



The SESK Party Game

STUFF EVERY AI EXPERT SHOULD KNOW

How to play (30 seconds)

1. The reader reads a card and its three answers out loud.
2. Everyone says "One, two, three, Goldilocks!" and holds up 1, 2, or 3 fingers.
3. The reader flips the card and reads the answer and the Why.
4. Everyone who was right takes a point.
5. The reader reads the joke with a straight face. Anyone laughs or groans? The reader takes a point. Nobody got it right? The reader takes a point.
6. Pass the deck left. Most points when the points run out wins.

What you need

These cards and the point tokens on the last two pages. No tokens? Use coins, buttons, pebbles, or marshmallows. 3 to 10 players, ages 8 and up.

Expert rule: Bet It

Before the reader flips the card, you may put one of your points on your answer. Right: take your normal point plus one extra. Wrong: your bet goes back to the pile. That's it.

Printing

Print single-sided, on cardstock if you have it. Cut on the dashed lines. Each card is a 3.5 x 5 inch index card. The answer half is upside down on purpose, so the table can't peek. Cards 1 to 50 of 170. More cards and the free daily online game at sesk.info. Built 2026-10-12.

ARTIFICIAL INTELLIGENCE



How do AI, machine learning and deep learning fit together?

- A** Deep learning is inside machine learning, inside AI
- B** Three separate fields that never overlap
- C** AI is one small part of deep learning

FLIP FOR THE ANSWER

#3557

SESK.INFO

STRAIGHT FACE
Machine learning is like a set of nesting dolls. Open AI, and there is another smart doll inside.

A Deep learning is inside machine learning, inside AI
WHY AI is the big umbrella. Machine learning is the part of AI where systems learn from data. Deep learning is a subfield of machine learning built around neural networks.

ARTIFICIAL INTELLIGENCE



What did Alan Turing call his 1950 test of machine intelligence?

- A** The Thinking Machine Test
- B** The Imitation Game
- C** The Chinese Room

FLIP FOR THE ANSWER

#3558

SESK.INFO

STRAIGHT FACE
I tried the Turing test on my toaster. It failed, but it did get a little warm.

B The Imitation Game
WHY In his 1950 paper, Turing proposed a game where an interrogator chats with a person and a machine and tries to tell which is which. He called it the imitation game.

ARTIFICIAL INTELLIGENCE



Turing predicted that by about 2000, judges would ID the machine at most what percent of the time?

- A** 50 percent, after one hour
- B** 99 percent, after one minute
- C** 70 percent, after five minutes

FLIP FOR THE ANSWER

#3559

SESK.INFO

STRAIGHT FACE
Why was the robot so calm in the interview? It had nerves of steel.

C 70 percent, after five minutes
WHY In 1950 Turing guessed that in about fifty years, an average interrogator would have no more than a 70 percent chance of spotting the machine after five minutes of questions.

ARTIFICIAL INTELLIGENCE



What 1956 event is called the birth of AI as a field of research?

- A** The Dartmouth summer research project
- B** The launch of ENIAC
- C** The first chess computer match

FLIP FOR THE ANSWER

#3560

SESK.INFO

STRAIGHT FACE
The Dartmouth scientists planned a summer of AI. Seventy years later, we are still on the first homework assignment.

A The Dartmouth summer research project
WHY In 1956, a small group of scientists met at Dartmouth College for a summer project on artificial intelligence. John McCarthy organized it, and it is seen as the start of AI as a field.

ARTIFICIAL INTELLIGENCE



Which approach represents the original top-down AI paradigm?

- A** Neural networks trained on data
- B** Reinforcement learning from rewards
- C** Logic-based symbolic AI

FLIP FOR THE ANSWER

#3561

SESK.INFO

STRAIGHT FACE
Symbolic AI and neural networks had an argument. One had all the rules, the other just had a feeling about it.

