VISUAL PERCEPTION

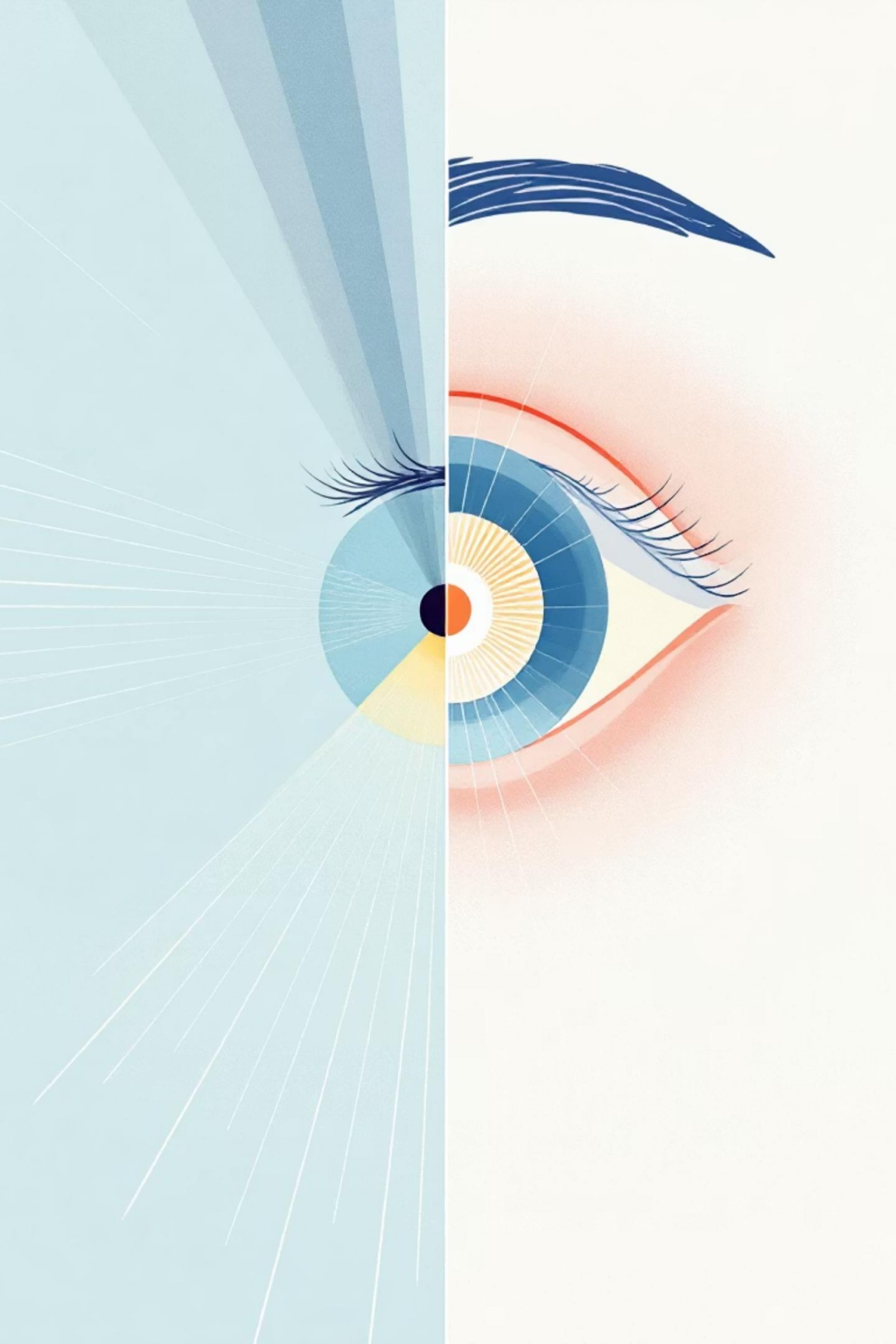## DIGITAL IMAGE PROCESSING

Dr. Irfan Ali, PhD
Assistant Professor (AI & MMG)

Aror University of Art, Architechture, Design & Heritage, Sukkur

Visual perception: Human vs. Computer Vision

Exploration of how we see and how machines learn to see. This lecture will compare the fascinating capabilities of human and computer vision, from processing light to perceiving depth and adapting to new challenges.

The Foundation of Sight: Light and Lenses

Human Visual System

Light enters the eye through the pupil, is focused by the lens onto the retina. The retina contains photoreceptor cells (rods and cones) that convert light into electrical signals.

