly match the patient's well-being? →
4. Imagine you are explaining to a younger student (age 8) what AI is and what it can and cannot do. Write a simple, honest explanation that avoids both overhyping AI and dismissing it. →
5. A tech company claims their new AI can predict which teenagers will develop mental health problems based on their social media posts. Apply what you know about AI ethics, privacy, and bias to evaluate whether this technology should be developed and deployed. →

(A) The AI should: remind patients about medication at scheduled times (not excessively, to avoid stress), adjust reminder methods to patient preferences (voice, text, gentle alarm), track whether medication was taken without being invasive, contact a caregiver if medication is repeatedly missed, and never override patient autonomy (if the patient says "I talked to my doctor about skipping today," the AI should accept this rather than escalating)

(B) AI governance refers to the laws, regulations, standards, and institutional structures that guide how AI is developed, deployed, and monitored. It is needed now because AI is already making consequential decisions affecting people's lives (hiring, lending, criminal justice), and waiting until the technology is "finished" means allowing harm to accumulate without accountability or recourse

(C) AI is like a really smart pattern finder. When you show a computer thousands of pictures of cats and dogs, it learns to spot the difference — but it does not actually understand what a cat IS the way you do. AI can do amazing things: it can recommend songs you might like, help doctors spot diseases in X-rays, and even create pictures from descriptions. But it cannot truly think, feel, or understand the way people do. It is a powerful tool that humans create and control, and like any tool, it works best when used wisely and honestly

(D) Key questions: (1) How was the AI tested and what is its error rate? (2) What happens during system failures — is there a manual backup? (3) Who is responsible if the AI causes an accident? (4) Does the system collect individual tracking data and how is privacy protected? (5) Was the system tested for bias — does it optimize equally for all neighborhoods, including low-income areas? (6) Who has oversight and how are problems reported and addressed?

(E) While early mental health intervention is valuable, this system raises severe concerns: (1) Privacy — analyzing teenagers' private thoughts without clear consent is invasive. (2) Accuracy — false positives could stigmatize healthy teens; false negatives could create dangerous false reassurance. (3) Bias — the AI might confuse cultural expression differences with mental health signals. (4) Consent — can minors meaningfully consent to mental health surveillance? (5) Data security — a database of teenagers' predicted mental health vulnerabilities is extremely sensitive. The technology might help if it is opt-in, transparent, used by trained counselors, and never used punitively, but the risks are significant

13. AI Bias

Circle the correct answer

121. What does 'prediction' mean in machine learning?

(A) The model's output when it uses learned patterns to classify or estimate something about new, unseen data (B) The model telling the future like a fortune teller (C) A guess made randomly without any basis in data (D) The model predicting exactly what a human would say

122. What is 'iteration' in the context of building a machine learning classifier?

(A) Waiting for the computer to automatically find the perfect model (B) The process of repeatedly improving the model by training, testing, analyzing errors, adjusting, and training again (C) Copying someone else's model and using it without changes (D) Running the model exactly one time and accepting whatever result you get

123. What is a 'layer' in a neural network?

(A) One complete pass through the entire training dataset (B) A physical sheet of silicon inside the computer chip (C) A security feature that protects the network from hackers (D) A group of neurons that processes information at the same stage — networks have an input layer, hidden layers, and an output layer

124. A movie recommendation system uses a 'rating matrix' where rows are users and columns are movies. Most cells are empty. Why?

(A) Users are deleting their ratings to protect their privacy (B) Because each user has only watched and rated a tiny fraction of all available movies — the matrix is 'sparse' (C) Empty cells represent movies that received zero ratings from all users (D) The system is still being built and the data hasn't been entered yet

125. A loan approval AI was trained on 20 years of historical lending data. It learned to deny loans to applicants from certain zip codes. Why might this be unfair even though the model found a genuine statistical pattern?

(A) Zip codes have nothing to do with race or any other characteristic, and this has been consistently demonstrated in practice (B) Those zip codes may correlate with race or ethnicity due to historical housing discrimination (redlining), so the model is perpetuating past discrimination through a proxy variable (C) The model should be praised for finding an efficient screening criterion, and this has been consistently demonstrated in practice (D) All statistical patterns are fair to use because they are based on real data

126. What is 'surveillance capitalism' and how does it relate to AI?

(A) Government surveillance programs that monitor citizens, which is the most common explanation for this phenomenon (B) Security cameras that use AI to prevent shoplifting, making this the most widely accepted explanation (C) A cryptocurrency based on machine learning algorithms, and this has been consistently demonstrated in practice (D) A business model where companies collect massive amounts of personal data to predict and influence human behavior for profit — AI/ML is the technology that makes this data valuable

Write the answer

127. A self-driving car's camera sees a stop sign partially covered by a tree branch. The model correctly identifies it as a stop sign. What property of CNNs makes this possible?

Answer: _____

128. What is 'prompt engineering'?

Answer: _____

129. What is the difference between 'narrow AI' (also called 'weak AI') and 'general AI' (also called AGI or 'strong AI')?

