

Lecture 1

How AI Learns the World

- What is AI?
- Three Ways of Learning
Supervised Learning, Unsupervised Learning, Reinforcement Learning
- Optimization

Supervised Learning

Features and Labels

Why is separate identification feasible?

Fruits of the same category are similar to each other.




Image Demonstration

Features:

Characteristics of the objects.

- Useful: colors, textures
- Unreliable: oval shape, shine

Labels: what we are trying to predict based on an input.

- Category of the fruits

Optimization

Definition

Optimization: Finding the best possible solution to a machine learning problem




Image Demonstration

Supervised Learning:
minimize the difference
between the label and the
generated output




Image Demonstration

Anomaly Detection:
minimize the costs/risks of
false results

Clustering: distinct clusters
while also having individual
data points within a cluster as
similar as possible.




Image Demonstration

Reinforcement Learning:
maximize the accumulated
reward received

Table S1. Subjective Questions: NASA TLX Form

Question ID	Question	Choice
1	How mentally demanding when you were using the online system to learn?	Very Low (1) to Very High (7)
2	How physically demanding when you were using the online system to learn?	Very Low (1) to Very High (7)
3	How hurried or rushed was the pace when you were using the online system to learn?	Very Low (1) to Very High (7)
4	How successful were you in accomplishing learning using the online system?	Perfect (1) to Failure (7)
5	How hard did you have to work to accomplish your learning using the online system?	Very Low (1) to Very High (7)
6	How insecure, discouraged, irritated, stressed, and annoyed were you when you were using the online system to learn?	Very Low (1) to Very High (7)

Table S2. Subjective Questions: Feedback Experience

Question ID	Question	Choice
1	I was aware of other students' engagement during the class.	Strongly Disagree (1) to Strongly Agree (5)
2	The feedback helps me understand the states of fellow students.	Strongly Disagree (1) to Strongly Agree (5)
3	The feedback helps me gain a sense of participation/belongingness.	Strongly Disagree (1) to Strongly Agree (5)
4	The feedback helps me follow the pace of the instructor during classes.	Strongly Disagree (1) to Strongly Agree (5)
5	I could be more engaged and less absent-minded with the feedback during classes.	Strongly Disagree (1) to Strongly Agree (5)
6	I could easily understand the feedback	Strongly Disagree (1) to Strongly Agree (5)
7	I feel I could learn better in the lectures with the feedback	Strongly Disagree (1) to Strongly Agree (5)
8	I would like to see this kind of visual feedback in the future when attending online classes.	Strongly Disagree (1) to Strongly Agree (5)

1 Supporting Information Text

2 Video Transcripts

3 **Video 1.** AI is already applied in everyday life, used in social media face filters, facial recognition, content recommendation, and
4 language translating technology. For example, AI is used for social media face filters by identifying key points on a person's
5 face, or face landmarks, such as the eyes, nose, and mouth. Once these points have been identified, the contours of the face can
6 be pinpointed, and the filter can be applied onto the user's face. AI also allows computers to solve problems through making
7 predictions and decisions. Global problems that can be solved with the assistance of AI include diagnosing diseases, protecting
8 endangered species, and providing support to places hit by natural disasters. The beginning of AI research dates back to the
9 1950s. Alan Turing, a notable computer scientist known for the breaking of the Enigma code in World War II, published a
10 paper suggesting the testing of a "thinking" machine that could hold a conversation with a human as if it were just another
11 person. Turing's papers and their discussion sparked the concept of artificial intelligence. While artificial intelligence was
12 an exciting, futuristic concept in the 1950s, by the 1970s, AI research funding was cut due to many reports stating a lack
13 of progress and success. The late 1970s became known as the first AI winter. The first AI winter ended in the early 1980s
14 after several new developments. One of these developments was the US increasing funding for AI again. Japan had started a
15 new fifth generation computer project and the US aimed to compete to become a world leader in computer science. Another
16 development was the introduction of "Expert Systems" which was quickly utilized by large global corporations. The 1980s
17 marked the end of the first AI winter with the promising introduction of Expert Systems. The Expert system operated within
18 a clearly defined area of knowledge and followed strict logic rules. Even with such restrictions, these programs proved useful to
19 corporations. However, despite the success of the Expert system, computer science experienced a second AI winter starting in
20 1987. Expert system computers were seen as slow, bulky, and not user friendly, and were becoming too expensive and difficult
21 to update compared to the new desktop computer. As consumers no longer needed to buy an expensive machine specialized for
22 running expert systems, this led to the collapse of the market for specialized AI hardware in 1987. Thus, an entire industry

