Big Idea #1: Perception	*Computers perceive the world using sensors.*	**Perception is the extraction of meaning from sensory information using knowledge.**	**The transformation from signal to meaning takes place in stages, with increasingly abstract features and higher level knowledge applied at each stage.**	**LO** = Learning Objective: what students should be able to do. **EU** = Enduring Understanding: what students should know.
Concept	**K-2**	**3-5**	**6-8**	**9-12**
Sensing (Living Things) **1-A-i**	**LO:** Identify human senses and sensory organs. **EU:** People experience the world through sight, hearing, touch, taste, and smell.	**LO:** Compare human and animal perception. **EU:** Some animals experience the world differently than people do. **Unpacked:** Bats and dolphins use sonar. Bees can see ultraviolet. Rats are have no color vision; dogs are red-green colorblind. Dogs and rats can hear higher frequences than humans.	**LO:** Give examples of how humans combine information from muliple modalities. **EU:** People can exploit correlations between senses, such as sight and sound, to make sense of ambiguous signals. **Unpacked:** In a noisy environment, speech is more understandable when the speaker's mouth is visible. People learn the sounds associated with various actions (such as dropping an object) and can recognize when the sound doesn't match their expectation.	N/A -- for AI purposes, this topic has already been adequately addressed in the lower grade bands. *Other courses, such as biology or an elective on sensory psychology, could go into more detail about topics such as taste, smell, proprioception, and vestibular organs.* *Possible enrichment material: look at optical illusions (Muller-Lyer illusion, Kanizsa triangle) and ask which ones are computer vision systems also subject to.*
Sensing (Computer Sensors) **1-A-ii**	**LO:** Locate and identify sensors (camera, microphone) on computers, phones, robots, and other devices. **EU:** Computers "see" through video cameras and "hear" through microphones.	**LO:** Illustrate how computer sensing differs from human sensing. **EU:** Most computers have no sense of taste, smell, or touch, but they can sense some things that humans can't, such as infrared emissions, extremely low or high frequency sounds, or magnetism.	**LO:** Give examples of how intelligent agents combine information from multiple sensors. **EU:** Self driving cars combine computer vision with radar or lidar imaging. GPS measurement, and accelerometer data to form a detailed representation of the environment and their motion through it.	**LO:** Describe the limitations and advantages of various types of computer sensors. **EU:** Sensors are devices that measure physical phenomena such as light, sound, temperature, or pressure. **Unpacked:** Cameras have limited resolution, dynamic range, and spectral sensitivity. Microphones have limited sensitivity and frequency response. Signals may be degraded by noise, such as a microphone in a noisy environment. Some sensors can detect things that people cannot, such as infrared or ultraviolet imagery, or ultrasonic sounds.
Sensing (Digital Encoding) **1-A-iii**	N/A	**LO:** Explain how images are represented digitally in a computer. **EU:** Images are encoded as 2D arrays of pixels, where each pixel is a number indicating the brightness of that piece of the image, or an RGB value indicating the brightness of the red, green, and blue components of that piece.	**LO:** Explain how sounds are represented digitally in a computer. **EU:** Sounds are digitally encoded by sampling the waveform at discrete points (typically several thousand samples per second), yielding a series of numbers.	**LO:** Explain how radar, lidar, GPS, and accelerometer data are represented. **EU:** Radar and lidar do depth imaging: each pixel is a depth value. GPS triangulates position using satellite signals and gives a location as longitude and lattitude. Accelerometers measure accleration in 3 orthogonal dimensions. **Unpacked:** Radar and lidar measure distance as the time for a reflected signal to return to the transceiver. GPS determines position by triangulating precisely timed signals from three or more satellites. Accelerometers use orthogonally oriented strain gauges to measure acceleration in three dimensions.

Big Idea #1: Perception	*Computers perceive the world using sensors.*	**Perception is the extraction of meaning from sensory information using knowledge.**	**The transformation from signal to meaning takes place in stages, with increasingly abstract features and higher level knowledge applied at each stage.**	**LO** = Learning Objective: what students should be able to do. **EU** = Enduring Understanding: what students should know.
Concept	**K-2**	**3-5**	**6-8**	**9-12**
Processing (Sensing vs. Perception) **1-B-i**	**LO:** Give examples of intelligent vs. non-intelligent machines and discuss what makes a machine intelligent. **EU:** Many machines use sensors, but not all use them intelligently. Non-intelligent machines are limited to simple sensing. Intelligent machines demonstrate perception. **Unpacked:** Cameras and phones can record and play back images and sounds, but extracting meaning from these signals requires a computer with artificial intelligence.	**LO:** Use a software tool such as a speech transcription or visual object recognition demo to demonstrate machine perception, and explain why this is perception rather than mere sensing. **EU:** Perception is the extraction of meaning from sensory signals. **Unpacked:** speech recognition and face detection are examples of perception. An automatic door activated by a pressure pad or ultrasonic sensor does not exhibit perception because it is just reacting to the raw signal rather than using kowledge to extract meaning from the signal.	**LO:** Give examples of diferent types of computer perception that can extract meaning from sensory signals. **EU:** There are many specialized algorithms for perceptual tasks, such as face detection, facial expression recognition, object recognition, obstacle detection, speech recognition, vocal stress measurement, music recognition, etc.	**LO:** Explain perception algorithms and how they are used in real-world applications. **EU:** Many devices and services rely on specialized perception algorithms, e.g., license plate readers, zip code readers, face-based phone unlocking, tagging people in Facebook posts, object identification (e.g., Google Lens), or voice-based customer service.