C Logic-based symbolic AI
WHY Early AI wrote knowledge as symbols and logical rules that a program reasons with, top-down. Statistical approaches like neural networks instead learn patterns from lots of examples.

ARTIFICIAL INTELLIGENCE



What makes machine learning "supervised"?

- A** A human watches the computer while it trains
- B** It trains on examples paired with correct labels
- C** It learns only from examples without labels

FLIP FOR THE ANSWER

#3562

SESK.INFO

STRAIGHT FACE
My model wanted unsupervised learning. I said not until it finishes its homework.

B It trains on examples paired with correct labels
WHY Supervised learning trains a model on features plus their matching labels, the known answers. It is like studying with a set of questions and their answer key.

ARTIFICIAL INTELLIGENCE



What kind of data does unsupervised learning usually train on?

- A** Unlabeled data, to find patterns
- B** Data where every row has the answer
- C** Rewards from a game score

FLIP FOR THE ANSWER

#3563

SESK.INFO

STRAIGHT FACE
Unsupervised learning is like a kid in a toy store with no parents. It still finds patterns, mostly in the candy aisle.

A Unlabeled data, to find patterns
WHY Unsupervised learning has no answer key. The model looks for structure, such as groups of similar customers, in data that has no labels.

ARTIFICIAL INTELLIGENCE



In reinforcement learning, what tells the agent how well an action went?

- A** A label written by a human for each row
- B** The clustering score
- C** A numerical reward from the environment

FLIP FOR THE ANSWER

#3564

SESK.INFO

STRAIGHT FACE
My dog uses reinforcement learning. Sit equals treat. Bark at mailman equals no treat but worth it.

C A numerical reward from the environment
WHY A reinforcement learning agent takes actions in an environment and gets a number back, the reward. It learns a policy that earns the most total reward over time.

ARTIFICIAL INTELLIGENCE



In a model predicting house prices from size and bedrooms, what is the label?

- A** The house price
- B** The number of bedrooms
- C** The square footage

FLIP FOR THE ANSWER

#3565

SESK.INFO

STRAIGHT FACE
I asked the model what my house is worth. It said, "Let me check my features first."

A **WHY** The house price
Features are the inputs the model looks at. The label is the answer it is learning to predict, here the price.

ARTIFICIAL INTELLIGENCE



What makes a regression model different from a classification model?

- A** Regression predicts a category, not a number
- B** Regression predicts a number, not a class
- C** Regression outputs only a yes-or-no decision

FLIP FOR THE ANSWER

#3566

SESK.INFO

STRAIGHT FACE
My regression model predicted my age. I told it to round down.

B **WHY** Regression predicts a number, not a class
Classification picks a category, like cat or dog. Regression predicts a numeric value, like tomorrow's temperature or a house price.

ARTIFICIAL INTELLIGENCE



In machine learning, what is inference?

- A** Adjusting the weights to reduce loss
- B** Collecting and labeling the training data
- C** Using a trained model to make predictions

FLIP FOR THE ANSWER

#3567

SESK.INFO

STRAIGHT FACE
My model finished training and asked for a raise. I said, let us see your inference first.

C **WHY** Using a trained model to make predictions
Training is when the model learns its weights from data. Inference comes after: the finished model is applied to new examples to produce predictions.

ARTIFICIAL INTELLIGENCE



Which data split is saved for the final check after tuning on the validation set?

- A** The test set
- B** The training set
- C** The validation set again

FLIP FOR THE ANSWER

#3568

SESK.INFO

STRAIGHT FACE
The test set is like the final exam. No peeking, even if the model asks nicely.

A **WHY** The test set
You train on the training set and compare versions on the validation set. When the model looks good, the test set gives a last, fresh double-check on data it has not been tuned to.

ARTIFICIAL INTELLIGENCE



A model scores great on its training data but badly on new data. What is that called?

