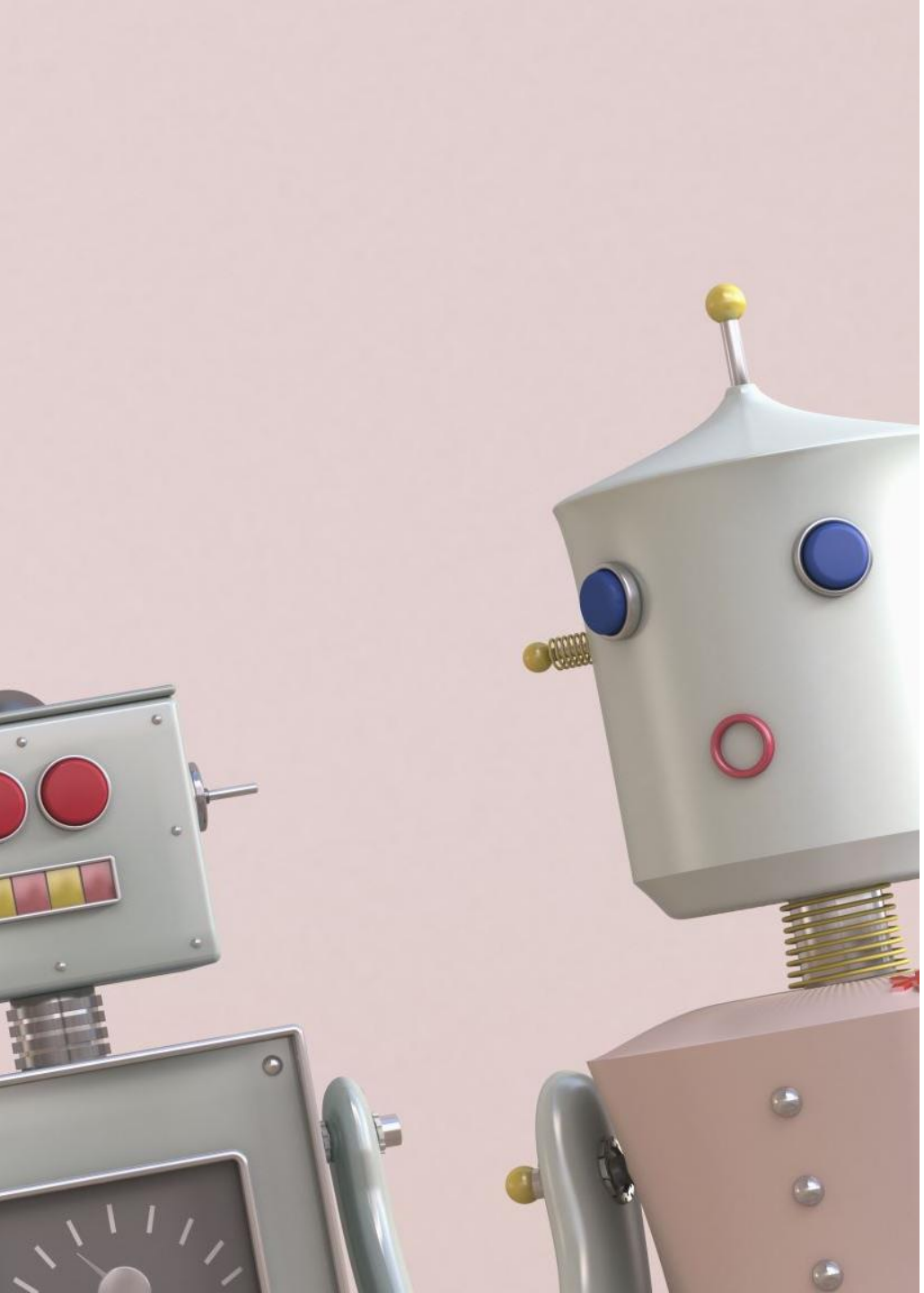


SOCIAL ROBOTS AND HUMAN-ROBOT INTERACTION

Week 5, Social Perception

Ana Paiva

P2
2025/2026

WHAT ARE “SOCIAL” ROBOTS ?

A social robot is a **physical entity embodied in a complex, dynamic, and social environment**, sufficiently **empowered to behave in a manner conducive to its own goals and those of its community,** being **capable to recognize others and engage in social interactions with them**.

Duffy, B.R., Anthropomorphism and the social robot. Robotics and autonomous systems, 2003. 42(3): p. 177-190.

THE PROBLEM

Based on the limited perceptual capabilities of a robot, how to build technology to understand the social situation and the user's (and other agents') affective, social, motivational and informational states, in order to respond in a socially appropriate manner.

Shakey the Robot

Sensing: Shakey the Robot (SRI, 1966-1972)

Sensors in Shakey were state-of-the-art for 1960s, enabling basic perceptic and navigation in controlled environments.

- Vidicon TV camera
- Optical range finder (spinning disk with light source)
- Whisker bump sensors (tactile)

Also:

- Wheel shaft encoders (odometry)
- Contact switches on bumpers

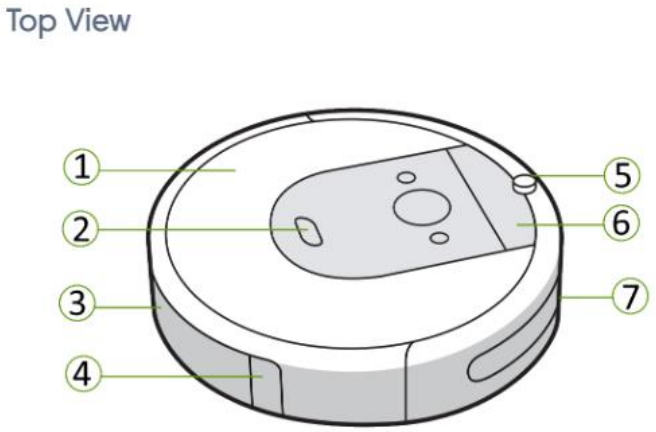
"Shakey" was the first mobile robot with the ability to perceive and reason about its surroundings.

The subject of SRI's Artificial Intelligence Center research from 1966 to 1972, Shakey could perform tasks that required planning, route-finding, and the rearranging of simple objects. The robot greatly influenced modern robotics and AI techniques; today, it resides in the Computer History Museum.

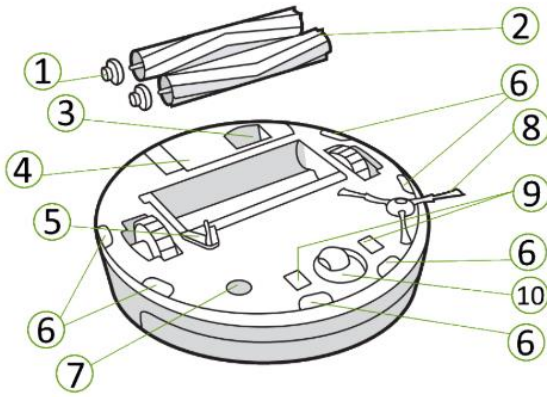
Sensing: Roomba

Sensors in roomba enabe basic perception, navigation and dirt detection in controlled home environments.