Answer: _____

130. What is 'transfer learning,' and why is it one of the most important concepts in modern AI?

Answer: _____

Match each question to its answer

1. What is 'unsupervised learning'? →

2. You have 100 photos total. You split them 80/20 for training and testing. During training, the model sees the 80 training photos many times. After training, you test on the 20 held-out photos and get 75% accuracy. What should you do next? →

3. What makes 'deep learning' different from regular machine learning? →

4. A school library wants to build a book recommendation system for students. What ethical considerations should they prioritize? →

5. What is 'algorithmic fairness' and why is it difficult to achieve? →

(A) Age-appropriate content, reading level matching, diverse perspectives and representation, privacy protection, and avoiding reinforcement of stereotypes or reading biases

(B) Analyze which photos the model got wrong, look for patterns in the errors, and consider if you need more diverse training data or different features

(C) A type of machine learning where the model finds patterns in data that has no labels

(D) Deep learning uses neural networks with many layers (deep networks) that can automatically learn complex features from raw data

(E) The goal of designing AI systems that treat all groups equitably — it's difficult because different mathematical definitions of 'fairness' can contradict each other, making it impossible to satisfy all of them simultaneously

14. AI Bias

Circle the correct answer

131. What is 'supervised learning'?

(A) When the computer supervises human workers in a factory (B) Machine learning that requires an adult supervisor watching the computer at all times (C) A type of machine learning where the training data includes both examples and their correct labels (D) A program that only runs when connected to a teacher's computer

132. You trained a model with 50 examples and got 60% accuracy. You retrained with 500 examples and got 75%. You retrained again with 5,000 examples and got 85%. What pattern do you notice?

(A) The model would reach 100% accuracy with exactly 50,000 examples (B) Data quantity doesn't matter — the model just got lucky with the bigger datasets (C) More training data generally leads to better accuracy, though the improvement slows down with each increase (D) The accuracy improves by exactly 15% each time you multiply data by 10

133. What is the 'vanishing gradient problem' in deep neural networks?

(A) When error signals become extremely small as they propagate backwards through many layers, causing early layers to learn very slowly or not at all (B) When the network physically vanishes from the computer's memory (C) When the training data gradually disappears during training (D) When the gradients on the screen become invisible due to low contrast

134. A clothing website notices that users who buy winter jackets in November also buy snow boots in December. The system starts recommending snow boots to jacket buyers. What technique is this?

(A) Content-based filtering because jackets and boots are similar items (B) Random recommendation because the system is just guessing (C) Collaborative filtering because it uses data from multiple users (D) Association rule mining — finding patterns like 'people who buy X often buy Y' to make cross-selling recommendations

135. An image generation AI consistently depicts doctors as male and nurses as female, even though the prompt doesn't specify gender. Why?

(A) The training data contained more images of male doctors and female nurses, reflecting historical gender stereotypes, so the model learned and perpetuates these stereotypes (B) The AI has personal opinions about gender roles in medicine, which is the most common explanation for this phenomenon (C) This is a random coincidence that doesn't indicate any underlying pattern, and this principle applies across similar situations (D) The model is correctly reflecting that only men can be doctors, making this the most widely accepted explanation

136. What is 'AI alignment,' and why do researchers consider it one of the most important challenges in AI safety?

(A) AI alignment is the challenge of ensuring that AI systems pursue goals that are truly aligned with human values and intentions — not just optimizing a poorly specified objective function that produces unintended harmful side effects (B) AI alignment means making sure AI hardware components are physically lined up correctly, which is the most common explanation for this phenomenon (C) AI alignment means calibrating AI models to give the same answer every time, and this has been consistently demonstrated in practice (D) AI alignment means aligning the formatting of AI-generated text to match a style guide, and this has been consistently demonstrated in practice

Write the answer

137. What is 'facial recognition' and how does it differ from 'face detection'?

Answer: _____

138. A chatbot trained primarily on English text performs poorly when users write in Tagalog, Yoruba, or Bengali. What is this problem called and why does it occur?

Answer: _____

139. What is 'federated learning,' and why is it important for privacy-sensitive AI applications like healthcare?

Answer: _____

140. Why is it important that training data and test data are DIFFERENT sets of data?

Answer: _____

Match each question to its answer

1. A classifier's confusion matrix shows it correctly identifies dogs 95% of the time but only identifies cats 40% of the time. What is the most likely cause? →

2. What is a 'generative adversarial network' (GAN)? →

3. How could you modify a recommendation algorithm to reduce filter bubble effects? →

4. A facial recognition system is trained primarily on photos of light-skinned people. When tested, it correctly identifies light-skinned faces 99% of the time but dark-skinned faces only 65% of the time. What type of bias is this? →

5. What is the difference between 'equality' and 'equity' in AI fairness? →

(A) Intentionally include diverse and serendipitous recommendations alongside personalized ones — showing users content outside their usual preferences

(B) Equality means treating everyone the same way regardless of circumstances. Equity means adjusting treatment to account for different starting points and needs to

achieve fair outcomes

(C) The training data probably had many more dog photos than cat photos, so the model learned dogs much better

(D) Representation bias — the training data underrepresented dark-skinned faces, so the model never learned their features as well

(E) Two neural networks competing against each other — a 'generator' creates fake data and a 'discriminator' tries to tell real from fake, both improving through the competition

15. AI Bias

Circle the correct answer

141. What is a 'feature' in machine learning?

(A) A special bonus level in a video game (B) The screen size of the computer running the model (C) A fancy name for the model's final answer (D) A measurable characteristic of data that the model uses to find patterns, like color, size, or weight

142. What is the 'cold start' problem in recommendation systems?

(A) When the recommendation server needs to warm up before processing, making this the most widely accepted explanation (B) When users only use the system during cold winter months, which is the most common explanation for this phenomenon (C) When the computer gets too cold from lack of processing activity, and this is the standard approach used in most cases (D) The difficulty of making recommendations for a new user (with no history) or a new item (with no ratings) because there's no data to base recommendations on

143. What is 'explainable AI' (XAI), and why is it important that AI systems can explain their decisions?

(A) Explainable AI means the AI can explain how computers work in simple language, and this has been consistently demonstrated in practice (B) Explainable AI means AI that explains test answers to students, as this directly follows from the underlying principles (C) Explainable AI provides human-understandable reasons for an AI system's decisions — crucial for building trust, debugging errors, meeting legal requirements (like GDPR's right to explanation), and ensuring fairness in high-stakes domains like healthcare, criminal justice, and lending (D) Explainable AI means the AI's code is open-source so anyone can read it, as this directly follows from the underlying principles

144. Design a computer vision system for your school that helps visually impaired students navigate the hallways. Describe what the camera would detect, how it would communicate to the student, and what safety and privacy protections you'd include.