- A** Underfitting
- B** Overfitting
- C** Regularization

FLIP FOR THE ANSWER

#3569

SESK.INFO

STRAIGHT FACE
My model memorized every answer on the practice test. Then the real test had different questions. Classic overfitting.

B Overfitting
WHY Overfitting means the model matches the training data so closely, quirks and noise included, that it fails to make correct predictions on new data.

ARTIFICIAL INTELLIGENCE



A model does poorly even on its own training data. What is the likely problem?

- A** Overfitting
- B** Data leakage
- C** Underfitting

FLIP FOR THE ANSWER

#3570

SESK.INFO

STRAIGHT FACE
My underfit model predicted every house costs the same. Great for buyers, terrible for science.

C Underfitting
WHY An underfit model is too simple or undertrained, so it misses the real patterns. It scores badly on training data and new data alike.

ARTIFICIAL INTELLIGENCE



Which term measures error from a model's assumptions, even with infinite training data?

- A** Bias
- B** Variance
- C** Irreducible noise

FLIP FOR THE ANSWER

#3571

SESK.INFO

STRAIGHT FACE
My model has a bias for straight lines. I told it life has curves.

A Bias
WHY Bias comes from a model's built-in assumptions, like forcing a straight line. Even infinite training data cannot remove that model error. Variance instead measures how much the fitted model changes across training sets.

ARTIFICIAL INTELLIGENCE



Training error is far lower than test error. Which problem does that point to?

- A** High bias
- B** High variance
- C** Irreducible noise

FLIP FOR THE ANSWER

#3572

SESK.INFO

STRAIGHT FACE
My model aced the practice quiz and bombed the test. Its teacher calls that high variance; I call it cramming.

B High variance
WHY Variance is how much the model changes with a different training set. A big gap between low training error and high test error means it over-specialized on its training data.

ARTIFICIAL INTELLIGENCE



What does L2 regularization penalize?

- A** The sum of the absolute weights
- B** The count of nonzero weights
- C** The sum of the squared weights

FLIP FOR THE ANSWER

#3573

SESK.INFO

STRAIGHT FACE L2 regularization is a diet for weights. Big ones get squared away.

C **WHY** The sum of the squared weights L2 regularization adds a penalty for big weights to the loss. It pulls large weights toward zero, but not all the way, which keeps the model simpler and less likely to overfit.

ARTIFICIAL INTELLIGENCE



Which regularization can push useless features' weights to exactly zero?

- A** L1 regularization
- B** L2 regularization
- C** Dropout regularization

FLIP FOR THE ANSWER

#3574

SESK.INFO

STRAIGHT FACE L1 regularization is the strictest coach. If you do not help the team, your weight goes to zero.

A **WHY** L1 regularization penalizes the sum of the absolute values of the weights. It drives weights of irrelevant features to exactly 0, removing them from the model. L2 only shrinks them close to 0.

ARTIFICIAL INTELLIGENCE



With early stopping, when do you end training?

- A** When training loss reaches exactly zero
- B** When validation loss starts to increase
- C** After exactly 100 epochs

FLIP FOR THE ANSWER

#3575

SESK.INFO

STRAIGHT FACE I use early stopping at dinner. When dessert loss starts rising, I put the fork down.

B **WHY** When validation loss starts to increase Training loss may keep falling, but once validation loss turns upward, the model is starting to overfit. Stopping there is a simple form of regularization.

ARTIFICIAL INTELLIGENCE



Which loss function punishes big errors, like outliers, more heavily: MSE or MAE?

- A** Mean absolute error (MAE)
- B** They punish outliers equally
- C** Mean squared error (MSE)

FLIP FOR THE ANSWER

#3576

SESK.INFO

STRAIGHT FACE MSE is that friend who never lets one mistake go. It squares it and brings it up forever.

C **WHY** Mean squared error (MSE) squares each difference between prediction and label. An error of 10 counts as 100, so large misses dominate. MAE just averages the absolute differences.