Sensors:
- Forward-facing camera (ceiling-facing for vSLAM models)
- Downward-facing optical floor tracking sensor (for navigation)
- 4 infrared cliff sensors (bottom-front)
- Bump sensors (tactile)
- Acoustic dirt sensor (under chassis)



1. Faceplate
2. Camera
3. Dust Bin and Filter
4. Bin Release Button
5. RCON Sensor
6. Handle
7. Light Touch Sensor



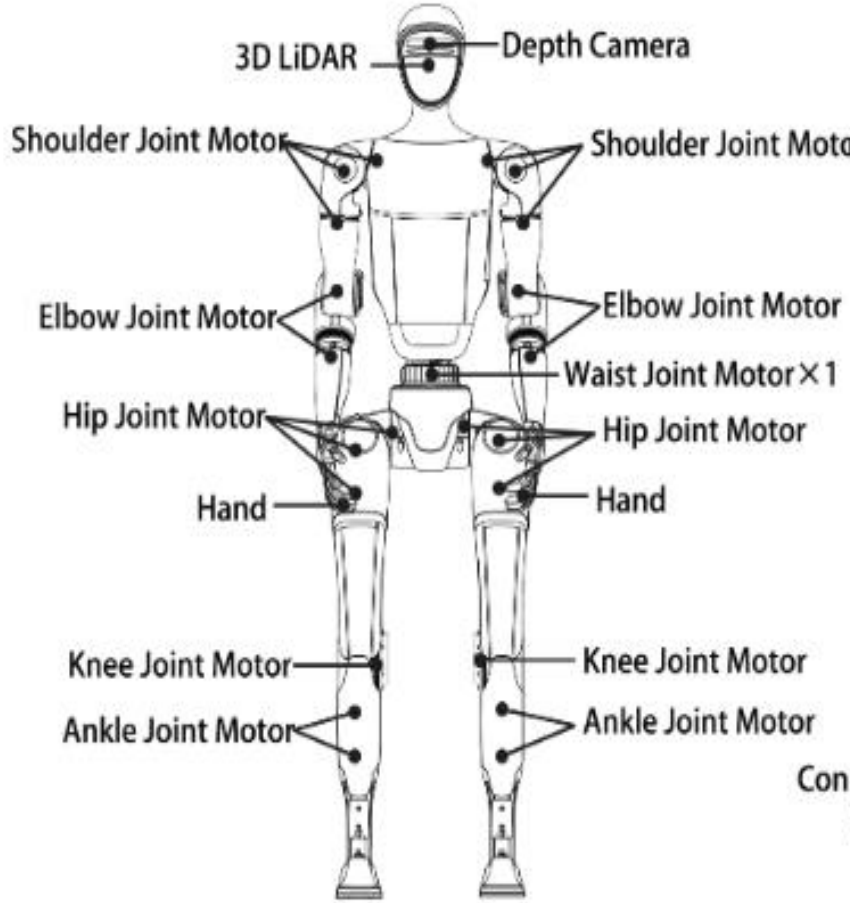
1. Brush collapse Std
2. Multi-Surface Brushes
3. Dirt Disposal Port
4. Dust Bin
5. Brush Frame Release Tab
6. Cliff Sensors
7. Floor Tracking Sensor
8. Edge-Sweeping Brush
9. Charging Contacts
10. Caster Wheel

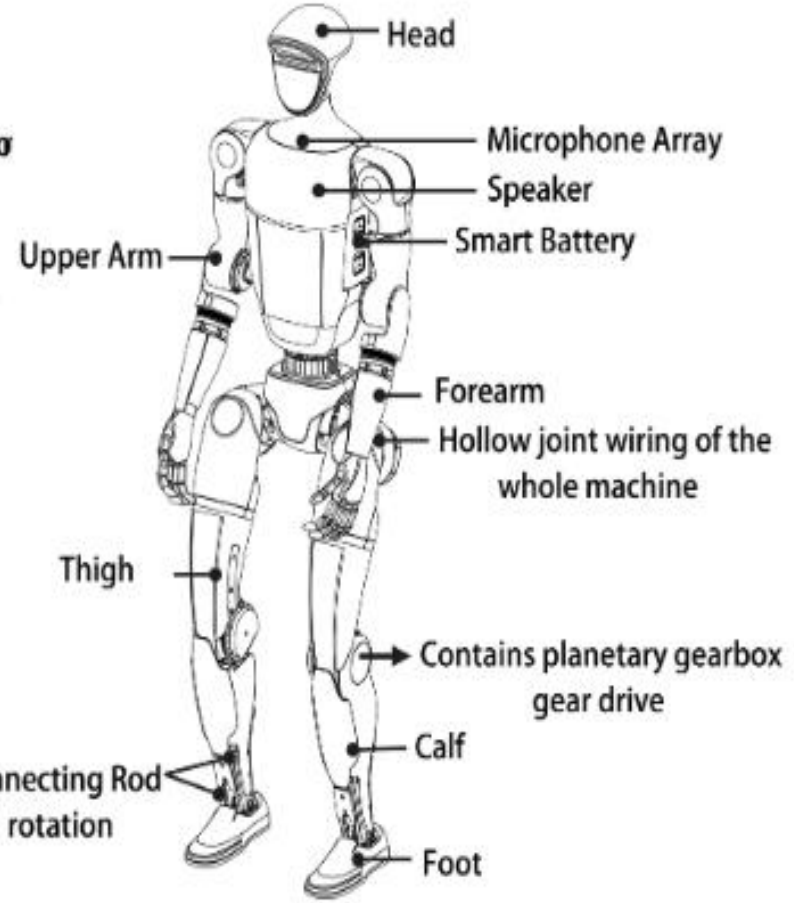
Sensing: Unitree G1 Humanoid

Sensors:
- Vision: Stereo RGB cameras and a depth camera
- LiDAR: 3D LiDAR (mounted on head)
- Microphone array

Other
- Force/Torque Sensors: 6-axis force/torque sensors in feet and ankles

The G1 basic version offers 23 degrees of freedom in total, allowing for precise motion and posture control through joint motors.





PERCEPTUAL CAPABILITIES: ROBOTICS' SENSORS

Different types of classifications: passive versus active sensors; simple versus complex sensors; exteroceptive versus proprioceptive sensors

Exteroceptive Sensors (Environment Perception)

- Vision: Cameras (2D/3D) to see and identify objects and surroundings.
- Range/Proximity: LiDAR (light), Radar (radio waves), Sonar (sound waves), Infrared (IR) to measure distance and detect nearby objects.
- Tactile/Contact: Touch sensors, force, and torque sensors to feel pressure, and measure forces on joints.
- Audio: microphones and sound sensors to monitor surroundings.
- Environmental: Temperature, humidity.

Mataric, M. J. (2007). The robotics primer. MIT press.