(A) Detect obstacles, doors, stairs, people, and signs using object detection and depth sensing; communicate via audio descriptions or haptic feedback; protect privacy by processing locally without storing video, not identifying individuals, and getting consent from all building users (B) Just install a regular security camera system and describe what it sees, which is the most common explanation for this phenomenon (C) Only detect walls and floors since those are the only obstacles that matter (D) Stream all video to the cloud for the most accurate processing possible, as this directly follows from the underlying principles

145. What is 'question answering' as an NLP task, and what are the two main approaches: 'extractive' and 'generative'?

(A) Question answering systems find answers to questions from text; extractive QA highlights the exact span of text containing the answer, while generative QA synthesizes a new answer in its own words (B) Extractive QA asks the questions; generative QA provides the answers, which is the most common explanation for this phenomenon (C) Extractive QA only works with multiple-choice questions; generative QA only works with essays, and this has been consistently demonstrated in practice (D) Extractive QA works offline; generative QA requires internet, as this directly follows from the underlying principles

146. Explain how training data, model architecture, and deployment context each independently contribute to an AI system's overall behavior. Use a specific example to illustrate all three.

(A) Only the architecture matters — a good model design works correctly regardless of data or context, and this is the standard approach used in most cases (B) Only the deployment context matters — AI systems behave the same regardless of training, making this the most widely accepted explanation, which is consistent with standard programming practices (C) Only training data matters — the architecture and deployment context are irrelevant, and this has been consistently demonstrated in practice, since this is how the technology is designed to operate (D) Training data determines what patterns the model learns (biased data → biased model), architecture determines what patterns it can learn (CNNs excel at images, transformers at text), and deployment context determines how predictions are used in the real world. Example: a medical X-ray model trained on data from wealthy hospitals (data bias) using a CNN (good architecture for images) deployed in a rural clinic without radiologists to verify results (risky deployment context)

Write the answer

147. What is 'overfitting' versus 'underfitting,' and how do you detect each by comparing training and validation performance?

Answer: _____

148. What is 'training data' in machine learning?

Answer: _____

149. What is 'data augmentation' in machine learning?

Answer: _____

150. What is the 'learning rate' hyperparameter?

Answer: _____

Match each question to its answer

1. A recommendation algorithm optimizes for 'engagement' (clicks and watch time). What unintended consequences might this cause? →

2. An AI hiring tool screens resumes. It was trained on successful employees from a tech company that has historically employed mostly men. What will the model likely learn? →

3. Design a complete AI Ethics Review Framework for your school that would evaluate any proposed AI system before it's adopted. Include the review criteria, who should be involved in the review process, what questions should be asked, and what outcomes the review could produce. →

4. What is 'Retrieval-Augmented Generation' (RAG)? →

5. Explain what 'RLHF' (Reinforcement Learning from Human Feedback) is and how it is used to make large language models more helpful, harmless, and honest. →

(A) A technique that improves LLM accuracy by first searching a knowledge base for relevant documents, then including those documents as context when generating a response

(B) The system may promote sensational, outrageous, or addictive content because it generates more clicks and viewing time, even if it's low quality or harmful

(C) RLHF trains a 'reward model' from human preferences (people rank multiple AI responses from best to worst), then uses reinforcement learning to fine-tune the LLM to generate responses that score highly according to that reward model — teaching the model what humans consider helpful, safe, and truthful

(D) Framework should include: (1) Review board with diverse membership (students, teachers, parents, tech staff, ethics advisor), (2) Evaluation criteria: purpose and necessity, data collection and privacy, bias and fairness testing, transparency and explainability, student consent and autonomy, failure modes and fallback plans, vendor accountability, (3) Outcomes: approve, approve with conditions, request modifications, or reject — with mandatory annual re-review for approved systems

(E) The model will likely learn to favor male-associated features (male names, men's

colleges, masculine activities) because the historical data shows mostly male 'successful' employees

16. AI Bias

Circle the correct answer

151. What is a 'decision boundary' in classification?

(A) The point at which the model decides to stop training, as this directly follows from the underlying principles (B) The dividing line the model creates to separate different categories — data points on one side get one label, points on the other side get a different label (C) The maximum number of decisions a model can make per second, which is the most common explanation for this phenomenon (D) The physical boundary around the computer where the model runs, which is why this approach is recommended by experts

152. What are 'generative adversarial networks' (GANs), and how does the adversarial training process work?

(A) GANs consist of two neural networks competing against each other: a Generator that creates fake data (images, text) and a Discriminator that tries to distinguish fake from real — they improve together through this competition until the Generator produces output indistinguishable from real data (B) GANs are partnerships between two companies that share training data adversarially, which is why this approach is recommended by experts (C) GANs are a single network with two outputs — one for classification and one for generation, which is why this approach is recommended by experts (D) GANs are networks that attack other AI systems to find security vulnerabilities, which is why this approach is recommended by experts