worth half a billion dollars was replaced in a single year. Research funding was once again cut. It wasn't until 1997 when IBM's Deep Blue defeated chess champion Garry Kasparov that the general public realized the power of AI. With the breakthrough in hardware equipment. The computer's processing power and storage capacity grew exponentially, enabling companies to store and process large amounts of data. Various companies and government agencies have successfully applied AI on a larger scale in different applications. This leads to a new round of growth in AI. The 2010s saw the rise of digital virtual assistants. Assistants like Alexa, Siri, and Google utilize natural language processing to act as a convenient source of information for users, doing many of the tasks a human assistant would. Virtual assistants like these are the future of AI. They are controlling autonomous cars and taking a physical form as a robot. Artificial intelligence has come a long way since Alan Turing's idea in a hypothetical paper, and it will continue into the future driving cars, walking around in everyday life, and more. Artificial Intelligence Definitions: But still, what is AI? AI usually does not have one explicit definition. However, it can be defined as a sector of computer science that allows computers to solve problems through making predictions and decisions. We have learned that researchers have developed many ways to enable computers to think and solve problems, and some are less successful while some are still striving these days. The sector of AI where AI currently is adopted in products is called machine learning, or ML. Machine learning teaches computers how to make problem-solving predictions and decisions by teaching them to learn from collected data. This is the state-of-the-art solution to achieving AI and will be our focus in the series of lectures. ML methods emphasize data, and we will see how data helps computers to learn about the world. ML will be our focus in the series of talks. As seen in the diagram, artificial intelligence is a branch of computer science, and machine learning is a branch of artificial intelligence. As our focus will be on machine learning, now let us have a look at this field in AI from a high level. Usually, machine learning problems and its related solutions can be divided into three categories. We will have a look at each category.

Video 2. The most straightforward way of learning is called supervised learning, also known as supervised machine learning. Supervised learning deals with the problems that are trying to find a relationship that guides an input to an output based on given examples of input-output pairs. Each input-output pair can be understood as a question-answer pair when you learn math. By showing the computer with many example pairs, we expect it can understand how to generate an output given some new input. Let's consider this example. In a fruit packaging factory, we would like to find an automated way to separate the different kinds of fruit, Apple, blueberry, and banana, into three different boxes for shipment. A digital system can take pictures of each fruit traveling down the conveyor belt and must use the image input to identify the fruit with its name for its proper shipment box destination. In this example, the input to the system is pictures of fruits, and the output expected from the system is the category of the fruit. The relationship we would like to find is how to identify the fruit category given the pictures taken. The input-output pairs we would provide to the system would then be the image of a fruit and a label saying what fruit is inside the image. Each time the picture-fruit-name pair is fed to the system, we expect the model to discover the underlying relationship between picture and fruits, and gradually the system improves the model's performance. This procedure of adjusting our system based on data is called learning or training. After training, we will apply the system to the conveyor belt, with new pictures of apples, blueberries, and bananas, to see if the model can provide the correct output. This is called prediction. So why is supervised learning feasible in this example? Why can images of apples be identified as one thing and images of blueberries can be identified separately? This is because the apples are similar to each other just like the blueberries are similar to each other. And apples and blueberries are different in some aspects. In fact, these similar or different aspects are what a system has to pay attention to. Such aspects are called features, which are characteristics of objects, they are selected to make all the inputs widely different from each other. In this instance, useful features are colors and textures, and unreliable features are "round shape" and "not sparkly" because both apples and blueberries are round and not sparkly. Both apples and blueberries share these characteristics and they would not be helpful in distinguishing the two. We may ask, what should good features look like? The samples have to be quite different in this feature. Color is a good feature because apples and blueberries are different in colors. In addition, the feature is supposed to be highly correlated with the target sample. A feature like the age of the fruit is irrelevant because fruits do not have age! Labels are the thing we are trying to predict based on an input. Using the fruit example, the label would be the type of fruit predicted based on the image input: apples, blueberries, or bananas. The system now notes useful features in the input image. If the fruit in the image is red, the algorithm would be able to label the inputted image as an apple. But computers do not understand quantitative descriptions such as colors or shapes. Computers only understand numbers. So we have to translate the previous ideas into mathematical descriptions. For example, consider how we would describe an apple. We would like first to describe its color. Then the first number in the vector will represent colors. We may use 0 to represent red, and 1 for navy blue, and 2 for green, etc. In the second number, we may describe the shape of the fruit. We can use 0 for a round shape, 1 for an oval shape, and 2 for a curved shape. This ordered collection of numbers is called a vector, and it describes the features we would like the system to take. The input data is usually represented by x and we use a subscript to distinguish each different feature. here, the x_1 would be color and x_2 would be the shape. We then translate the labels. In this example, we can use 0 for apples, 1 for blueberries, and 2 for bananas. But the output can also be real numbers. For example, we can ask the model to generate the weight of the apple given the image of the apple, so we could further understand the quality of apples and sell them at different prices. Now we can use mathematical equations to describe this relationship. We can start with a very simple math equation: $y = 1 + 0.5 x_1 + 0.5 x_2$. We use the input to calculate the output generated by our system, that would be 1, 1.5, 2, and 2. We notice the model does not provide the correct answer now but it is fine. We can change the equation to make the output from the model to be closer to the label. This is the goal of supervised learning : to minimize the difference between the generated output from the model with the label.