ROBOTS VERSUS HUMANS PERCEPTION CAPABILITIES

Humans sensory systems: Vision; Audition; Smell; Taste; Touch

Plus: temperature perception; pain perception; proprioception (perception of the body muscles, etc); interoception (hunger, thirst, heartbeat, breathing, etc); …

In terms of individual sensors (receptors)

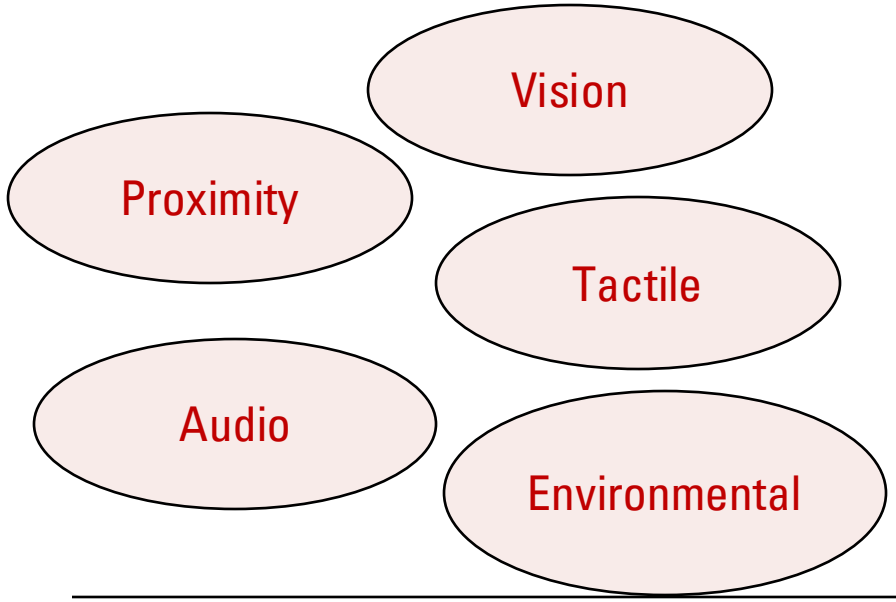
-> the human body has hundreds of millions of individual sensors….<-

-> recent humanoid robots can have about 120- 150 sensors…<-

FROM SENSORS TO SOCIAL PERCEPTIONS

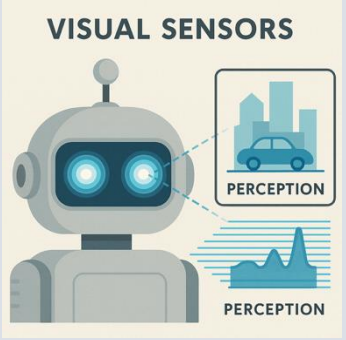
In humans, social perception is how to form impressions, representations and make inferences about others (humans). It is how to perceive and understand individuals and groups in a social context.

How to go from sensors to understanding of individuals (humans) and groups in a social context

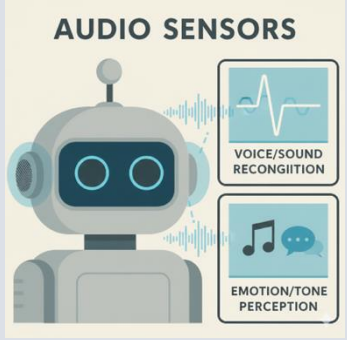


Identity,
Facial expressions,
Gaze direction (attention),
Body posture,
Gestures,
Proximity,
Action,
Relationships.
Emotions
Attitudes…

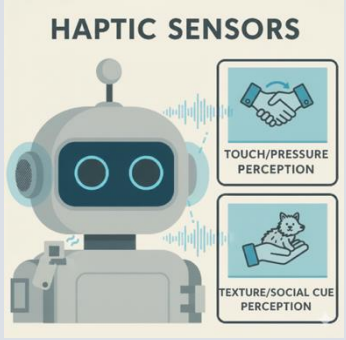
EMBODIMENT: FROM SENSORS -> PERCEPTION



Hardware for Vision:
Stereo RGB and depth
3D LiDAR
What (perception of):
Objects, People,
Identity,
Facial expressions,
Gaze direction, Gestures
Body posture,
Proximity…..



Hardware for Audio
Microphones,
arrays of microphones
Perception of: Speech content,
tone,
pitch,
volume,
non-verbal vocalizations (laughs, sighs).

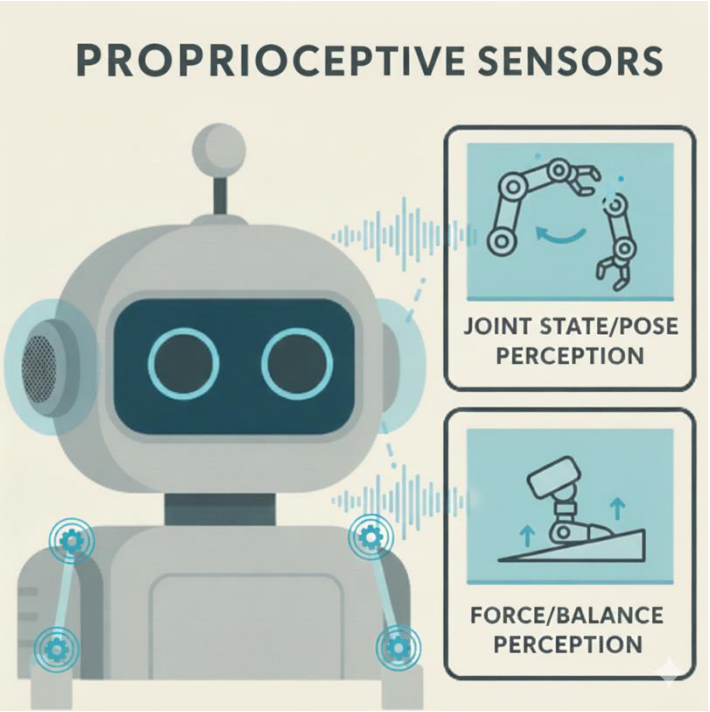


Hardware for Haptic
Touch sensors,
Force-Torque sensors.
Perception of
body location
pressure,
stroke,
hit, etc.