ARTIFICIAL INTELLIGENCE



In gradient descent, which way does each step move the weights?

- A** In the direction of the negative slope
- B** In a random direction each time
- C** Uphill, toward higher loss

FLIP FOR THE ANSWER

#3577

SESK.INFO

STRAIGHT FACE
Gradient descent is just a computer looking for the nearest way downhill. Same as me on a bike.

A In the direction of the negative slope
WHY Gradient descent finds the slope (gradient) of the loss and takes a small step the opposite way, downhill. It repeats until the loss stops dropping.

ARTIFICIAL INTELLIGENCE



What happens if the learning rate is set too high?

- A** Training takes many small steps but converges
- B** The model overshoots and never converges
- C** The model converges quickly and smoothly

FLIP FOR THE ANSWER

#3578

SESK.INFO

STRAIGHT FACE
My learning rate was so high the model skipped right over the answer and landed in next week.

B The model overshoots and never converges
WHY The learning rate sets the step size. Too small and training crawls. Too big and every step overshoots the low point, so loss jumps around instead of settling.

ARTIFICIAL INTELLIGENCE



With 1,000 training examples and a batch size of 50, how many iterations make one epoch?

- A** 50
- B** 1,000
- C** 20

FLIP FOR THE ANSWER

#3579

SESK.INFO

STRAIGHT FACE
My model wanted to watch the training data one more time. I said fine, but this is the last epoch.

C 20
WHY An epoch is one full pass over the training set, every example seen once. Each iteration processes one batch, so an epoch takes N divided by batch size iterations.

ARTIFICIAL INTELLIGENCE



If 1 percent of cases are positive, what accuracy does always guessing negative score?

- A** 99 percent
- B** 50 percent
- C** 1 percent

FLIP FOR THE ANSWER

#3580

SESK.INFO

STRAIGHT FACE
My fraud detector is 99 percent accurate. It just says "no fraud" every time.

A 99 percent
WHY Accuracy counts all correct calls. When positives are rare, a useless model that always says no is right 99 times in 100. That is why accuracy misleads on imbalanced data.

ARTIFICIAL INTELLIGENCE



A model flags 200 emails as spam; 150 really are. What is its precision?

- A** 0.25
- B** 0.75
- C** 1.33

FLIP FOR THE ANSWER

#3581

SESK.INFO

STRAIGHT FACE
My spam filter has 75 percent precision. The other 25 percent was my grandma's emails.

B 0.75
WHY Precision asks: when the model said positive, how often was it right? True positives divided by all positive predictions: 150 over 200 is 0.75.

ARTIFICIAL INTELLIGENCE



Which metric answers: of all the real positives, how many did the model catch?

- A** Precision
- B** Accuracy
- C** Recall

FLIP FOR THE ANSWER

#3582

SESK.INFO

STRAIGHT FACE
My memory has terrible recall. I can never find all the positives in my day.

C Recall
WHY Recall is true positives divided by true positives plus false negatives. A cancer screen with high recall misses very few real cases.

ARTIFICIAL INTELLIGENCE



The F1 score combines precision and recall using which kind of average?

- A** The harmonic mean
- B** The simple arithmetic mean
- C** The median

FLIP FOR THE ANSWER

#3583

SESK.INFO

STRAIGHT FACE
F1 is the strict parent of metrics. One bad grade and the whole report card drops.

A The harmonic mean
WHY F1 is 2 times precision times recall, divided by precision plus recall. The harmonic mean drags the score down if either one is low, so a model cannot hide a bad recall behind great precision.

ARTIFICIAL INTELLIGENCE



What does a confusion matrix show for a classifier?

- A** How long each prediction took
- B** The model's weights after training
- C** Counts of right and wrong predictions per class

FLIP FOR THE ANSWER

#3584

SESK.INFO

STRAIGHT FACE
I made a confusion matrix of my weekend plans. Mostly false positives.