Video 3. But in many cases, often large amounts of data are available, but no labels are present. For example, medical devices such as a CAT scanner, MRI scanner, or an EKG, produce streams of data but these are entirely unlabeled. They do not contain the information of which part is indicating what disease. In these cases obtaining labeled data is difficult, costly, or impossible, and so supervised learning methods are not possible. This gives rise to unsupervised learning. Unsupervised learning describes a family of algorithms that learns patterns from unlabelled data. This is in contrast to supervised learning techniques where a model is given a training set of inputs and a set of observations and must learn a function from the inputs to the observations. In unsupervised learning, only the inputs are available, and a model must look for interesting patterns in the data. The machine is forced to build a compact internal representation of its given input data and use such representation to complete the requested tasks. The two main subcategories of unsupervised learning we will be discussing today are anomaly detection and clustering. One widely adopted use case for unsupervised learning is anomaly detection. It is generally understood to be the identification of rare items, events, or observations that deviate significantly from the majority of the data and do not conform to normal behavior. Looking at the graph to the left, most of the data points follow a nonlinear, increasing line. The one exception is the point at (30,6000). This point deviates significantly from the pattern of the rest of the data, and would be considered an anomaly. One example of the use of anomaly detection is with fraud. With the common use of credit cards, financial fraud has become a major problem. Unauthorized or fraudulent transactions can sometimes be recognized as a break or an anomaly from the user's normal pattern of usages, such as large volume transactions, or rapid buying sprees. Credit card transaction data can be fed into an anomaly detection algorithm in the form of a vector, which is a collection of numbers, such as transaction amount, transaction time of day, transaction location, and time since the previous transaction. The outliers can then be flagged to the bank as potentially fraudulent. In these cases, the bank can either unilaterally block the card or request the user to authenticate the transaction in another way. Anomaly detection algorithms, however, may not always be accurate. Sometimes, a normal datapoint may be flagged as an anomaly, a false positive, or an anomalous data point is mistaken as normal and overlooked, a false negative. For example, when looking at cancer diagnoses, a false positive would be if cancer was detected in a patient that does not have cancer. A false negative would be if cancer was not detected in a patient that does have cancer. These inaccuracies can have consequences. In this situation, a false positive would cause unnecessary stress and worry for the healthy patient. A false negative would be far more damaging, as a sick cancer patient may not seek treatment, mistaking themselves as healthy. A false positive, however, is not always the more damaging error. In addition, the severity of errors is not always the same: a false negative COVID test is less serious than a false negative cancer diagnosis. It is important to assess the situation in which an anomaly detection algorithm is being applied to. The algorithm can then be adjusted to favor a false negative over a false positive, or vice versa. The end goal of an anomaly detection algorithm is to be as accurate as possible, while also minimizing the risk/cost of false results. Clustering is a machine learning problem that aims to group data that have similar properties or features. Recall the previous medical imaging example, often large amounts of data are available, but no labels for the data are given. Doctors have a large amount of imaging data, but they do not know whether the obtained images are associated with what kind of disease. In 2019, a team of researchers in the UAE, Egypt, and Australia conducted a meta-study of clustering algorithms on Alzheimer's disease data, and reported that it was possible to identify subgroups which corresponded to the stage of the disease's progression. A number of clustering methods have been applied to datasets and they found that all of the clustering algorithms investigated brought a new level of insight into the various subtypes of Alzheimer's patients. The clustering techniques allow medical practitioners to identify patterns across patients which would otherwise be difficult to find by eye. The blue dots on the left diagram represent all of the input feature vectors. The separate green, blue, and red dots on the right diagram represent the same inputs, but now they have been grouped into three separate clusters. You want the clusters to be distinct from one another, while also having individual data points within a cluster as similar as possible. In other words, you want to maximize the distance between clusters, and minimize the distance between each datapoint and the average point of the cluster to which it is assigned. This is the goal of clustering problems.