EMBODIMENT: INTERNAL SENSING

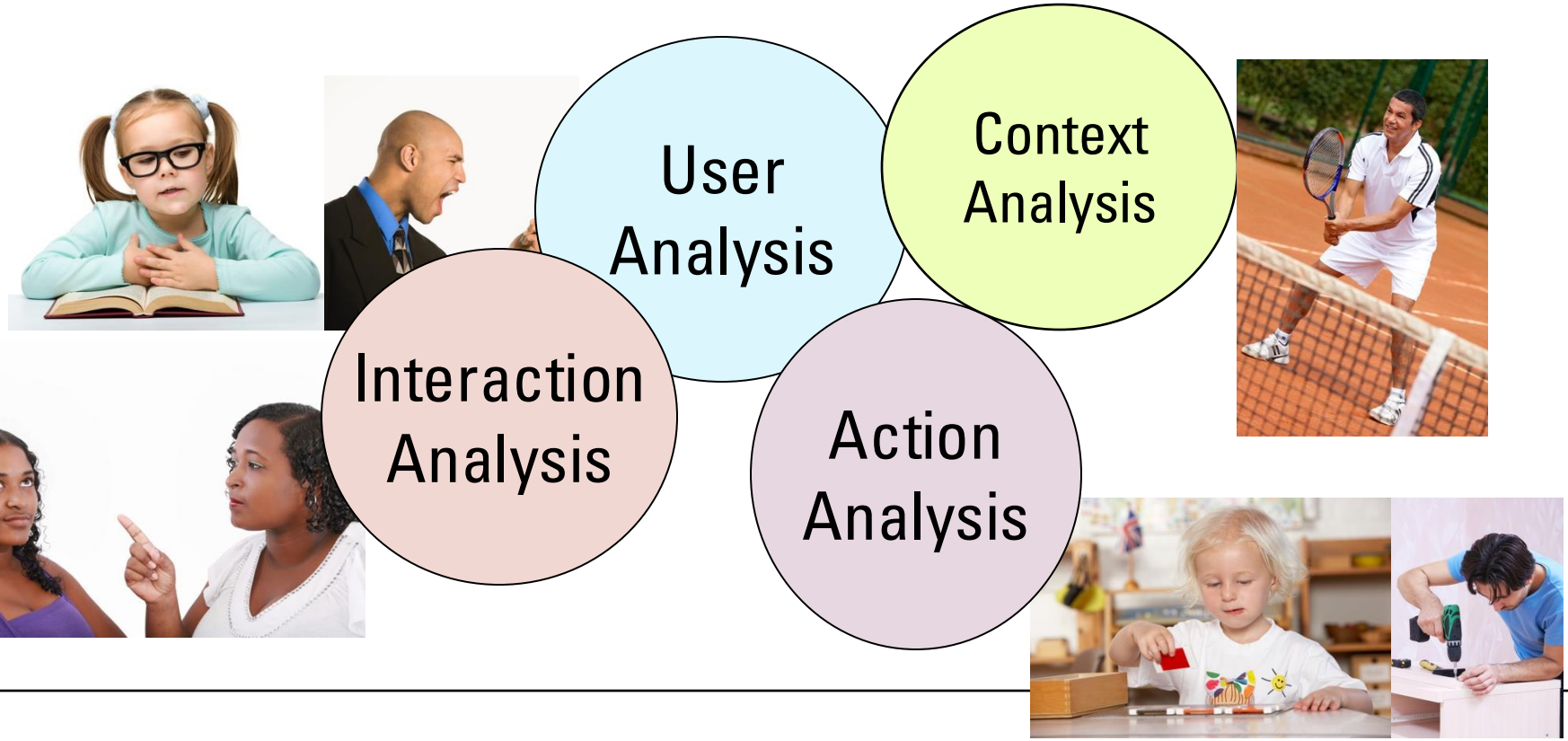
Proprioceptive Sensors (Internal State)

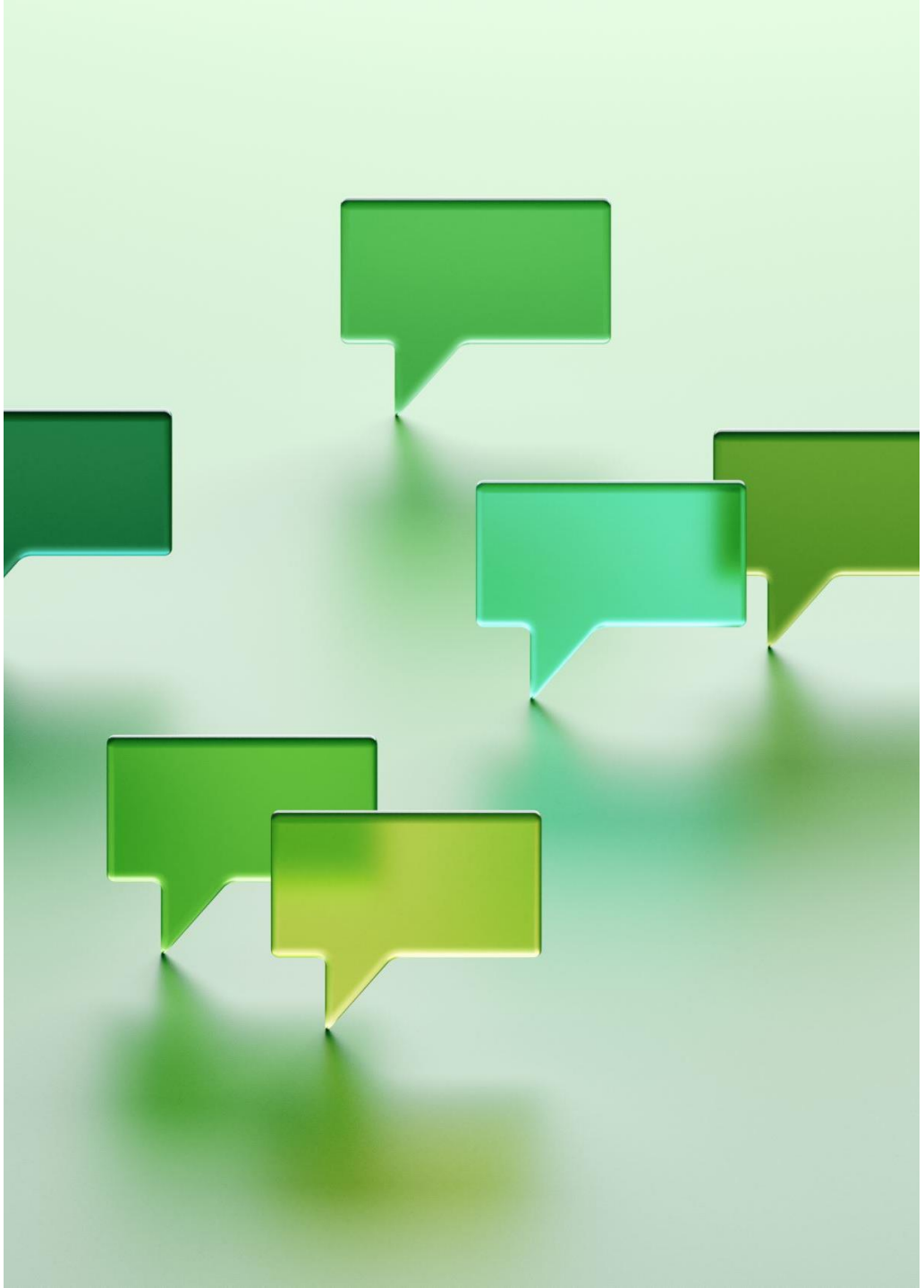
- Motion/Position: Gyroscopes (orientation), Accelerometers (acceleration, tilt) to track movement and orientation.
- For positioning: GPS…