Processing (Feature Extraction) **1-B-ii**	**LO:** Give examples of features one would look for if one wanted to recognize a certain class of objects (e.g., cats) in an image. **EU:** The visual features of an object include its subparts, textures, and colors. **Unpacked:** to recognize cats one would look for ears, paws, whiskers, and a cat-shaped nose and tail; for textures that look like fur; and for coloration patterns typical of cats.	**LO:** Illustrate how face detection works by extracting facial features. **EU:** Face detectors use special algorithms to look for eyes, noses, mouths, and jaw lines. **Unpacked:** Facial recognition goes one step further and tries to determine whose face has been detected. Recognition is based on quantifiable properties such as distance between the eyes or shape of the jaw line.	**LO:** Illustrate the concept of feature extraction from images by simulating an edge detector. **EU:** Locations and orientations of edges in an image are features that can be detected by looking for specific arrangements of light and dark pixels in a small (local) area.	**LO:** Explain how features are extracted from waveforms and images. **EU:** A speech spectrogram shows the energy present in a waveform in various frequency bands. Formants are auditory features defined as regions of concentrated energy in the spectrogram. Feature extraction from images begins with detecting edges in the image, or intensity gradients at multiple scales. **Unpacked:** Different formant patterns are associated with different speech sounds, i.e., different vowels and consonants.
Processing (Abstraction Pipeline: Language) **1-B-iii**	**LO:** Describe the different sounds that make up one's spoken language, and for every vowel sound, give a word containing that sound. **EU:** In order for a computer to understand speech, it has to be able to recognize the sounds from which words are constructed. **Unpacked:** There are 15 vowel sounds in American English: 5 short, 5 long, and 5 "other". Words for the 5 short vowels are: bid, bed, bad, bog, and bug.	**LO:** Illustrate how sequences of sounds can be recognized as candidate words, even if some sounds are unclear. **EU:** Going from sounds to words is one step in the abstraction pipeline for speech understanding. **Unpacked:** consider the problem of guessing a four letter word given only partial information about the sound in each position, e.g., the first sound is either "f" or "d", and the second sound is either "l" or "n". Knowledge about the constraints between adjacent sounds in a word can be used to narrow down the possibilities. In this case, only "fl" is a valid word-initial sequence in English.	**LO:** Illustrate how sequences of words can be recognized as phrases, even if some of the words are unclear, by looking at how the words fit together. **EU:** Information at higher levels of representation can be used to resolve ambiguities in lower levels of the language abstraction pipeline. **Unpacked:** in a three-word phrase, if the first word might be "seat" or "sea" or "see", the second word might be "the" or "a" or "of", and the third word might be "moody" or "movie", then the most likely phrase is "see the movie" because it's both grammatical and statistically common. Alternatives such as "seat a moody" sound similar, but are neither grammatical nor statistically common.	**LO:** Illustrate the abstraction hierarchy for speech understanding, from waveforms to sentences, showing how knowledge at each level is used to resolve ambiguities in the levels below. **EU:** The spoken language hierarchy is: waveforms -> articulatory gestures -> sounds -> morphemes -> words -> phrases -> sentences. **Unpacked:** To go from noisy, ambiguous signals to meaning requires recognizing structure and applying domain knowledge at multiple levels of abstraction. A classic example: the sentences "How to recognize speech" and "How to wreck a nice beach" are virtually identical at the waveform level.

Big Idea #1: Perception	*Computers perceive the world using sensors.*	**Perception is the extraction of meaning from sensory information using knowledge.**	**The transformation from signal to meaning takes place in stages, with increasingly abstract features and higher level knowledge applied at each stage.**	**LO** = Learning Objective: what students should be able to do. **EU** = Enduring Understanding: what students should know.