Video 4. Computer vision is the field of computer science that focuses on creating digital systems that can process, analyze, and make sense of visual data (images or videos) in the same way that humans do. The concept of computer vision is based on teaching computers to process an image at a pixel level and understand it. Technically, machines attempt to retrieve visual information, handle it, and interpret results through special software algorithms. Computer vision trains machines to perform a variety of functions, such as how to tell objects apart, how far away they are, whether they are moving, and whether there is something wrong with an image. But it has to do it with cameras, data, and algorithms and way faster than humans do. For example, a system trained to inspect products or watch a production asset can analyze thousands of products in less than a minute, noticing imperceptible defects or issues for humans. Computer vision is used in industries ranging from [energy and utilities] to [manufacturing and automotive]. The market is continuing to grow. It is expected to reach around 50 billion dollars by this year. Scientists and engineers have been trying to develop ways for machines to see and understand visual data for about 60 years. Experimentation began in 1959 when neurophysiologists showed a cat an array of images, attempting to correlate a response in its brain. They discovered that it responded first to hard edges or lines, and scientifically, this meant that image processing starts with simple shapes like straight edges. At about the same time, the first computer image scanning technology was developed, enabling computers to digitize and acquire images. In 1959, scientists developed equipment that allowed transforming images into grids of numbers—in the language machines could understand. And it's because of their work that we now can process digital images in various ways. One of the first digitally scanned photos was the image of the scientist's infant son. It was just a small 5cm by 5cm photo captured as 30k pixels. In 1989, a young French scientist Yann LeCun applied a back-propagation style learning algorithm. After working on the project for a few years, LeCun released LeNet-5—the first modern convolutional network that introduced some of the essential ingredients we still use in CNNs today. In 2006, the

Pascal VOC project was launched. It provided a standardized dataset for object classification. The founders also ran an annual competition that allowed evaluating the performance of different methods for object class recognition in 2012, a team from the University of Toronto entered a convolutional neural network model into an image classification competition and that changed everything. The model, called AlexNet, similar to LeCun's LeNet-5, achieved an error rate of 16.4%, which overperformed all other methods at that time. This was a breakthrough moment for CNNs. However, convolutional neural networks have been around since the 1980s. So why did it take so long for them to become popular? Thanks to the development of Graphics Processing Units, we can parallelize the computation and significantly reduce the time needed for computation. Also, it was around 2012 when large, labeled, high-dimensional visual datasets became available to the computer vision research community. These two factors helped CNNs to boost. Here are a few common tasks that computer vision systems can be used. The first common task is called Object classification. The system parses visual content and classifies the object on a photo/video to the defined category. For example, the system can find a cat when presented with the first image. The second figure shows the task called Object localization. This requires a bit more than the classification. Not only the system needs to generate the class label, it should also identify a bounding box that covers the detected object. Then the third figure shows the Object Detection task. This becomes more difficult as it requires the system to be able to process multiple objects in the image at the same time. The fourth task is called Object tracking. It can be considered object detection over video. The system takes an initial set of multiple objects detected and then tracks each object as they move around frames in a video while maintaining the identity of each object. Before we discuss how computers process images, we need to understand how images are stored in computers. A digital image stored in computers is a collection of pixels. A pixel or picture element is the smallest addressable element in the image. For each pixel, it is assigned a color when we are dealing with color images. Here we show an example of a gray-scale image. In computers, the image is stored in a format similar to a table. Each row and column specifies a pixel, and each pixel is associated with a value representing the level of brightness. For gray-scale images, the value to each pixel is just the brightness at each pixel. But for colorful images, each pixel contains three numbers representing the red, green, and blue values of each color. For simplicity, we will mainly discuss grey-scale images in the following slides.