SOCIAL PERCEPTION

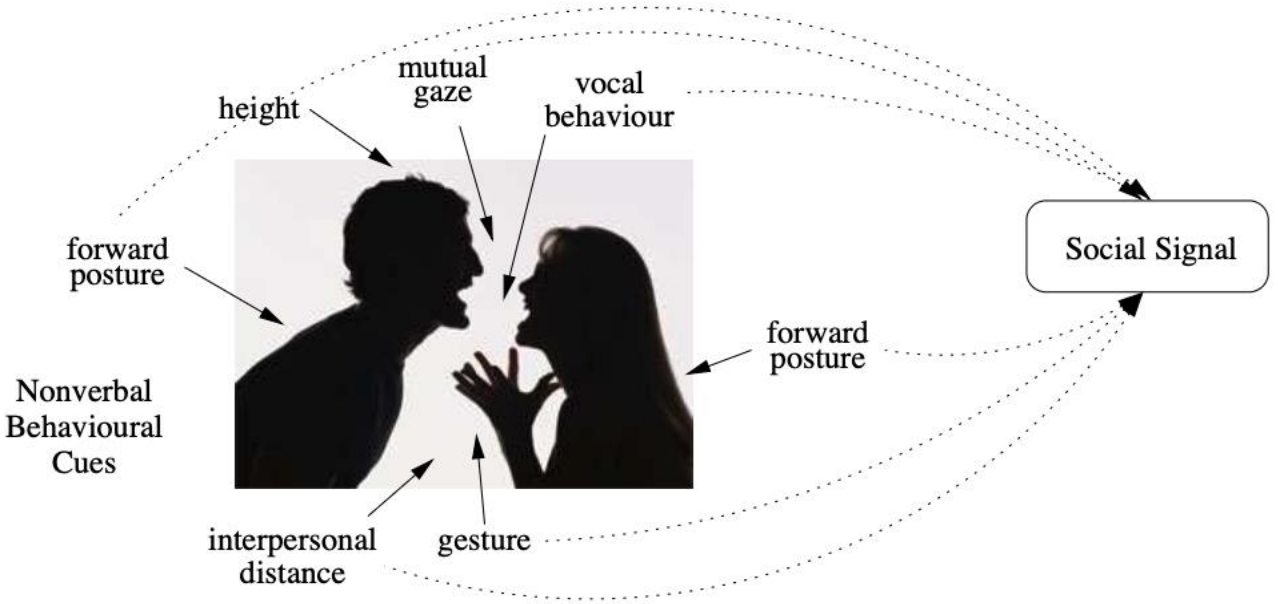
HOW TO GO FROM SENSORS TO UNDERSTANDING OF INDIVIDUALS (HUMANS) AND GROUPS IN A SOCIAL CONTEXT





SOCIAL SIGNALS PROCESSING

Goal: analyse and generate nonverbal cues (like facial expressions, gaze, gestures, tone) to build socially intelligent machines and robots



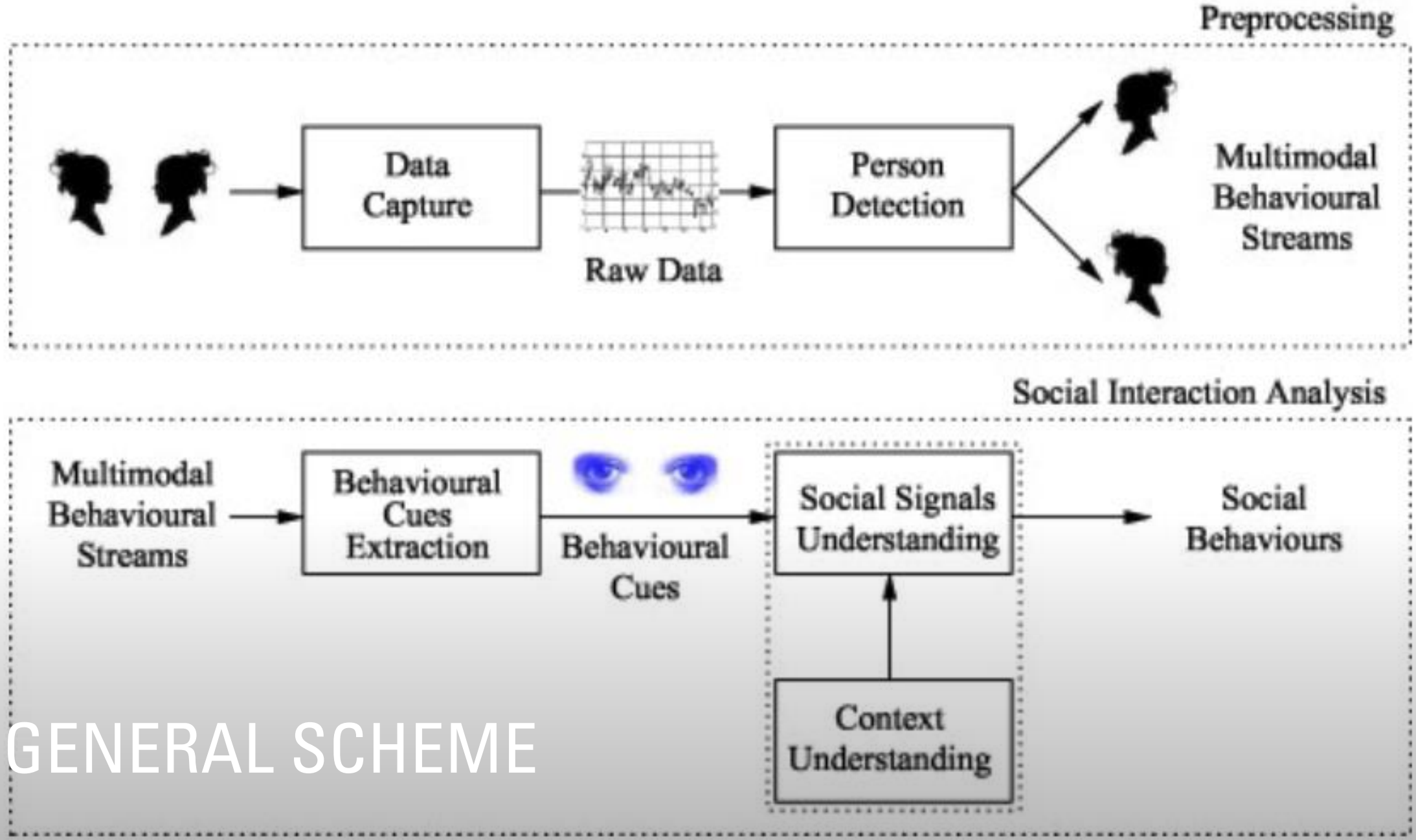
Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and vision computing, 27(12), 1743-1759.