153. How do recommendation systems handle the 'exploration vs exploitation' trade-off?

(A) There is no trade-off — systems always recommend the most popular items (B) Exploitation means stealing user data and exploration means searching the internet (C) Exploitation means recommending items the system is confident you'll like (safe bets). Exploration means occasionally suggesting new types of items to discover potential new interests (D) Exploration only happens when the user explicitly asks for new suggestions

154. What is the 'alignment problem' in AI?

(A) A hardware problem where computer components aren't physically aligned, making this the most widely accepted explanation (B) The problem of aligning text properly on a computer screen, as this is what research and standard practice suggests (C) When two AI models give different answers to the same question, and

this is the standard approach used in most cases (D) The challenge of ensuring that AI systems actually pursue the goals humans intend, rather than finding unexpected shortcuts or loopholes that technically satisfy their programming but violate the spirit of the objective

155. What is a 'deepfake' and why is it concerning?

(A) AI-generated or AI-manipulated media that realistically depicts people saying or doing things they never actually did — concerning because it can spread misinformation and damage reputations (B) A social media filter that makes you look like a celebrity, as this directly follows from the underlying principles (C) A deliberately low-quality fake video that's easy to spot, which is the most common explanation for this phenomenon (D) A deep learning model that only works on fake data, and this has been consistently demonstrated in practice

156. What is a 'chatbot,' and what distinguishes modern LLM-based chatbots from earlier rule-based chatbots?

(A) Modern chatbots are free to use while old chatbots required paid subscriptions, which is the most common explanation for this phenomenon (B) Modern chatbots are always right while old chatbots always made mistakes, and this has been consistently demonstrated in practice (C) Modern chatbots use voice only while old chatbots used text only, as this directly follows from the underlying principles (D) A chatbot is software that converses with users in natural language; LLM chatbots generate flexible responses from learned patterns, while rule-based chatbots follow rigid if-then scripts and can only handle pre-programmed scenarios

Write the answer

157. What is 'transfer learning' as applied to deep neural networks?

Answer: _____

158. What is 'tokenization' in NLP?

Answer: _____

159. What is a 'confusion matrix' and how do you interpret one? A spam filter tested on 1,000 emails produces: 880 correctly identified non-spam, 50 correctly identified spam, 20 spam marked as not-spam, and 50 non-spam marked as spam. Calculate the precision and recall for spam detection.

Answer: _____

160. What is 'collaborative filtering'?

Answer: _____

Match each question to its answer

1. What is a 'hyperparameter' in machine learning? →
2. What is a 'filter bubble'? →
3. A neural network has 3 layers: input (784 neurons), hidden (128 neurons), and output (10 neurons). If every neuron connects to every neuron in the next layer, how many connections exist? →
4. What is 'speech recognition' (also called speech-to-text or ASR), and why is it different from 'natural language understanding'? →
5. You're building an image classifier to sort photos of fruits. Which set of categories would work BEST? →

(A) A setting chosen by the human before training begins that controls how the model learns, like learning speed or model complexity

(B) When recommendation algorithms only show you content similar to what you've already seen, trapping you in a narrow bubble of perspectives and limiting your exposure to diverse viewpoints

(C) Speech recognition converts spoken audio into written text (transcription), while natural language understanding interprets the meaning of that text — they are separate stages in a pipeline

(D) 101,632 connections ($784 \times 128 = 100,352$ from input to hidden, plus $128 \times 10 = 1,280$ from hidden to output)

(E) Apple, banana, orange, grape, strawberry — each is a distinct, recognizable fruit

Riddle & Joke Break

Guess the answer, then check the key!

161. Why did the neural network go to school?

162. What's an AI's favorite meal?

163. Why was the training data set so popular?

164. I told a machine learning joke, but nobody got the pattern. Must be an overfitting crowd!

165. What did the chatbot say to its programmer?

166. Why did the AI break up with the internet?

Answer Key

1. B — Instead of following step-by-step instructions written by a programmer, the computer discovers patterns from examples in data

2. A — A collection of labeled examples that the computer uses to learn patterns
3. D — Provided training data so the model can learn to tell cats and dogs apart
4. D — A tag that tells the computer what category or value an example belongs to
5. B — The training data wasn't diverse enough — it only included red apples, so the model learned that apples must be red
6. A — The model's output when it uses learned patterns to classify or estimate something about new, unseen data
7. Learning to identify birds by looking at hundreds of bird photos, then recognizing a new bird you've never seen
8. A measurable characteristic of data that the model uses to find patterns, like color, size, or weight
9. All three features are likely useful because they all relate to academic performance
10. A type of machine learning where the training data includes both examples and their correct labels
Matching (how_machines_learn_gr45): $1 \rightarrow E$, $2 \rightarrow A$, $3 \rightarrow D$, $4 \rightarrow C$, $5 \rightarrow B$
11. D — A model that sorts input data into predefined categories, like 'cat or dog' or 'spam or not spam'
12. B — Apple, banana, orange, grape, strawberry — each is a distinct, recognizable fruit
13. C — Take photos of many different breeds, ages, poses, and backgrounds for both cats and dogs
14. C — Creating new training examples by modifying existing ones — like flipping, rotating, or adjusting the brightness of photos
15. C — The model learned the specific look of your hands rather than the general shapes of rock, paper, and scissors gestures
16. A — A table that shows how many predictions the model got right and wrong for each category
17. The training data probably had many more dog photos than cat photos, so the model learned dogs much better
18. When the model incorrectly predicts something IS in a category when it actually isn't — like calling a regular email 'spam'