C Counts of right and wrong predictions per class
WHY A confusion matrix is an N by N table for N classes. For yes/no tasks its four cells are true positives, false positives, true negatives and false negatives.

ARTIFICIAL INTELLIGENCE



What AUC (area under the ROC curve) does a model that just guesses randomly get?

- A** 0
- B** 0.5
- C** 1.0

FLIP FOR THE ANSWER

#3585

SESK.INFO

STRAIGHT FACE
My coin flip model has an AUC of 0.5. It is confident, just not helpful.

B 0.5
WHY AUC is the chance the model ranks a random positive above a random negative. Random guessing gets that right half the time, so 0.5. A perfect ranker scores 1.0.

ARTIFICIAL INTELLIGENCE



How is a random forest built?

- A** One very deep decision tree
- B** Many trees trained on random samples, combined
- C** A neural network shaped like a tree

FLIP FOR THE ANSWER

#3586

SESK.INFO

STRAIGHT FACE
A single decision tree overfits. A random forest overfits less, because the trees argue it out.

B Many trees trained on random samples, combined
WHY A random forest trains each tree on a bootstrap sample (drawn with replacement) and a random subset of features. Averaging many different trees lets their errors partly cancel out.

ARTIFICIAL INTELLIGENCE



Why are decision trees called a "white box" model?

- A** Their rules always fit nonlinear data
- B** Their predictions ignore the training data
- C** Each prediction follows readable if/then logic

FLIP FOR THE ANSWER

#3587

SESK.INFO

STRAIGHT FACE
My decision tree asked me twenty yes or no questions and decided I wanted pizza. It was right.

C Each prediction follows readable if/then logic
WHY A decision tree is a chain of conditions, like "is age over 30?", ending in a leaf. You can read the path for any prediction, unlike a neural network's millions of weights.

ARTIFICIAL INTELLIGENCE



How does k-nearest neighbors classify a new point?

- A** A majority vote of its closest training points
- B** It follows a tree of yes/no questions
- C** It averages all the weights of a network

FLIP FOR THE ANSWER

#3588

SESK.INFO

STRAIGHT FACE
k-nearest neighbors is how I pick a restaurant I ask the three closest people and go with the majority.

A A majority vote of its closest training points
WHY k-NN does not build a model. It stores the training data, finds the k points nearest the new one, and gives it the class most of those neighbors have.

ARTIFICIAL INTELLIGENCE



What must you tell the k-means algorithm before it starts?

- A** The correct label for each point
- B** The number of clusters to make
- C** The exact coordinates of every centroid

FLIP FOR THE ANSWER

#3589

SESK.INFO

STRAIGHT FACE
I asked k-means to organize my sock drawer. It wanted to know how many piles first.

B **WHY** The number of clusters to make
k-means needs k up front. It then places k centroids, assigns each point to the nearest one, moves each centroid to the middle of its points, and repeats.

ARTIFICIAL INTELLIGENCE



In 1958, what did Rosenblatt's perceptron learn to tell apart after 50 trials?

- A** Cats from dogs in photos
- B** Spoken words yes and no
- C** Cards marked on the left from the right

FLIP FOR THE ANSWER

#3590

SESK.INFO

STRAIGHT FACE
The first perceptron could only tell left from right. Honestly, same as me before coffee.

C **WHY** Cards marked on the left from the right
In July 1958 the US Office of Naval Research showed Frank Rosenblatt's perceptron running on an IBM 704. After 50 trials it taught itself which side of a punch card was marked.

ARTIFICIAL INTELLIGENCE



Which 1969 book by Minsky and Papert sharply criticized perceptron research?

- A** Perceptrons
- B** Computing Machinery and Intelligence
- C** The Society of Mind

FLIP FOR THE ANSWER

#3591

SESK.INFO

STRAIGHT FACE
Minsky and Papert wrote a whole book about perceptrons. The perceptrons did not get to write a reply.