SOCIAL PERCEPTION

1. **Person detection & Identification**
2. **Social behaviours and emotions**

Using

- **Face**
  - Face recognition
  - Face detection and tracking
  - Facial expression (emotions) analysis
  - Gaze tracking
- **Body**
  - Body detection and tracking
  - Hand tracking
  - Recognition of posture, gestures and activity
- **Vocal nonlinguistic signals**
  - Estimation of auditory features such as pitch, intensity, and speech rate
  - Recognition of nonlinguistic vocalizations like laughs, cries, sighs, and coughs



GENERAL SCHEME

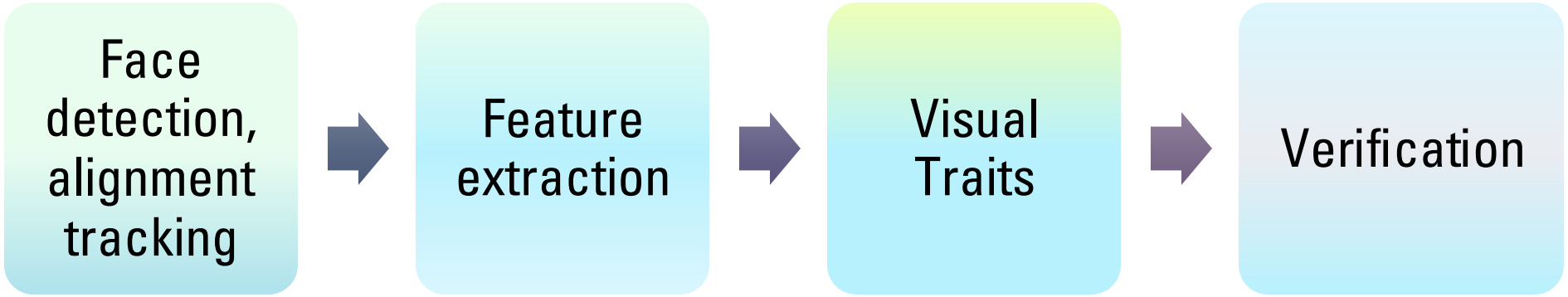
1. PERSON DETECTION & IDENTIFICATION

- **Face**: Facial Recognition (Identification through Face) Facial identification is highly robust and widely used. The process generally involves detection of face, alignment, feature extraction, and matching.

- **Voice**: Speaker recognition (or voice biometrics) identifies who is speaking, not what is being said (which is speech recognition).

- **Gait and Posture**: Gait recognition analyzes the way a person walks, while posture/pose recognition deals with the static shape of the body. These are often used for identification at a distance where face and voice data are unavailable.

FACE DETECTION & RECOGNITION: REAL TIME USE



Face detection, alignment & normalization- Techniques are used to geometrically normalize the face (e.g., center the eyes, adjust the head tilt) before feature extraction.

Extract Features: For each face image low-level features are extracted (for example normalized pixel values, image gradient directions) these vectors to form a large feature vector F(I).

Visual Traits: For each extracted feature vector, "trait vector" are built for the face. These classifiers may be focused on attributes such as gender, age, and race, which provide strong cues about a person's identity.

Verification: To decide if face matches one already in the system (calculating if the new user is he same person, can be done by comparing their trait vectors using a final classifier D.

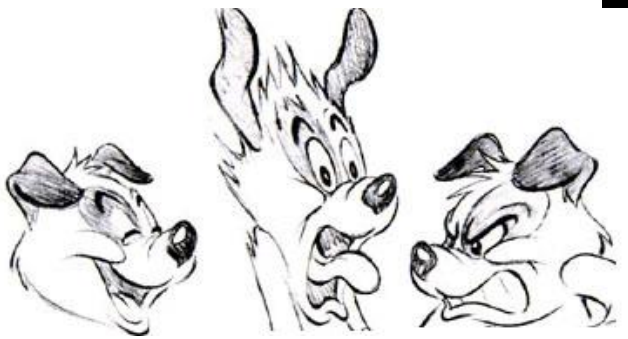
IDENTIFICATION WITH VOICE

- Speaker recognition (or voice biometrics) identifies ***who*** is speaking, not ***what*** is being said (which is speech recognition). It is divided into text-dependent (speaker must say a specific phrase) and text-independent (speaker can say anything).
- Humans have characteristics:
  - **Pitch/Fundamental Frequency** (F0): Average speed of vocal cord vibration. Vocal Tract Shape. Speaking Rate: Phonemes/words per minute.
  - **Pronunciation Patterns:** Dialect, accent, unique word stress.
  - **Voice Quality:** Breathiness, roughness, tension.

RECOGNISING EMOTIONS IN A FACE

What does a facial expression show?

- The internal physical state of a person
- An indication of what he/she is going to do next
- The plans, expectations and memory
- The emotional state



EXPRESSIONS: EMOTIONS AND MOOD RECOGNITION

Markers of Emotional Expression

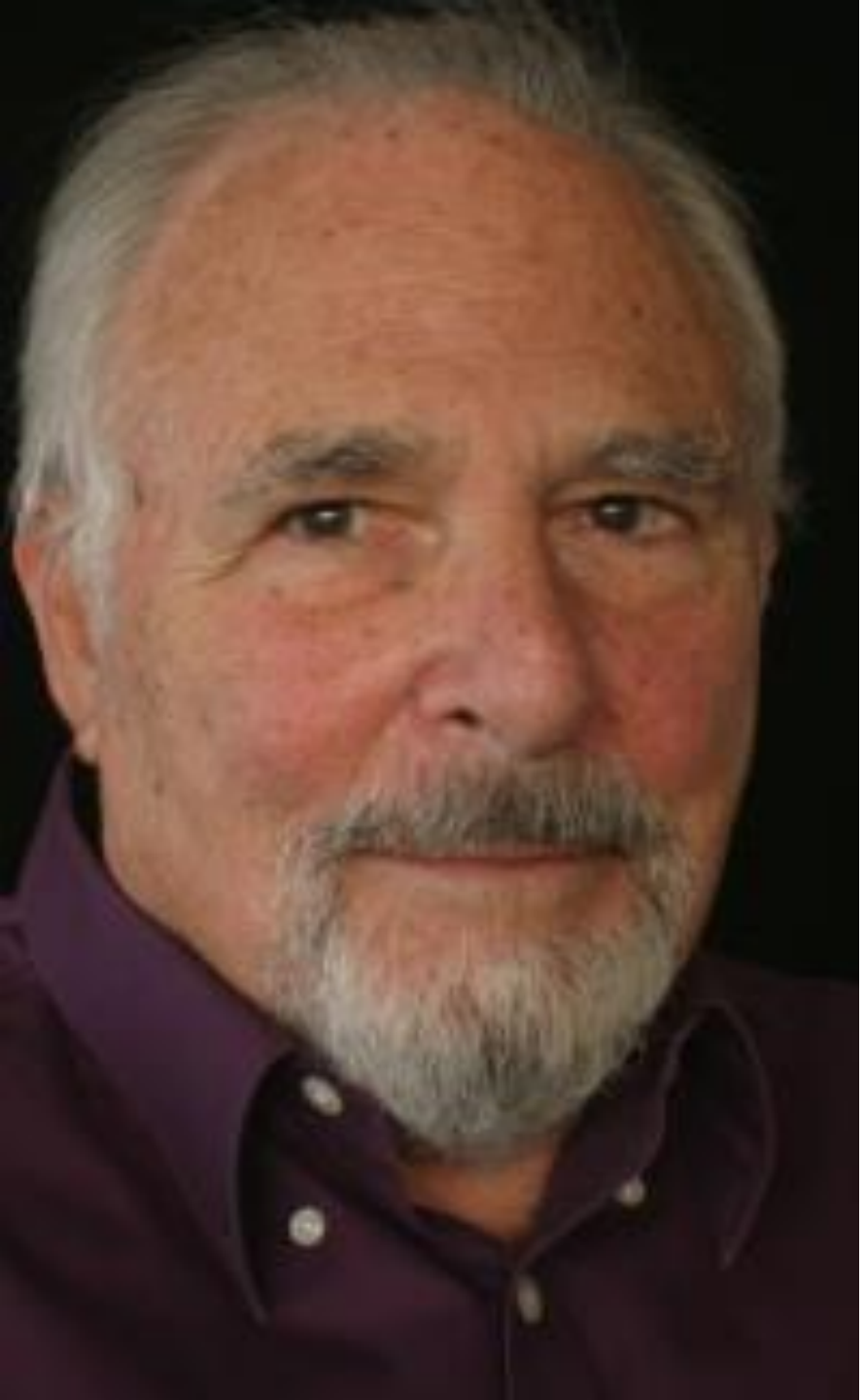
Involuntary muscle actions that people cannot deliberately produce/suppress