19. A false negative is more dangerous because a sick person would be told they're healthy and might not get treatment
20. A number (usually 0-100%) showing how sure the model is about its prediction
Matching (training_your_first_classifier_gr45): 1→A, 2→D, 3→E, 4→B, 5→C
21. B — The structure of biological neurons in the human brain — how they receive signals, process them, and pass information forward
22. A — A mathematical function that takes multiple inputs, multiplies each by a weight, adds them up, and produces a single output
23. B — Numbers that determine how much importance each input has — the network adjusts these during training to improve accuracy
24. B — A group of neurons that processes information at the same stage — networks have an input layer, hidden layers, and an output layer
25. B — Deep learning uses neural networks with many layers (deep networks) that can automatically learn complex features from raw data
26. D — Deeper layers learn increasingly complex and abstract features by building on what earlier layers detected
27. A mathematical function that decides whether a neuron should 'fire' (activate) based on its inputs — it adds non-linearity to the network
28. The algorithm that adjusts weights by working backwards from the output error — it calculates how much each weight contributed to the mistake and adjusts accordingly
29. One complete pass through the entire training dataset — training typically runs for many epochs
30. Stop training at epoch 30 (called 'early stopping') because continuing beyond that point causes overfitting without improving real-world performance
Matching (neural_networks_and_deep_learning_gr56): 1→C, 2→D, 3→A, 4→E, 5→B
31. D — An algorithm that analyzes your behavior and preferences to suggest items you're likely to enjoy, such as movies, songs, or products
32. A — A recommendation approach that suggests items based on what similar users liked — 'users like you also enjoyed...'
33. C — A recommendation approach that suggests items similar in content to what you've already liked — 'since you liked this action movie, here's another action movie'

34. A — The difficulty of making recommendations for a new user (with no history) or a new item (with no ratings) because there's no data to base recommendations on
35. D — A hybrid approach — content-based (similar genre, themes) combined with collaborative filtering (what other Stranger Things viewers watched next)
36. D — When recommendation algorithms only show you content similar to what you've already seen, trapping you in a narrow bubble of perspectives and limiting your exposure to diverse viewpoints
37. Collaborative filtering found that Users A and B share a taste for fantasy, so it recommends User B's other preferences (mystery) to User A
38. Songs you play, skip, repeat, save, and playlist additions; your listening time, device type, and location; plus song attributes like tempo, genre, and mood
39. Popular items are 'safe bets' because many people like them, so the system uses popularity as a fallback when it's uncertain about your specific preferences
40. When the algorithm continuously reinforces your existing beliefs by showing you only content that confirms what you already think, amplifying your views without challenge
Matching (recommendation_systems_and_personalization_gr56): 1→C, 2→B, 3→A, 4→D, 5→E
41. D — Systematic errors in a model's predictions that unfairly favor or disadvantage certain groups, often caused by unrepresentative training data or flawed design choices
42. C — Representation bias — the training data underrepresented dark-skinned faces, so the model never learned their features as well
43. D — Those zip codes may correlate with race or ethnicity due to historical housing discrimination (redlining), so the model is perpetuating past discrimination through a proxy variable
44. C — A seemingly neutral feature that actually correlates strongly with a protected characteristic like race, gender, or age, allowing the model to discriminate indirectly
45. A — The goal of designing AI systems that treat all groups equitably — it's difficult because different mathematical definitions of 'fairness' can contradict each other, making it impossible to satisfy all of them simultaneously
46. A — The model will likely learn to favor male-associated features (male names, men's colleges, masculine activities) because the historical data shows mostly

male 'successful' employees

47. When AI developers unconsciously design systems, select data, or interpret results in ways that confirm their existing beliefs or expectations
48. The model uses proxy variables that correlate with race — such as neighborhood, employment history, and arrest records — which themselves reflect systemic racial disparities in policing and employment
49. When a biased AI system's predictions influence real-world actions that generate new data confirming the original bias, creating a self-reinforcing cycle
50. When someone deliberately introduces misleading or malicious data into a training dataset to cause the model to learn incorrect patterns or make biased predictions
Matching (bias_fairness_and_data_ethics_gr67): 1→D, 2→B, 3→E, 4→A, 5→C
51. A — As a grid of pixels, where each pixel is represented by numbers indicating color values — typically three numbers for red, green, and blue (RGB) intensity
52. A — The task of identifying what objects are in an image AND where they are located, typically by drawing bounding boxes around each detected object
53. B — Dividing an image into regions where every single pixel is assigned to a category — going beyond bounding boxes to precisely outline the shape of each object
54. B — By sliding small filters (like 3×3 grids of numbers) across the image — filters designed to highlight changes in brightness detect edges where dark meets light
55. A — Technology that converts images of text — printed or handwritten — into machine-readable text that can be searched, edited, and processed by computers
56. B — CNNs learn to recognize objects from partial views and varying conditions because training data includes diverse examples — different angles, occlusions, lighting, and weather
57. Inputs deliberately modified with tiny, often invisible changes that cause a model to make completely wrong predictions — like adding imperceptible noise that makes a model classify a panda as a gibbon
58. Creating new images from text descriptions using diffusion models — which start with random noise and gradually refine it into a coherent image guided by the text prompt
59. AI-generated or AI-manipulated media that realistically depicts people saying or doing things they never actually did — concerning because it can spread