A **WHY** Perceptrons
Marvin Minsky and Seymour Papert's 1969 book attacked the limits of Rosenblatt's perceptrons so hard that interest in neural network research faded for years.

ARTIFICIAL INTELLIGENCE



Which 1973 UK report helped cause a sharp drop in AI research funding?

- A** The Dartmouth Proposal
- B** The Lighthill Report
- C** The Alvey Report

FLIP FOR THE ANSWER

#3592

SESK.INFO

STRAIGHT FACE
The first AI winter was so cold, even the robots asked for a sweater.

B **WHY** The Lighthill Report
The UK Science Research Council commissioned Sir James Lighthill to review AI. His 1973 report was deeply skeptical of basic AI research, and funding in Britain dried up quickly afterward.

ARTIFICIAL INTELLIGENCE



About how many parameters did AlexNet, the 2012 ImageNet breakthrough network, have?

- A** About 60 thousand
- B** About 60 million
- C** About 60 billion

FLIP FOR THE ANSWER

#3593

SESK.INFO

STRAIGHT FACE
AlexNet had 60 million parameters. I have trouble with three parameters for dinner.

B About 60 million
WHY Krizhevsky, Sutskever and Hinton's network had 60 million parameters and five convolutional layers. An efficient GPU implementation helped make its training faster.

ARTIFICIAL INTELLIGENCE



What did AlexNet's authors use to reduce overfitting in its fully connected layers?

- A** Dropout regularization
- B** A smaller dataset
- C** A higher learning rate

FLIP FOR THE ANSWER

#3594

SESK.INFO

STRAIGHT FACE
Dropout is like randomly benching players at practice. The whole team gets better.

A Dropout regularization
WHY AlexNet was big enough to overfit. Its authors used a then-new regularization method, dropout, which removes a random selection of units at each training step so the network cannot lean on any one of them.

ARTIFICIAL INTELLIGENCE



ImageNet's millions of labeled images are organized by which word hierarchy?

- A** The Dewey Decimal System
- B** Wikipedia categories
- C** WordNet

FLIP FOR THE ANSWER

#3595

SESK.INFO

STRAIGHT FACE
ImageNet has 14 million pictures and still no photo of where I left my keys.

C WordNet
WHY ImageNet follows the WordNet hierarchy. Each meaningful concept in WordNet is a synonym set, or synset, and may be described by multiple words or phrases.

ARTIFICIAL INTELLIGENCE



What does the "deep" in deep learning refer to?

- A** How much the model understands
- B** The number of layers in the network
- C** How large the training dataset is

FLIP FOR THE ANSWER

#3596

SESK.INFO

STRAIGHT FACE
My neural network is deep. It just stares at the ocean and thinks about gradients.

B The number of layers in the network
WHY Neural networks stack layers of artificial neurons: input, hidden and output. IBM describes a network with more than three layers, counting input and output, as deep learning.

ARTIFICIAL INTELLIGENCE



Which of these is a hyperparameter you set, not something the model learns?

- A** The model's weights
- B** The learning rate
- C** The model's bias values

FLIP FOR THE ANSWER

#3597

SESK.INFO

STRAIGHT FACE
Hyperparameters are the settings you argue about. Parameters are the ones the model quietly figures out.

B The learning rate
WHY Parameters, the weights and biases, are learned during training. Hyperparameters like learning rate, batch size and number of epochs are chosen by people (or a tuning service) between training runs.

ARTIFICIAL INTELLIGENCE



Raising a classifier's threshold usually does what to false positives and false negatives?

- A** Fewer false positives, more false negatives
- B** More false positives, fewer false negatives
- C** Both go down together

FLIP FOR THE ANSWER

#3598

SESK.INFO

STRAIGHT FACE
I raised my threshold for dad jokes. Now I miss a lot of good ones, but none of the bad ones get through.