Usually last a couple of seconds

smile with enjoyment - 10 seconds

polite smile without emotion (exceptionally brief ¼ second or it can be enforced during long periods of time)

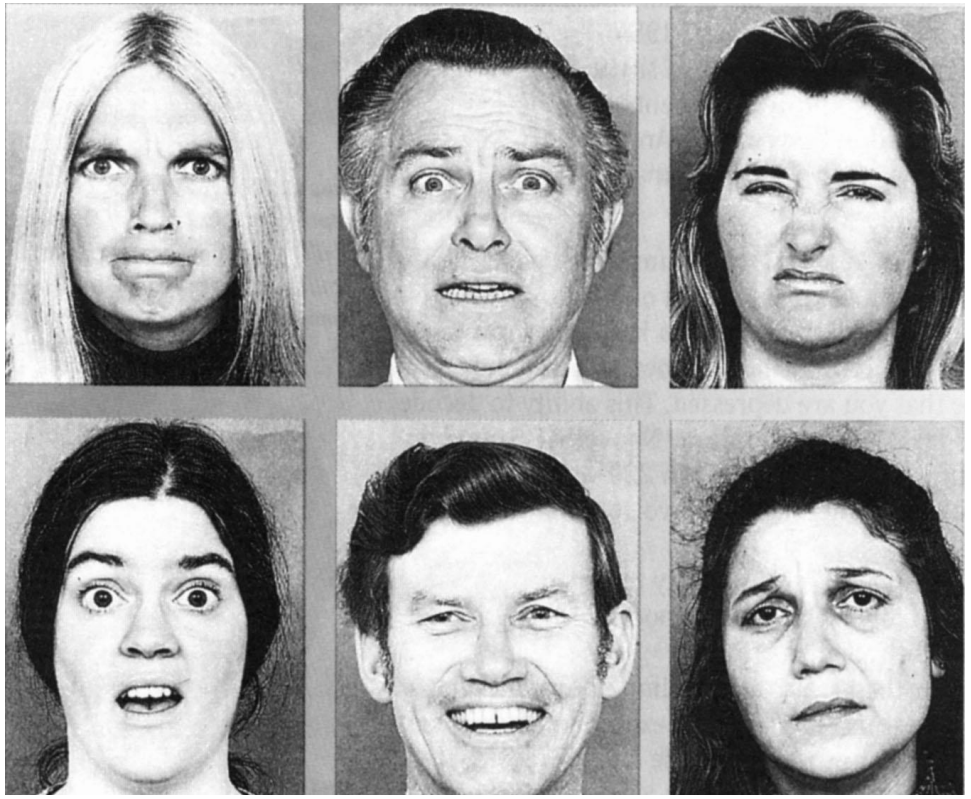
Duchenne smile-spontaneous smile that involves the contraction of major muscles. Research shows Duchenne smiles leads to perception of people as "trustworthy, likable, and emotionally healthy."



RECOGNISING EMOTION EXPRESSIONS: PAUL EKMAN

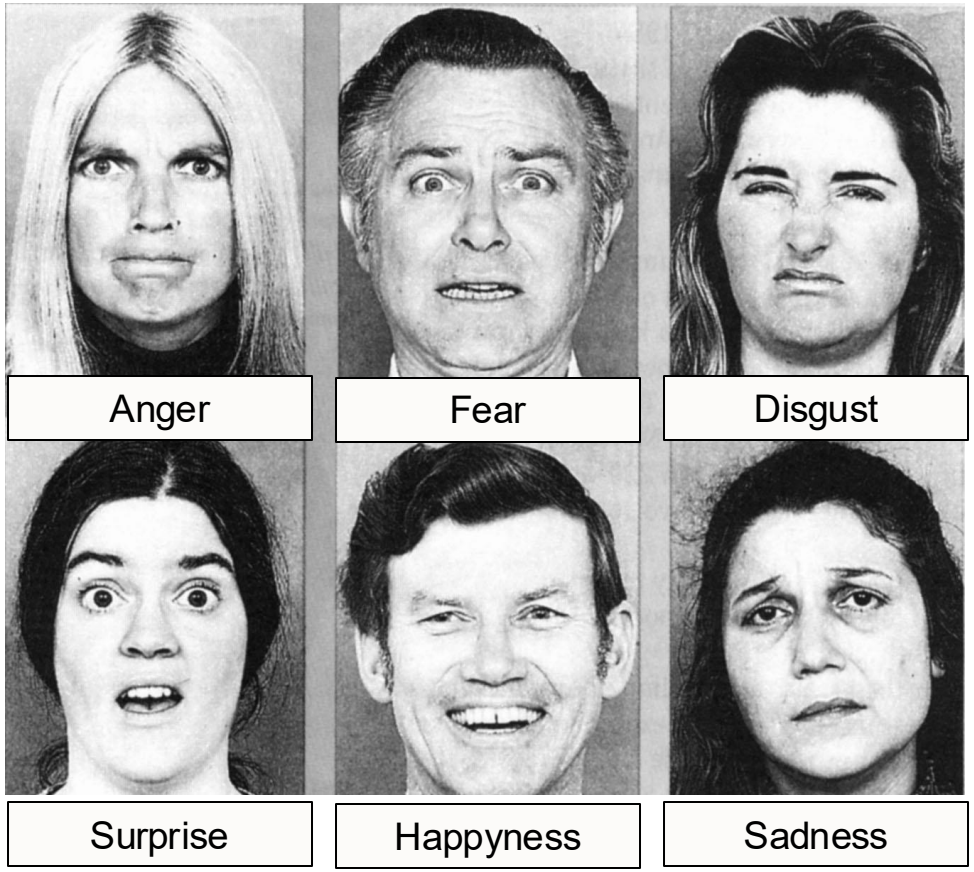
- **Paul Ekman** (February 15, 1934 – November 17, 2025) was an American psychologist and professor at the University of California, San Francisco, who was a pioneer in the study of emotions and their relation to facial expressions.[1][2] He was ranked 59th out of the 100 most eminent psychologists of the twentieth century in 2002 by the *Review of General Psychology*.[3]