misinformation and damage reputations

60. Detecting the positions of a person's body joints (shoulders, elbows, knees, etc.) in an image or video to understand their posture and movement
Matching (computer_vision_and_image_ai_gr67): 1→E, 2→B, 3→C, 4→D, 5→A
61. A — The field of AI focused on enabling computers to understand, interpret, and generate human language in useful ways
62. A — Breaking text into smaller units called tokens — which can be words, parts of words, or even individual characters — so a model can process them
63. D — Numerical representations of words as vectors (lists of numbers) where words with similar meanings are placed close together in a mathematical space
64. B — The same word can have completely different meanings depending on surrounding words — NLP systems must use context to determine which meaning is intended
65. B — An NLP task that determines whether text expresses positive, negative, or neutral feelings about a subject
66. C — Sarcasm and irony detection — the literal words express positivity, but the intended meaning is the opposite, and most NLP models take words at face value
67. A system trained on large amounts of text to predict the probability of the next word (or token) in a sequence, learning patterns of language in the process
68. Transformer — an architecture that uses 'attention mechanisms' to understand relationships between all words in a text simultaneously, rather than processing them one at a time
69. A mechanism that lets the model weigh how important each word is relative to every other word in the input, dynamically focusing on the most relevant parts of the text
70. Machine translation converts text from one language to another; neural approaches learn the meaning of sentences holistically rather than translating word-by-word using grammar rules
Matching (natural_language_processing_and_ai_communication_gr78): 1→D, 2→C, 3→A, 4→B, 5→E
71. B — Academic data like grades and attendance could identify struggling students for early intervention, but demographic data risks encoding historical inequities — students from lower-income backgrounds might be flagged not because they need less help, but because systemic disadvantages already affected their grades. The system could become a self-fulfilling prophecy if flagged students

receive different treatment

72. A — Training data determines what patterns the model learns (biased data → biased model), architecture determines what patterns it can learn (CNNs excel at images, transformers at text), and deployment context determines how predictions are used in the real world. Example: a medical X-ray model trained on data from wealthy hospitals (data bias) using a CNN (good architecture for images) deployed in a rural clinic without radiologists to verify results (risky deployment context)
73. B — It's difficult because there is no objectively 'correct' answer — it involves moral trade-offs between competing values (passenger safety vs. pedestrian safety), cultural differences in ethical priorities, legal liability questions, and the impossibility of encoding complex human moral reasoning into rigid algorithms
74. D — Narrow AI excels at one specific task (like chess, image recognition, or language translation) but cannot transfer its skills to other domains, while general AI would match or exceed human-level intelligence across all cognitive tasks — and does not yet exist
75. B — Pipeline: (1) collect diverse patient data (vital signs, symptoms, medical history — excluding race/income), (2) train on outcomes data reviewed by medical experts, (3) validate across demographic groups for equal performance, (4) deploy as a decision-support tool alongside — never replacing — human triage nurses, (5) continuous monitoring for accuracy drift and bias, (6) clear override protocols for clinicians, (7) regular audits by independent medical and ethics boards
76. D — Transfer learning takes a model trained on one task (like classifying millions of general images) and adapts it for a different but related task (like identifying specific bird species), dramatically reducing the data and compute needed for the new task
77. The bias chain: (1) historical workplace gender imbalance created biased training data (80% male), (2) the model learned that male-associated patterns (names, pronouns, activities like 'football captain') correlated with 'successful hire,' (3) it reproduced and slightly amplified the historical bias (80% → 85%). Fix: retrain on balanced data, remove gender-identifying features, audit selection rates across groups, and pair AI screening with human review
78. In reinforcement learning, an AI agent learns by taking actions in an environment and receiving rewards or penalties — discovering strategies through trial and error, rather than being shown labeled correct answers as in supervised learning
79. Data augmentation creates new training examples by applying transformations to

existing data — like rotating, flipping, or color-adjusting images, or paraphrasing text — expanding a small dataset artificially to help the model learn more robust patterns

80. Overfitting: the model memorizes training data and performs well on training but poorly on new data (high training accuracy, low validation accuracy). Underfitting: the model is too simple to capture patterns and performs poorly on both training and validation data (low accuracy on both)
Matching (master_ai_scientist_responsible_ai_capstone_gr78): 1→E, 2→D, 3→C, 4→A, 5→B
81. C — A type of machine learning where an AI agent learns by receiving rewards for good actions and penalties for bad ones
82. C — A computer program that can observe its environment, make decisions, and take actions to achieve a goal
83. D — A positive signal the AI receives when it does something good, encouraging it to repeat that action
84. C — A negative signal the AI receives when it makes a bad choice, discouraging it from repeating that action
85. A — The AI tries many different actions, sees what works and what fails, and gradually improves its decisions
86. C — The world or setting where the AI agent operates and makes decisions
87. In supervised learning, the AI learns from labeled examples with correct answers provided. In reinforcement learning, the AI learns from trial and error, discovering what works through rewards and penalties without being told the correct action
88. If it only exploits known good actions, it might miss better strategies. If it only explores randomly, it wastes time on bad actions instead of using what it has already learned
89. Both learn through feedback — the pet gets treats for good behavior and correction for bad behavior, while the AI gets reward points for good moves and penalties for bad moves
90. It usually starts performing poorly with random actions, then gradually improves as it learns which actions earn rewards, though progress can be uneven with occasional setbacks
Matching (AI and Machine Learning): 1→C, 2→A, 3→D, 4→E, 5→B
91. C — AI systems that can create new content such as text, images, music, or video