Video 5. In 2019, NHS Foundation Trust provided data on 1.6 million patients to Alphabet's DeepMind, without taking any permission from the patients to share their private data. Keeping in view the privacy concerns of patients, Google canceled the plan to broadcast chest X-ray scans on account of concerns of patients as they were carrying personally identifiable information. The second example is a data set containing above 10 million images of 100,000 celebrities created by Microsoft called MS Celeb. This has been removed since celebrities expressed their concerns that they were not even aware of being included. Besides this unconscious revealing of personal data, there is a type of personal data that we are uploading on social media by ourselves. For instance, we visit a restaurant or a store and take several images of the food or product we like and post them online. Most of this data is transferred to cloud computers that have considerably enhanced the probability of tracking this personal information. With these examples in mind, we can have a formal discussion about privacy in the context of AI. Imagine the following scenario. You know that you often forget where you parked your car, so you use an app you downloaded called "Find my Car." The app takes a photo of your car and then geocodes the photo, enabling you to easily find the right location when you come to retrieve your car. However, this example illustrates a variety of privacy concerns. The first concern is called Data persistence, where data exists longer than the human subjects that created it. The images of your car photo may store for years and even you have forgotten its existence. The second concern is called Data repurposing, which means data are being used beyond their originally imagined purpose. For example, in a decade's time, parking habits may be part of the data used by grocery retailers to analyze an individual's preference for shopping. Finally, the concern called Data spillovers is associated with the case when data are collected on people who are not the target of data collection. The photo may record other people and they may be identifiable through facial recognition, or incidentally captured cars may be identifiable through license plate databases. In addition to these general concerns over data, in the context of AI, more concerns have grown over its ability to make an unbelievable precise prediction of someone. ML algorithms can easily infer sensitive information from insensitive data. For example, Google has been very successful in gathering private data. The main factor behind its success is that while searching on the internet to get some information, people most of the time can't hide their interests. Even if someone tries to hide sensitive private issues, he can't search about his interest unless the terms are not entered in the search bar. Nonetheless, our most intimate interests that everyone wants to keep private are not private anymore and are collected online. Political views, ethnic identity, and overall health can also be predicted from activity logs and several other metrics. As concerns grow, many Privacy-preserving techniques that can largely conceal the identity of persons or groups are now a standard staple in AI systems. These techniques must balance the need for privacy and the usability of the system. We first discuss the technology that is called Differential privacy. Before trying to understand this method, let us consider this example. Assume your parent is trying to get the score of your final exam this semester, but your teacher is trying to keep the privacy of students. The only information your parent can receive is the average, median and some other general information about the score. In this way, your parent learns some information about the final, but your privacy is preserved. This is the idea used by Differential privacy: the information requester will not be able to be 100% sure to infer any specific individual. But please note this protection is offered by the number of students in your class. If the class contains a hundred people, each person's score contributes just 1%. However, if the class contains scores from only a single person, that person's individual data contributes to 100% of the class data. The key insight of differential privacy is that as the query is made on the data of fewer and fewer people, more noise needs to be added to the result to produce the same amount of privacy. In this example, if your parent asks the average score of students with your last name, to protect your privacy, your teacher must add some noise in the result. otherwise, your scores will very likely be identified. Though this partially solves the problem. Individuals are not

Table S3. Post-Test Questions in Video 1

ID	Question	Choice
1	Which of the following could become the applications of AI: (1). Social media face filters (2). Facial recognition (3). Content recommendation (4). Language translation (5). Emotion recognition (6). English translation	A: (1)(2)(3)(4), B: (1)(2)(3)(4)(5), C: (1)(2)(3)(4)(6), D: (1)(2)(3)(4)(5)(6)
2	AI is used in social media face filters by identifying [] on the person's face.	A: eyes, B: key points, C: key lines, D: key shapes
3	In the 1950s, [] suggests the possibility of a "thinking machine".	A: Yann LeCun, B: Geoffrey Hinton, C: Alan Turing, D: Yoshua Bengio
4	How many AI winters from 1950s to 2010s (60 years)	A: 2, B: 3, C: 4, D: 5
5	From 1980s to 1990s, [] , followed by the second AI winter, led to success.	A: thinking machines, B: expert systems, C: machine learning, D: digital virtual assistants
6	From 1990s to 2000s, [] got further development due to the advances of computing resources and hardware.	A: expert systems, B: machine learning, C: digital virtual assistants, D: chat boxes
7	Which of the following are digital virtual assistants? (1). Alexa, (2). Siri, (3). Google Assistant	A: (1)(2)(3), B: (1)(3), C: (2)(3), D: (1)(2)
8	Which of the following is CORRECT?	A: Difficulties in AI progress in the 1970s result in the first AI winter, B: AI develops quickly even in the first AI winter, thanks to the development of powerful computing hardware resources, C: Machine learning systems operate within a clearly defined area of knowledge and follow strict logic rules., D: The digital virtual assistants today have already had the same ability like humans
9	AI allows computers to solve problems by making [] and decisions.	A: assumptions, B: predictions, C: observations, D: tests
10	Machine learning could teach computers to solve problems via learning from collections of [].	A: dataset, B: problems, C: other computers, D: user behaviors
11	Which of the following is correct to describe the relationship among computer science (CS), artificial intelligence (AI), and machine learning (ML)? (X < Y means X is included by Y)	A: AI < ML < CS, B: ML < AI < CS, C: CS < ML < AI, D: ML = AI < CS
12	Which of the following represents three ways of machine learning?	A: Representation Learning, Meta Learning, Reinforcement Learning, B: Supervised Learning, Unsupervised Learning, Representation Learning, C: Reinforcement Learning, Supervised Learning, Unsupervised Learning, D: Feature Engineering, Contrastive learning, Representation Learning