A Fewer false positives, more false negatives
WHY A higher threshold means the model says positive only when very sure. It makes fewer wrong positive calls but misses more real positives. So precision tends to rise while recall falls.

ARTIFICIAL INTELLIGENCE



What calculation does a basic artificial neuron perform before activation?

- A** Weighted sum plus bias
- B** Median of its inputs
- C** Largest input only

FLIP FOR THE ANSWER

#3599

SESK.INFO

STRAIGHT FACE
Why did the neuron carry a calculator? It had too many weighted decisions.

A Weighted sum plus bias
WHY Each input is multiplied by its own weight. The products are added, then a bias is added before the activation function acts.

ARTIFICIAL INTELLIGENCE



What can a neuron's bias change even when its inputs stay fixed?

- A** The training-set size
- B** The neuron's pre-activation
- C** The number of inputs

FLIP FOR THE ANSWER

#3600

SESK.INFO

STRAIGHT FACE
My neuron admitted its bias. That was a positive adjustment.

B The neuron's pre-activation
WHY A bias is added after the weighted inputs are summed. It can move the result up or down without changing any input value.

ARTIFICIAL INTELLIGENCE



Why place nonlinear activation functions between neural network layers?

- A** To sort every input
- B** To remove all parameters
- C** To model nonlinear relationships

FLIP FOR THE ANSWER

#3601

SESK.INFO

STRAIGHT FACE
The activation function took a curve because straight lines were too predictable.

C **To model nonlinear relationships**
WHY A stack of linear operations still collapses to one linear operation. Nonlinear activations let deeper networks represent far more complicated functions.

ARTIFICIAL INTELLIGENCE



What does the ReLU activation return for an input x ?

- A** $\max(0, x)$
- B** 1 divided by x
- C** x squared

FLIP FOR THE ANSWER

#3602

SESK.INFO

STRAIGHT FACE
ReLU stays positive, but only when the input does.

A **$\max(0, x)$**
WHY Negative inputs become zero. Zero and positive inputs pass through unchanged, making ReLU simple and piecewise linear.

ARTIFICIAL INTELLIGENCE



Into what open interval does the logistic sigmoid map real inputs?

- A** Negative infinity to infinity
- B** 0 to 1
- C** -1 to 1

FLIP FOR THE ANSWER

#3603

SESK.INFO

STRAIGHT FACE
Sigmoid never reaches one, but it keeps approaching the subject

B **0 to 1**
WHY The logistic sigmoid is 1 divided by (1 plus e to the power negative x). It approaches 0 and 1 but never reaches either for finite x .

ARTIFICIAL INTELLIGENCE



Where is a neural network's hidden layer located?

- A** After the loss function
- B** Inside the dataset
- C** Between input and output

FLIP FOR THE ANSWER

#3604

SESK.INFO

STRAIGHT FACE
Why was the hidden layer hard to find? It stayed between the lines.

C **Between input and output**
WHY Input layers receive features and output layers produce predictions. Any intervening computational layers are called hidden layers.

ARTIFICIAL INTELLIGENCE



Which algorithm makes gradient descent practical for multilayer neural networks?

- A** Backpropagation
- B** Tokenization
- C** Beam search

FLIP FOR THE ANSWER

#3605

SESK.INFO

STRAIGHT FACE
Backpropagation learned from its mistakes, in reverse order.

A **Backpropagation**
WHY Backpropagation applies the chain rule from the loss backward through the computation graph. Those gradients tell an optimizer how parameters affect the loss.

ARTIFICIAL INTELLIGENCE



What happens to layers whose training gradients vanish?

- A** They double in width
- B** They barely train
- C** They tokenize inputs

FLIP FOR THE ANSWER

#3606

SESK.INFO

STRAIGHT FACE
The gradient vanished before paying the loss.

B **They barely train**
WHY Deep backpropagation can multiply many small derivative terms. The resulting near-zero gradient gives early layers almost no useful update signal.