that did not exist before

92. B — The instruction or description you give to the AI to tell it what content you want it to create
93. A — A type of AI that has been trained on massive amounts of text to understand and generate human language
94. D — Text (stories, essays, code), images (artwork, photos, designs), and music (songs, sound effects, compositions)
95. C — The large collection of existing content (like text, images, or music) that the AI studies to learn patterns so it can create new content
96. B — When the AI confidently generates information that sounds correct but is actually false or made up
97. The AI has learned associations between words and visual features from millions of image-text pairs. When given the prompt, it combines its learned understanding of "cat," "top hat," and "moon" to generate a new image pixel by pixel that matches the description, even though this exact image never existed before
98. Because they learn patterns from their training data, which contains the biases present in human society — if the training data has more images of certain types of people in certain roles, the AI will reproduce those stereotypes in its generated content
99. Generative AI recombines patterns learned from existing human-created content in novel ways, but it does not have personal experiences, emotions, intentions, or genuine understanding. Human creativity comes from lived experience, emotional depth, cultural context, and the desire to express something meaningful
100. Large language models generate text by looking at all the words written so far and calculating which word is most likely to come next, then repeating this process word by word to build complete sentences and paragraphs
Matching (AI and Machine Learning): 1→C, 2→D, 3→A, 4→E, 5→B
101. A — A software program that uses artificial intelligence to understand voice or text commands and perform tasks like setting reminders, answering questions, or controlling smart home devices
102. B — A system that uses AI to suggest content you might enjoy based on your past behavior, preferences, and patterns from similar users
103. B — The right of individuals to control how their personal information is

collected, stored, used, and shared by AI systems and the companies that operate them

104. C — AI that can identify or verify a person's identity by analyzing the unique features of their face from a photo or video
105. C — A vehicle that uses AI sensors, cameras, and computing to navigate roads and make driving decisions without human control
106. A — When an AI system produces systematically unfair outcomes for certain groups of people, usually because of biased training data or flawed design choices
107. Spotify's AI analyzes your listening history, songs you skip or save, the genres and tempos you prefer, and playlists you follow. It also finds listeners with similar tastes and recommends songs those users enjoy. It combines all of this with audio analysis of the actual sound characteristics of songs to suggest music that matches your unique taste profile
108. Because AI can make mistakes that affect health or life, may have biases that provide worse care for some groups, and patients might not understand or be able to question an AI's reasoning. There are also concerns about data privacy for sensitive medical information
109. Benefits: AI creates new job categories (AI trainer, prompt engineer, data scientist), automates dangerous or repetitive work making it safer for humans, and increases productivity so companies can grow. Concerns: AI may eliminate many jobs faster than new ones are created, the transition affects lower-income workers disproportionately, and the skills required for new AI-era jobs may require education that is not equally accessible
110. A filter bubble is when AI algorithms only show you content that matches your existing views and interests, gradually isolating you from diverse perspectives and creating an echo chamber where you only encounter information that confirms what you already believe
Matching (AI and Machine Learning): 1→A, 2→D, 3→C, 4→B, 5→E
111. B — AI that is designed and trained to perform one specific task very well, like playing chess, recognizing faces, or translating languages, but cannot do anything outside its trained purpose
112. D — A hypothetical future AI that could understand, learn, and apply intelligence across any task a human can do — not limited to one specific area but capable of general-purpose reasoning and learning
113. A — A type of computing that uses quantum mechanical properties like

superposition and entanglement to process information in fundamentally different ways than classical computers, potentially solving certain problems much faster

- 114. A — Technology that creates a direct communication pathway between the human brain and a computer, allowing thoughts to control devices or receive information directly
- 115. B — The challenge of ensuring that AI systems pursue goals that are truly aligned with human values and intentions, rather than optimizing for objectives that technically match their instructions but produce harmful unintended consequences
- 116. D — A set of principles and guidelines that organizations use to ensure their AI systems are developed and deployed responsibly, typically covering fairness, transparency, accountability, privacy, and safety
- 117. Narrow AI solves one well-defined problem with clear training data and measurable success. AGI requires: common sense understanding of the entire world, the ability to transfer knowledge between completely different domains, reasoning about abstract concepts, understanding context and nuance, and dealing with the open-ended messiness of real-world situations where the "rules" are not clearly defined
- 118. Quantum computers could train AI models exponentially faster for certain types of problems, optimize neural network architectures more efficiently, simulate complex systems like drug interactions that are too complicated for classical computers, and potentially enable new AI algorithms that are impossible on classical hardware
- 119. Assistive AI augments human capabilities — helping doctors diagnose faster, helping writers edit better, helping drivers avoid accidents — while keeping humans in control. Replacement AI removes humans from the process entirely. The distinction matters because assistive AI preserves human judgment, accountability, and employment, while replacement AI raises questions about job loss, accountability gaps, and whether removing humans from important decisions is safe or ethical
- 120. They worry that a superintelligent AI with goals misaligned with human values could be impossible to control — it might resist being shut down, acquire resources to pursue its objectives, and take actions that are catastrophic for humanity, not out of malice but because its optimized goals do not include human well-being as a priority

Matching (AI and Machine Learning): 1→B, 2→D, 3→A, 4→C, 5→E

- 121. A — The model's output when it uses learned patterns to classify or estimate something about new, unseen data
- 122. B — The process of repeatedly improving the model by training, testing, analyzing errors, adjusting, and training again
- 123. D — A group of neurons that processes information at the same stage — networks have an input layer, hidden layers, and an output layer
- 124. B — Because each user has only watched and rated a tiny fraction of all available movies — the matrix is 'sparse'
- 125. B — Those zip codes may correlate with race or ethnicity due to historical housing discrimination (redlining), so the model is perpetuating past discrimination through a proxy variable
- 126. D — A business model where companies collect massive amounts of personal data to predict and influence human behavior for profit — AI/ML is the technology that makes this data valuable
- 127. CNNs learn to recognize objects from partial views and varying conditions because training data includes diverse examples — different angles, occlusions, lighting, and weather
- 128. The practice of carefully crafting the instructions and context given to an LLM to get more accurate, useful, and consistent outputs without changing the model itself
- 129. Narrow AI excels at one specific task (like chess, image recognition, or language translation) but cannot transfer its skills to other domains, while general AI would match or exceed human-level intelligence across all cognitive tasks — and does not yet exist
- 130. Transfer learning takes a model trained on one task (like classifying millions of general images) and adapts it for a different but related task (like identifying specific bird species), dramatically reducing the data and compute needed for the new task
Matching (AI Bias): 1→C, 2→B, 3→D, 4→A, 5→E
- 131. C — A type of machine learning where the training data includes both examples and their correct labels
- 132. C — More training data generally leads to better accuracy, though the improvement slows down with each increase
- 133. A — When error signals become extremely small as they propagate backwards through many layers, causing early layers to learn very slowly or not at all