205 identified, but the data we uploaded to the server may suffer from Data persistence and spillover. A solution to this problem is
 206 to encrypt the data. Encryption is the process of encoding information. This process converts the original representation of the
 207 information, known as plaintext, into an alternative form known as ciphertext. Ideally, only authorized parties can decipher a
 208 ciphertext back to plaintext and access the original information. Encryption does not itself prevent interference but denies
 209 the intelligible content to a would-be interceptor. However, this stops us from utilizing any service provided by the system.
 210 The privacy is preserved, but we again receive nothing helpful. Scientists then try to find a way to encrypt the data, yet still
 211 allow computation. This is termed homomorphic encryption. It permits users to perform computations on its encrypted data
 212 without first decrypting it. These computations are left in an encrypted form. When decrypted, the result is identical to that
 213 produced had the operations been performed on the unencrypted data. Homomorphic refers to a special correspondence, and
 214 in this context, it means the calculation on encrypted data is the same when it is being decrypted. For example, let's consider
 215 the easiest encryption called the Caesar cipher, which replaces a letter with another letter with a fixed number of positions
 216 down the alphabet. For example, with a shift of 3, "A" will be replaced by "D" and "B" will be replaced by "E". Letter "Z"
 217 will be replaced by "C." Consider our calculation is to concatenate two words. The encrypted words are uploaded to some
 218 remote computer for this computation, and the result is transmitted back. We then decipher the result and find this is just the
 219 result if we conduct the same operation with the plain text, but the good thing is the remote computer can not tell what is
 220 being transmitted. In this way, both privacy and utility are preserved.

Table S4. Post-Test Questions in Video 2

ID	Question	Choice
1	Supervised learning is a task of learning one [] that maps one [] to one [] based on example pairs.	A: output, function, input, B: function, input, output, C: output, input, function, D: input, function, output
2	Which of the following is CORRECT?	A: Labels are necessary in supervised learning., B: Labels could be optional in supervised learning., C: Whether labels are necessary or optional depends on the tasks in supervised learning., D: No information about labels is provided in the lecture.
3	Assume that we will predict the fruit name from its image using supervised learning. Which of the following is WRONG?	A: The image serves as the input., B: The fruit name serves as the label., C: Supervised learning models could not predict input image which is unseen before., D: Supervised learning models will learn the relationship between images and fruit names..
4	Which of the following tasks is NOT a supervised learning problem?	A: A system learns to distinguish images with known categories such as cars or bicycles., B: Detect abnormal behaviors of motors from a large dataset of their running speed., C: Predict whether a piece of audio is spoken by a specific user from a labeled dataset., D: Recognize human activities according to human motion data and activity list..
5	Assume that we will predict the category of one fruit from its characteristics. Then the category means [] and its characteristics means [].	A: label, features, B: input, output, C: output, function, D: label, function.
6	Assume that we will predict the fruit name from its image. Then what kind of features below could be extracted from images for prediction? (1). Color (2). Shape (3). Texture (4). Brightness (5). Weight (6). Smell (7). Taste	A: (1)(2)(3)(4), B: (1)(2)(3)(7), C: (1)(2)(3)(4)(5)(6), D: (1)(2)(3)(4)(5)(6)(7).
7	In the fruit recognition task, the reason why separate identification is feasible is that fruits of the same category are [] to each other.	A: the same, B: similar, C: distinguishable, D: totally different.
8	Which of the following is WRONG?	A: In a task to distinguish between apples and oranges, color will be a useful feature because two kinds of fruits have different colors., B: In a task to distinguish between apples and pears, shape may not be a reliable feature because two kinds of fruits probably have similar shape., C: In a task to distinguish between apples and pears, useful features should be similar across all apples., D: In a task to distinguish between apples and pears, useful features should be similar between apples and pears..
9	The goal of supervised learning models is to [] the difference between the generated output and the label.	A: maximize, B: minimize, C: iterate, D: examine.
10	In a fruit recognition task, if the input feature is (1,2) and our supervised learning model is $y = 1 + x_1 + x_2$, then the output is [].	A: 4, B: 3, C: 2, D: 1.
11	In a fruit recognition task, assume that the useful features are color and shape. Which of the following is WRONG?	A: A fruit with different colors will be represented by different values., B: A fruit with different shapes will be represented by different values., C: Supervised learning models will learn the relationship between features and labels for fruit recognition., D: The goal of supervised learning models is to learn the difference between features and labels for fruit recognition..