Concept	**K-2**	**3-5**	**6-8**	**9-12**
Processing (Abstraction Pipeline: Vision) **1-B-iv**	**LO:** Demonstrate figure/ground segmentation by identifying the foreground figures and the background in an image. **EU:** Visual scenes have a structure that includes a background and a foreground, with foreground objects partially occluding (blocking) the background. **Unpacked:** Understanding that scenes have structure to them is a way to approach the problem of machine perception of images. Computers have to do foreground/background segmentation in order to pick out the objects in an image.	**LO:** Illustrate how the outlines of partially occluded (blocked) objects in an image differ from the real shapes of the objects. **EU:** Understanding complex scenes requires taking into account the effects of occlusion when attempting to recognize objects. **Unpacked:** consider two cereal boxes, one in front of and partially occluding the other. Students could be asked to draw (on a separate sheet of paper) the outline of each box as it appears in the image. The occluding box will be a rectangle, but the occluded box will be a concave polygon. Students could then be asked to draw the true outline of the box as a dotted line in the original image. They should realize that they are inferring the true outline; the box could be defective and have a different shape in the region they can't see.	**LO:** Describe how edge detectors can be composed to form more complex feature detectors, e.g., for letters or shapes. **EU:** The progression from signal to meaning takes place in stages, with increasingly complex features extracted at each stage. **Unpacked:** Example: detecting an "A" by looking for a combination of three oriented edges. Edges are detected by looking at pixels.	**LO:** Demonstrate how perceptual reasoning at a higher level of abstraction draws upon earlier, lower levels of abstraction. **EU:** Scenes are composed of objects, which are composed of surfaces and boundaries. Boundaries are marked by contours, which are composed of edges, which are made up of pixels. Relationships between objects in a scene, such as one object occluding another, are inferred from the arrangement of their surfaces and boundaries.
Domain Knowledge (Types of Domain Knowledge) **1-C-i**	**LO:** Describe some things an intelligent agent must "know" in order to make sense of a question. **EU:** To understand spoken requests, computers must know our vocabulary and pronunciation conventions, and they must be able to distinguish a question from a command. **Unpacked:** Understanding a spoken query such as "Will it rain today?" requires all the above knowledge.	**LO:** Demonstrate how a text to speech system can resolve ambiguity based on context, and how its error rate goes up when given ungrammatical or meaningless inputs. **EU:** Speech recognition systems are trained on millions of utterances, allowing them to distinguish common from uncommon sequences of words, which helps them select the most likely interpretation of the signal. **Unpacked:** Compare the transcription of "the jockey reined in the horse" vs. "the king reigned in the horse". Or test the system on "which witch is which" or "two ways to go is one too many". To explore grammatical influences, compare the transcription accuracy of a sentence read with normal word order vs. the same sentence read with the word order (not the individual words) reversed, e.g., "see the view" vs. "view the sea".	**LO:** Classify a given image (e.g., "traffic scene", "nature scene", "social gathering", etc.) and then describe the kinds of knowledge a computer would need in order to understand scenes of this type. **EU:** Domain knowledge for vision includes knowing what kinds of objects are likely to appear in a scene, where they are likely to appear in relation to other objects, and how occlusions and shadows can alter object appearances. **Unpacked:** In a traffic scene, cars appear on roads, some traffic signs appear alongside of roads but not in the road, some signs appear above the road, and pedestrians appear on sidewalks, in crosswalks, and occasionally on roads. In a nature scene, the top of the image is likely to be blue sky and the bottom of the image is likely to be green grass or trees.	**LO:** Analyze one or more online image datasets and describe the information the datasets provide and how this can be used to extract domain knowledge for a computer vision system. **EU:** Domain knowledge in AI systems is often derived from statistics collected from millions of sentences or images. **Unpacked**: sample image databases: ImageNet: https://image-net.org/ Coco: http://cocodataset.org/#explore Word prediction when typing texts or emails is an example of the use of statistical prediction similar to what is found in high level perception systems. Analyzing large collections of images produces statistics about what kinds of objects are likely to co-occur in a scene.

Big Idea #1: Perception	*Computers perceive the world using sensors.*	**Perception is the extraction of meaning from sensory information using knowledge.**	**The transformation from signal to meaning takes place in stages, with increasingly abstract features and higher level knowledge applied at each stage.**	**LO** = Learning Objective: what students should be able to do. **EU** = Enduring Understanding: what students should know.
Concept	**K-2**	**3-5**	**6-8**	**9-12**
Domain Knowledge (Inclusivity) **1-C-ii**	**LO:** Discuss why intelligent agents need to be able to understand languages other than English. **EU:** Speech recognition systems need to accomodate different languages because many different types of people will use them. **Unpacked:** Alexa settings provide for multiple dialects of English (US, British, Canadian, Indian), and numerous European and Asian languages.	**LO:** Discuss how domain knowledge needs to be broad enough to encompass all the groups an application is intended to serve. **EU:** Speech recognition systems need to accomodate different types of accents and alternative pronunciations. Music recognition systems must know about different musical genres.	**LO:** Describe how a vision system might exhibit cultural bias if it lacked knowledge of objects not found in the culture of the people who created it. **EU:** Domain knowledge must take multiple cultures into account if an AI application is to serve diverse groups. **Unpacked:** A self-driving car that only knows about American traffic signs will not be able to recognize traffic signs in Europe or Asia.	**LO:** Describe some of the technical difficulties in making computer perception systems function well for diverse groups. **EU:** Dark or low contrast facial features are harder to recognize than bright, high contrast features. Children's speech is in a higher register and less clearly articulated than adult speech.