- **Rods:** Detect light and dark, essential for night vision.
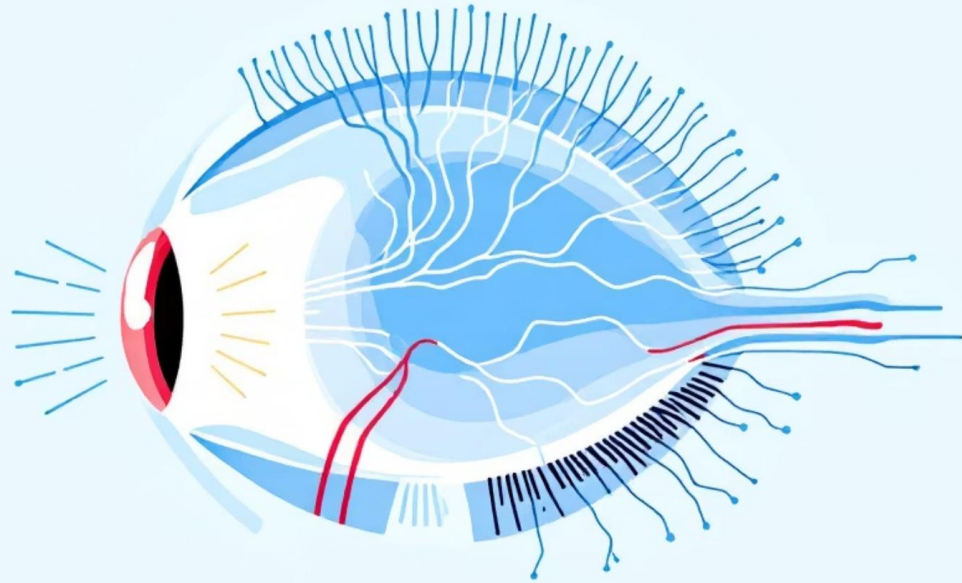- **Cones:** Detect color and fine details, responsible for day vision.

Computer Vision System

Cameras act as the "eyes," capturing light through a lens onto an image sensor (CCD or CMOS). Each pixel on the sensor records the intensity and color of light hitting it, converting it into digital data.

- **Lens:** Focuses light, similar to the human eye's lens.
- **Sensor:** Converts light into electronic signals, similar to the retina.
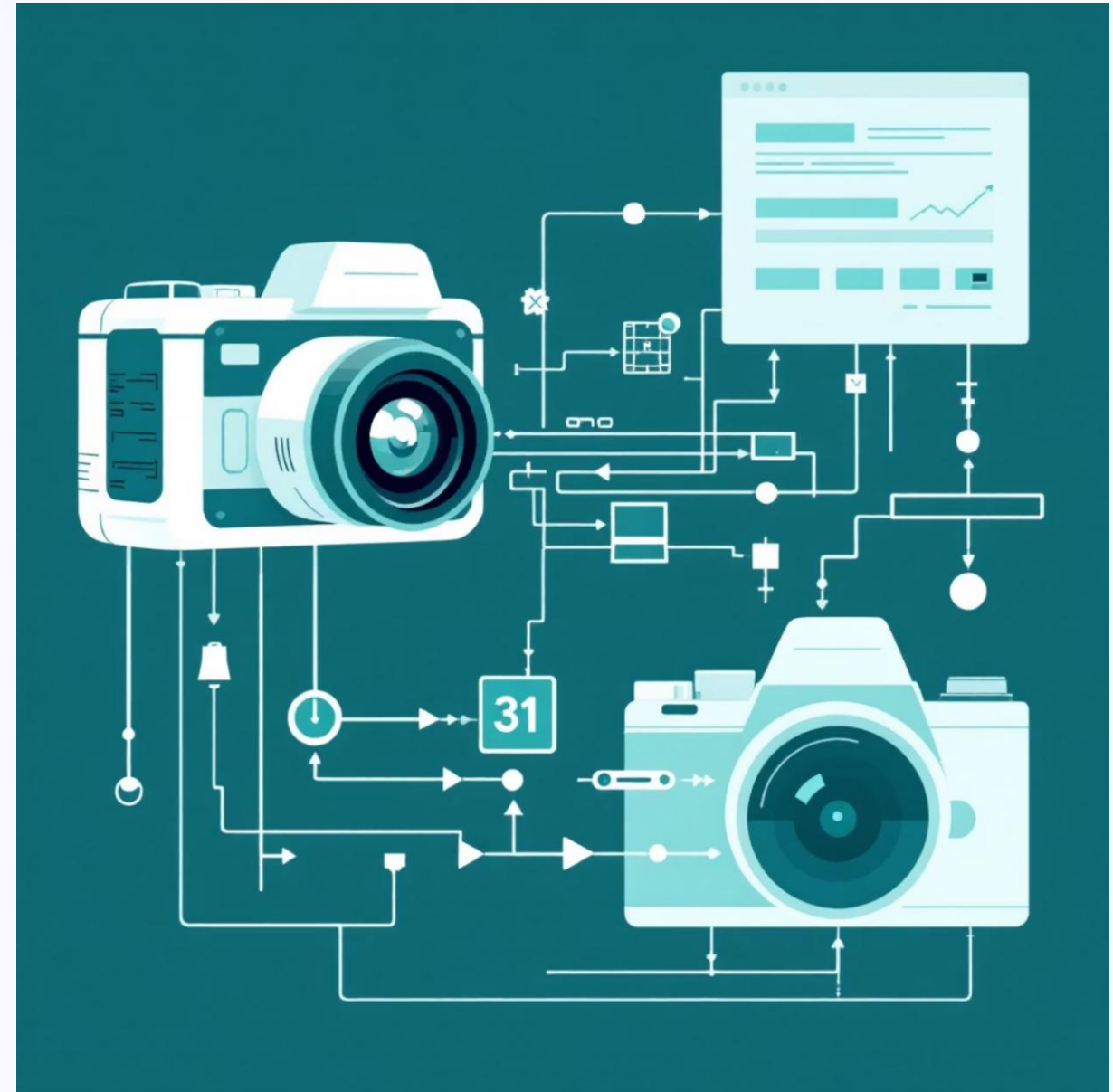
Mechanisms of Sight: How Each System Works