Table S5. Post-Test Questions in Video 3

ID	Question	Choice
1	Unsupervised learning is to analyze and group [] datasets by learning the patterns among data.	A: labeled, B: unlabeled, C: clustered, D: balanced.
2	Which of the following tasks is an unsupervised learning problem?	A: A computer calculates the digits of π , a mathematical constant that is close to 3.14. B: A system learns to distinguish images with known categories such as cars or bicycles. C: A security system can detect intrusive data flow that differs from the normal pattern. D: The chess software competes with the player and gradually plays better.
3	Which of the following is WRONG?	A: Supervised learning requires a teacher to tell the system the correct answer., B: Unsupervised learning could directly learn by examples, with the correct answer., C: Training a model to recognize fruit categories from fruit images with labels is a supervised learning problem., D: Training a model to distinguish fruit groups from fruit images without labels is an unsupervised learning problem..
4	Anomaly detection means the identification of [] items, events, or observations that deviate significantly from the [] of the data and do not conform to [] behavior.	A: rare, minority, normal, B: rare, majority, normal, C: most, minority, traditional, D: most, majority, traditional.
5	The goal of anomaly detection is to [] the risk/cost associated with [] positive and negative.	A: minimize, true, B: minimize, false, C: maximize, true, D: maximize, false.
6	An anomaly detection system has labeled the following samples. Which sample is a false positive? Predicted: Positive (Sample 1), Positive (Sample 2), Negative (Sample 3), Negative (Sample 4), Negative (Sample 5). Actual: Negative (Sample 1), Positive (Sample 2), Negative (Sample 3), Negative (Sample 4), Positive (Sample 5)	A: Sample 1, B: Sample 2, C: Sample 3, D: Sample 5.
7	Which of the following is NOT an example of anomaly detection?	A: Estimate whether the machine is running normally from its working logs., B: Distinguish whether the stock volatility is abnormal according to the financial curve., C: Estimate whether one boy is happy from his face image., D: Distinguish between normal and abnormal working status of a motor in its working data..
8	Which of the following is WRONG?	A: Supervised learning aims to minimize the difference between generated output and labels., B: Unsupervised learning aims to find some distinct groups of data without labels., C: Anomaly detection is a kind of unsupervised learning model., D: Anomaly detection aims to detect traditional behaviors from a dataset..
9	Clustering is a machine learning problem that aims to group data that have [].	A: similar, features, B: similar, categories, C: different, features, D: different, categories.
10	Which of the following is NOT an example of clustering?	A: Find potential diseases from a large amount of medical image dataset without labels., B: Cluster a dataset of human activities and get different groups., C: Divide a human face dataset into several groups containing specific properties., D: Predict user emotion from a dataset with user face and emotion status..
11	Which of the following is WRONG?	A: Clustering does not have a specific label to determine whether the prediction is correct or not., B: Clustering is a kind of machine learning model., C: Clustering could find potential correlations/properties in the dataset automatically., D: Clustering is another supervised learning model compared with anomaly detection..

Table S6. Post-Test Questions in Video 4

ID	Question	Choice
1	Computer vision is the field of computer science that focuses on creating digital systems that can process, analyze, and make sense of [] in the same way that humans do.	A: images, B: videos, C: audio, D: images or videos.
2	Which of the following may utilize computer vision techniques? (1). Use a camera to check potential issues on the surface of products (2). Estimate the freshness of apples from pictures (3). Estimate whether a car is speeding via a camera (4). Determine whether a piece of audio is spoken by a specific person	A: (1)(2)(3), B: (1)(2)(4), C: (2)(3)(4), D: (1)(2)(3)(4).
3	The first computer image scanning technique was developed in the [].	A: 1940s, B: 1960s, C: 1980s, D: 1990s.
4	[] attempted perceptual grouping in 1997.	A: Yann LeCun, B: Geoffrey Hinton, C: Alan Turing, D: Yoshua Bengio.
5	One breakthrough in computer vision happened at the University of Toronto in 2012, which achieved an error rate of [] in image classification.	A: 6.4%, B: 10.4%, C: 12.4%, D: 16.4%.
6	In 2006, [] project was launched to provide researchers with a standard dataset for image classification task competition.	A: MNIST, B: Pascal VOC, C: CIFAR, D: Image scanning.
7	Which of the following is WRONG?	A: In the 1960s, neurophysiologists first showed a cat an array of images, attempting to correlate a response in its brain., B: The first computer image scanning technique was used to show a photo of an infant., C: A complete computer vision system not only needs a computer to process image data, but also requires a sensing device to capture input signals (data)., D: If a computer vision system could achieve higher image classification accuracy than humans, then it means its intelligence is superior to human intelligence..
8	Which of the following are computer vision common tasks? (1). Image Classification (2). Image Tracking (3). Object Localization (4). Object Detection (5). Object Tracking	A: (1)(2)(4)(5), B: (1)(2)(3)(5), C: (1)(3)(4)(5), D: (1)(2)(3)(4)(5)(6).
9	Which of the following is closer to the image classification task?	A: Judge whether an input image is an apple, B: Find different objects from an input image, C: Find one specific object from an input image, D: Estimate the number of dogs in an input image.
10	Tom would like to track the movements of a dog in a picture. Which of the following techniques may be used? (1). Image Classification (2). Image Tracking (3). Object Localization (4). Object Detection (5). Object Tracking	A: (1)(5), B: (4)(5), C: (1)(4)(5), D: (3)(4)(5).
11	Tom would like to count the number of apples and dogs in a picture. Which of the following techniques are necessary to be used? (1). Image Classification (2). Image Tracking (3). Object Localization (4). Object Detection (5). Object Tracking	A: (4), B: (4)(5), C: (3)(4), D: (1)(4).
12	Which of the following is WRONG?	A: An image is processed as a matrix in computer vision., B: The minimum unit is one pixel in an image in computer vision., C: The matrix of an image could represent color, transparency, brightness, etc., D: Each pixel of an image represents one number which is related to image color..