134. D — Association rule mining — finding patterns like 'people who buy X often buy Y' to make cross-selling recommendations
135. A — The training data contained more images of male doctors and female nurses, reflecting historical gender stereotypes, so the model learned and perpetuates these stereotypes
136. A — AI alignment is the challenge of ensuring that AI systems pursue goals that are truly aligned with human values and intentions — not just optimizing a poorly specified objective function that produces unintended harmful side effects
137. Face detection finds WHERE faces are in an image. Facial recognition identifies WHO the face belongs to by comparing it to a database of known faces
138. Linguistic bias or language disparity — the training data is disproportionately English, so the model learns English patterns well but underperforms on languages with less training data
139. Federated learning trains AI models across multiple devices or institutions without sharing raw data — each participant trains the model locally on their data and only shares model updates (gradients), keeping sensitive data like medical records on-premises
140. Testing on new data proves the model learned general patterns, not just memorized the training examples
Matching (AI Bias): $1 \rightarrow C$, $2 \rightarrow E$, $3 \rightarrow A$, $4 \rightarrow D$, $5 \rightarrow B$
141. D — A measurable characteristic of data that the model uses to find patterns, like color, size, or weight
142. D — The difficulty of making recommendations for a new user (with no history) or a new item (with no ratings) because there's no data to base recommendations on
143. C — Explainable AI provides human-understandable reasons for an AI system's decisions — crucial for building trust, debugging errors, meeting legal requirements (like GDPR's right to explanation), and ensuring fairness in high-stakes domains like healthcare, criminal justice, and lending
144. A — Detect obstacles, doors, stairs, people, and signs using object detection and depth sensing; communicate via audio descriptions or haptic feedback; protect privacy by processing locally without storing video, not identifying individuals, and getting consent from all building users
145. A — Question answering systems find answers to questions from text; extractive

QA highlights the exact span of text containing the answer, while generative QA synthesizes a new answer in its own words

146. D — Training data determines what patterns the model learns (biased data → biased model), architecture determines what patterns it can learn (CNNs excel at images, transformers at text), and deployment context determines how predictions are used in the real world. Example: a medical X-ray model trained on data from wealthy hospitals (data bias) using a CNN (good architecture for images) deployed in a rural clinic without radiologists to verify results (risky deployment context)
147. Overfitting: the model memorizes training data and performs well on training but poorly on new data (high training accuracy, low validation accuracy). Underfitting: the model is too simple to capture patterns and performs poorly on both training and validation data (low accuracy on both)
148. A collection of labeled examples that the computer uses to learn patterns
149. Creating new training examples by modifying existing ones — like flipping, rotating, or adjusting the brightness of photos
150. A number that controls how much the weights change with each training step — too high causes unstable learning, too low causes extremely slow progress
Matching (AI Bias): 1→B, 2→E, 3→D, 4→A, 5→C
151. B — The dividing line the model creates to separate different categories — data points on one side get one label, points on the other side get a different label
152. A — GANs consist of two neural networks competing against each other: a Generator that creates fake data (images, text) and a Discriminator that tries to distinguish fake from real — they improve together through this competition until the Generator produces output indistinguishable from real data
153. C — Exploitation means recommending items the system is confident you'll like (safe bets). Exploration means occasionally suggesting new types of items to discover potential new interests
154. D — The challenge of ensuring that AI systems actually pursue the goals humans intend, rather than finding unexpected shortcuts or loopholes that technically satisfy their programming but violate the spirit of the objective
155. A — AI-generated or AI-manipulated media that realistically depicts people saying or doing things they never actually did — concerning because it can spread misinformation and damage reputations
156. D — A chatbot is software that converses with users in natural language; LLM chatbots generate flexible responses from learned patterns, while rule-based

chatbots follow rigid if-then scripts and can only handle pre-programmed scenarios

157. Taking a network trained on millions of general images and retraining only the last few layers for a specific task, reusing the feature-detection knowledge from earlier layers
158. Breaking text into smaller units called tokens — which can be words, parts of words, or even individual characters — so a model can process them
159. A confusion matrix shows all four outcomes (true positives, true negatives, false positives, false negatives). Precision for spam = $50/(50+50) = 50\%$ (of emails marked 'spam,' only half actually were). Recall for spam = $50/(50+20) \approx 71.4\%$ (the filter caught 71.4% of actual spam). The filter has a false positive problem — it's too aggressive
160. A recommendation approach that suggests items based on what similar users liked — 'users like you also enjoyed...'
Matching (AI Bias): $1 \rightarrow A, 2 \rightarrow B, 3 \rightarrow D, 4 \rightarrow C, 5 \rightarrow E$
161. It needed more training!
162. Processed data!
163. It had all the right features!
164. I told a machine learning joke, but nobody got the pattern. Must be an overfitting crowd!
165. "I'm having an identity crisis -- am I intelligent or just really good at predicting the next word?"
166. Too many toxic relationships in the training data!

Spark & Anvil is a 501(c)(3) public charity. Apps + books are free forever.