While both systems aim to recognize objects, patterns, and interpret scenes, they employ fundamentally different mechanisms and technologies.

The Human Eye & Brain

Cameras & Algorithms

Biological vs. Technological Architecture

The core difference lies in their fundamental building blocks and operational processes.

Human Visual System

Biological Basis: Relies on specialized biological organs and neural networks.

- **Eyes:** Cornea, lens, retina, photoreceptors capture light.
- **Optic Nerve:** Transmits electrical signals to the brain.
- **Brain:** Interprets signals, forms perceptions, and integrates with memory and emotion.

Computer Vision System

Hardware & Software Basis: Uses engineered components and computational models.

- **Cameras/Sensors:** Capture visual data (RGB, depth, thermal).
- **Processors (GPUs):** Perform rapid computations for algorithms.
- **Algorithms (AI/ML):** Analyze, identify, and interpret data patterns.
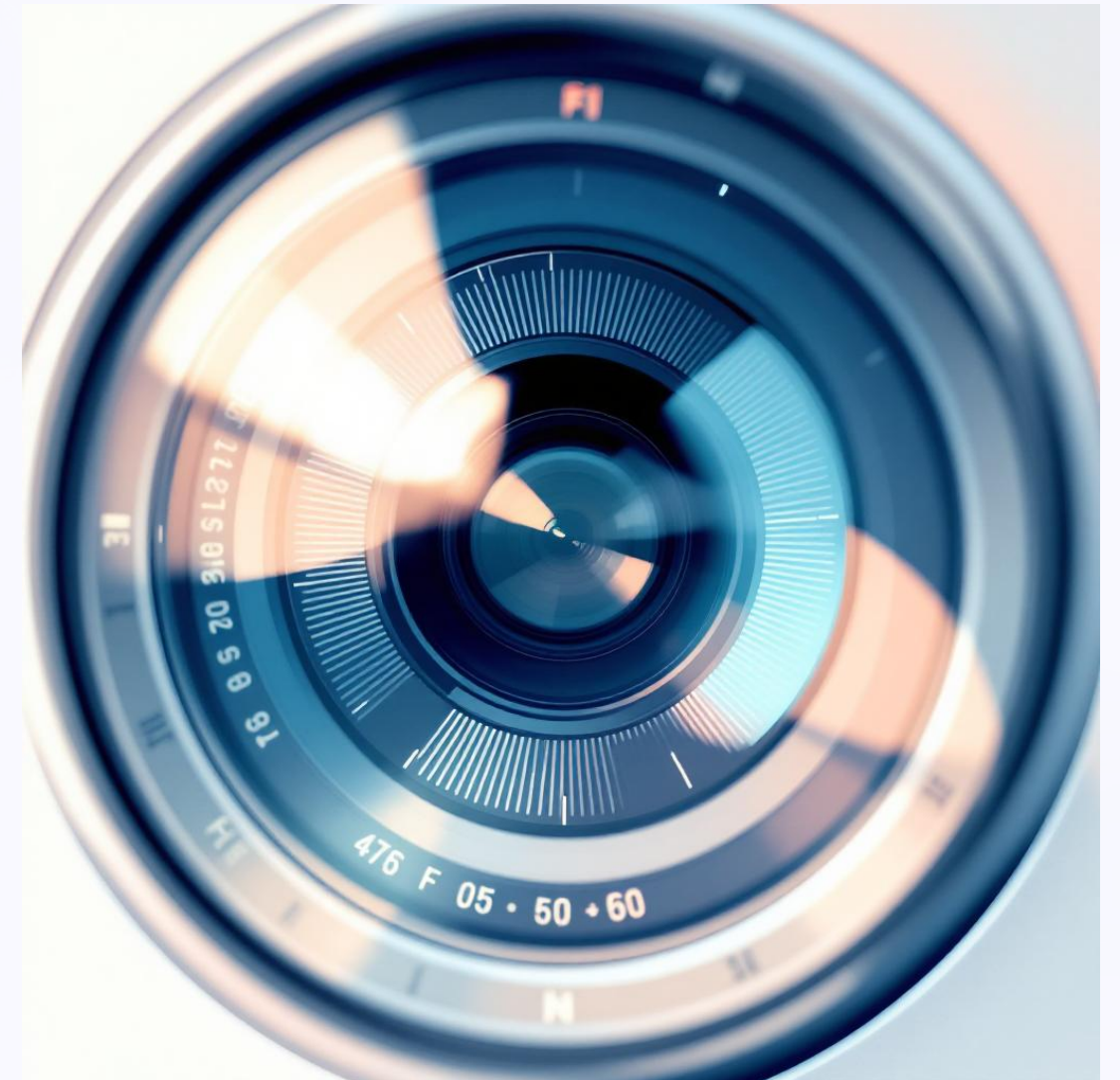
4

The Lens: Gathering Light

A lens is an optical device that **gathers and focuses light** from a scene onto a sensor or film. Its primary components include multiple glass elements arranged to correct aberrations and achieve sharpness. Key figures of a lens include:

Focal Length: Measured in millimeters (mm), this determines the angle of view and magnification. Shorter focal lengths (e.g., 24mm) offer a wider view, while longer ones (e.g., 200mm) narrow the view and magnify distant subjects.

Aperture (f-number): The size of the opening that lets light pass through. A lower f-number (e.g., f/1.8) indicates a wider opening, allowing more light in and creating a shallower depth of field (blurry background).