Table S7. Post-Test Questions in Video 5

ID	Question	Choice
1	[] is one of the major sources for high-tech companies to collect personal information.	A: posters, B: physical examinations, C: social media (SNS), D: flyers.
2	10 million celebrity images shared by [] were removed on request.	A: DeepMind, B: Microsoft, C: Meta, D: Apple.
3	Which of the following is CORRECT?	A: Researchers should collect user data with consent, even if the data collection process will not hurt users., B: High-tech companies could gather personal health information without consent, on condition that they could make sure the information will not be released to the public., C: We could directly use images shared in the social media for non-commercial purposes such as research purposes., D: We could only collect very limited personal health information if we do not have consent..
4	Bob uploads a photo of himself in a crowd of people on an SNS. This image is then uploaded and stored to a remote server, but Bob is concerned this might violate the background peoples' privacy. Which of the following options best describes Bob's concern?	A: Data persistence, B: Data repurposing, C: Data spillovers, D: Sensitive information inferred from insensitive data..
5	Tom wants to find medicine suggestions for his fever by searching on the Internet. Which kind of privacy issue may occur in this scenario?	A: Data persistence, B: Data repurposing, C: Data spillovers, D: Sensitive information inferred from insensitive data..
6	Jack uploaded his personal data to a website for registration, which however was stored in the website for several years. Which kind of privacy issue may occur in this scenario?	A: Data persistence, B: Data repurposing, C: Data spillovers, D: Sensitive information inferred from insensitive data.
7	Tom signed a consent form in a study that aimed to use his data to train a model for human activity recognition. However, his data was then further utilized to investigate other medical research topics without notifying Tom. Which kind of privacy issue may occur in this scenario?	A: Data persistence, B: Data repurposing, C: Data spillovers, D: Sensitive information inferred from insensitive data..
8	In homomorphic encryption, the calculation could be operated directly over the encrypted data, and the result is [] after decryption.	A: the same, B: similar, C: different in most cases, D: totally different.
9	Tom's father wants to know about Tom's exam score. Which of the following protects Tom's privacy using a differential privacy method? (1). The teacher only tells Tom's father the average score of all students. (2). The teacher tells Tom's father the score range is from 80 to 90 for all students with the first name "Tom". (3). The teacher tells Tom's father the encrypted exam score of Tom. (4). The teacher does not tell Tom's father the exam score of Tom.	A: (1)(2), B: (1)(3), C: (1)(2)(3), D: (1)(2)(3)(4).
10	We now use Caesar Cipher for encryption, i.e. replace the letter with a shifted letter. Specifically, A, B, C, . . . , Y, Z will be replaced by D, E, F, . . . , B, C. In this case, if the raw data is "World" on our device, which kind of encrypted data will be received in the remote computer?	A: Koor, B: Zruog, C: Khoog, D: World.
11	Assume we encrypt a word by multiplying it by 2. For example, the number 123 is encrypted as 246. Which of the following math operations holds the homomorphic property using this encryption?	A: Addition, B: Multiplication, C: Both operations, D: None of the above.
12	Which of the following is WRONG?	A: Differential privacy protects user privacy via adding some noise into the results so that others could not obtain certain information., B: Encryption protects user privacy via encoding private data in specific manners during transmission., C: Differential privacy could be used for remote data transmission as well, just like encryption and decryption., D: Both differential privacy and encryption could protect user privacy..