Image Stabilization: A mechanism that compensates for camera shake, crucial for sharp images in low light or with long focal lengths.

The Camera: Capturing the Image

The camera body houses the sensor, shutter, and image processing unit. It's responsible for converting the focused light into a digital file or exposing film. Key figures of a camera include:

Image Sensor Size: Larger sensors (e.g., full-frame) capture more light and detail, leading to better image quality, especially in low light. Common sizes include APS-C, Micro Four Thirds, and full-frame.

Megapixels (MP): The total number of pixels on the sensor, defining the resolution of the image. More megapixels allow for larger prints and more cropping flexibility.

ISO: A camera setting that controls the sensor's sensitivity to light. Higher ISO values are used in dark conditions but can introduce digital noise.

Shutter Speed: The duration the sensor is exposed to light. Fast shutter speeds (e.g., 1/1000s) freeze motion, while slow speeds (e.g., 1/30s) can create motion blur or capture more light.
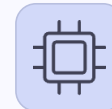
Decoding the Image: Processing and Interpretation

Human Brain

The optic nerve transmits signals from the retina to the brain, where complex neural networks interpret shapes, colors, movements, and patterns. This processing happens almost instantaneously and subconsciously.

Computer Processors

Digital image data is processed by algorithms that perform tasks like edge detection, feature extraction, and pattern recognition. This requires significant computational power and pre-programmed rules or learned models.

While the human brain excels at intuitive, holistic interpretation, computer vision systems rely on structured, algorithmic analysis to make sense of visual data.

Depth Perception and 3D Vision

Human Visual System

Humans use binocular vision (two eyes) to perceive depth, allowing them to estimate distances and interpret 3D space. The brain merges slightly different images from both eyes to create a detailed 3D understanding of the world. This is known as stereopsis.

Computer Vision System

Can simulate depth perception through stereo vision (two or more cameras), LiDAR sensors (Light Detection and Ranging), or specialized depth sensors (e.g., infrared). It requires complex algorithms to estimate depth based on visual cues like parallax or object shape, often mapping points in a 3D space.

Humans have a natural, seamless understanding of depth, while computers rely on specific technologies to simulate it with mathematical precision.

2D and 3D vision system

2D vision perceives objects in only two dimensions: **width** and **height**. Our eyes capture images on our retinas, which are inherently 2D surfaces. When we look at a photograph or a screen, we are primarily experiencing 2D vision. There is no information about depth or distance, making everything appear flat.

- Perceives width and height.
- Lacks depth information.
- Common in art and digital displays.

3D vision, or stereoscopic vision, allows us to perceive objects in three dimensions: **width**, **height**, and crucially, **depth**. This is achieved because our two eyes are slightly separated, each capturing a slightly different perspective of the same scene. Our brain then combines these two images, interpreting the discrepancies as depth and giving us a sense of spatial awareness.

- Perceives width, height, and depth.
- Relies on binocular disparity.
- Crucial for navigation and object manipulation.

2D Versus 3D vision system

Learning and Adaptability: Humans vs. Machines

Human Visual System

Humans adapt quickly to new environments, understand complex scenes, and recognize objects even in unusual contexts (e.g., recognizing a face in low light or a partially obscured object). Visual learning happens through continuous experience and natural learning processes, allowing for remarkable generalization.

Computer Vision System

Relies on machine learning algorithms that require large datasets for training. Systems can improve over time with more data (e.g., deep learning for object detection). However, they are generally less adaptable to completely new or unseen environments without significant retraining or adjustment of the underlying algorithms.

Humans can generalize better and adapt to novel situations with ease, while computer vision often requires continuous refinement and extensive data exposure for comparable performance.

11

Human Learning

Humans learn through **experience, intuition, and complex reasoning**. A child learning to ride a bike might fall many times, but each attempt provides immediate feedback, allowing them to adapt their balance and coordination. This process involves a deep understanding of cause and effect, often without explicit rules.

- Relies on intuition and emotional context.
- Adapts to novel situations with generalized understanding.
- Learns from very few examples (e.g., one bad experience).

Machine Learning

Machines learn by **identifying patterns in vast datasets** and applying algorithms. For example, a machine learning model can be trained on millions of images of bikes and people to recognize them. Its "understanding" is derived from statistical correlations rather than inherent comprehension.

- Requires large amounts of labeled data for training.
- Excels at specific, repetitive tasks once trained.
- Struggles with abstract concepts or unforeseen scenarios.

Real-World Applications

Both human and computer vision play crucial roles in our daily lives, each with unique strengths.

Human Driving

Navigating complex traffic, reacting to unexpected events, understanding social cues.

Autonomous Vehicles

Using cameras, LiDAR, and radar to detect obstacles, lane markings, and traffic signs.

Medical Diagnosis

Physicians interpreting X-rays, MRIs, and patient symptoms for a holistic diagnosis.

AI in Healthcare

Analyzing medical images for early detection of diseases, assisting with surgical precision.

Human Car driving and Autonomous car driving

Strengths & Limitations

Human Vision

- **Strengths:** Contextual understanding, emotional interpretation, creativity, adaptability to novelty.

- **Limitations:** Subject to biases, fatigue, limited memory retrieval, can be fooled by illusions.

Computer Vision

- **Strengths:** Precision, speed, consistency, tireless operation, ability to process vast datasets.

- **Limitations:** Requires large training data, struggles with unseen scenarios, lacks common sense and intuition, ethical concerns.

Summary

From basic light reception to complex interpretation, human and computer vision systems operate on fundamentally different, yet equally powerful, principles.

- **Biological vs. Algorithmic:** Human vision is intuitive and adaptive; computer vision is precise and data-driven.

- **Complementary Strengths:** Neither is inherently "better"; they excel in different domains.

- **Future Collaboration:** The most exciting advances will likely come from integrating both human insight and AI precision.