UNIVERSALITY OF FACIAL EXPRESSIONS



- First test of the universal hypothesis
  - 3000 photos of different people
  - 6 basic emotions

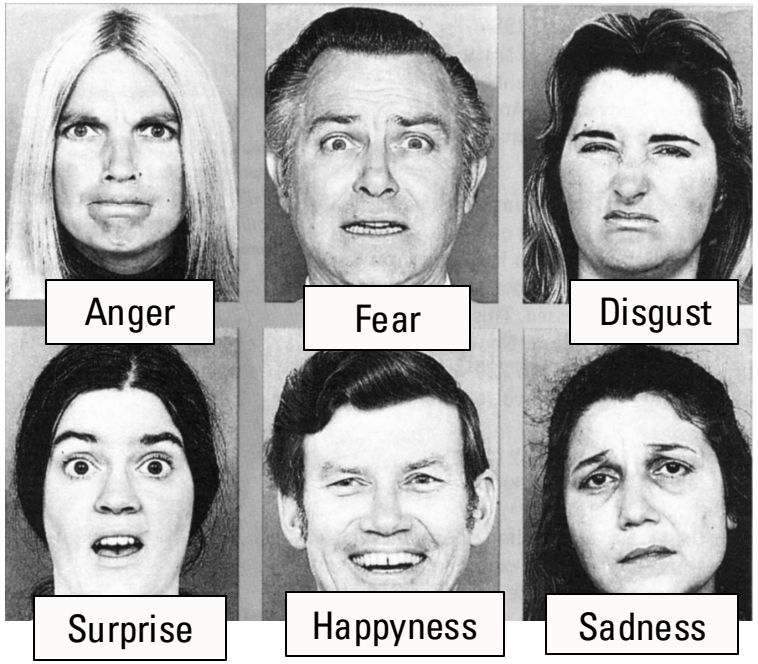
UNIVERSALITY OF FACIAL EXPRESSIONS



- First test of the universal hypothesis
  - 3000 photos of different people
  - 6 basic emotions

UNIVERSALITY OF FACIAL EXPRESSIONS: EKMAN'S STUDIES

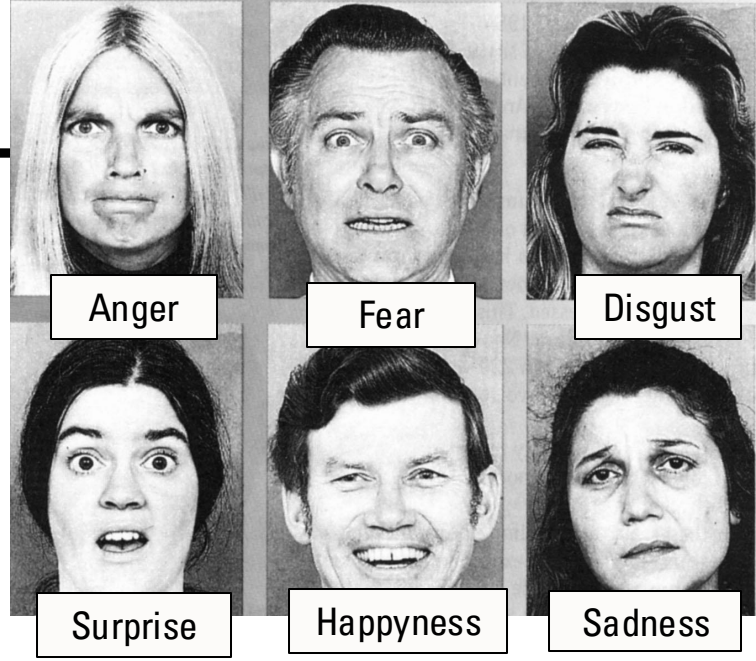
- First test of the universal hypothesis
  - Photos showed to participants across countries
    - Japan, Brasil, Argentina, Chile, U.S
  - Participants were asked to select the emotion term that better matched the displayed emotion
    - From a list of 6 terms
  - Accuracy rates of 80-90%
    - In all countries



Critics:

Participants had seen U.S. television and movies and might have learned american labels for the expressions

UNIVERSALITY OF FACIAL EXPRESSIONS: EKMAN'S STUDIES



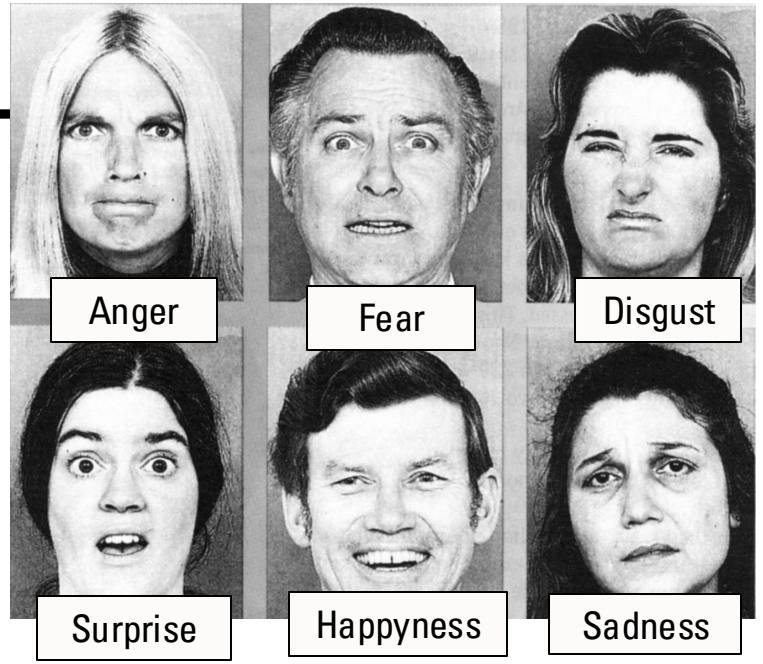
Second experiment: Ekman travelled to Papua, new Guinea (Fore):

- lived 6 months with people of the Fore tribe
- did not see any movie or magazine
- did not speak English
- minimal exposure to westerners

Two tasks were designed to evaluate the universality hypothesis

	Fore participants judging western photos		U.S students judging Fore expressions
	Adults	Children	
Anger	84	90	51
Disgust	81	85	46
Fear	80	93	18
Happiness	92	92	73
Sadness	79	91	68
Surprise	68	98	27

HOW TO MEASURE?



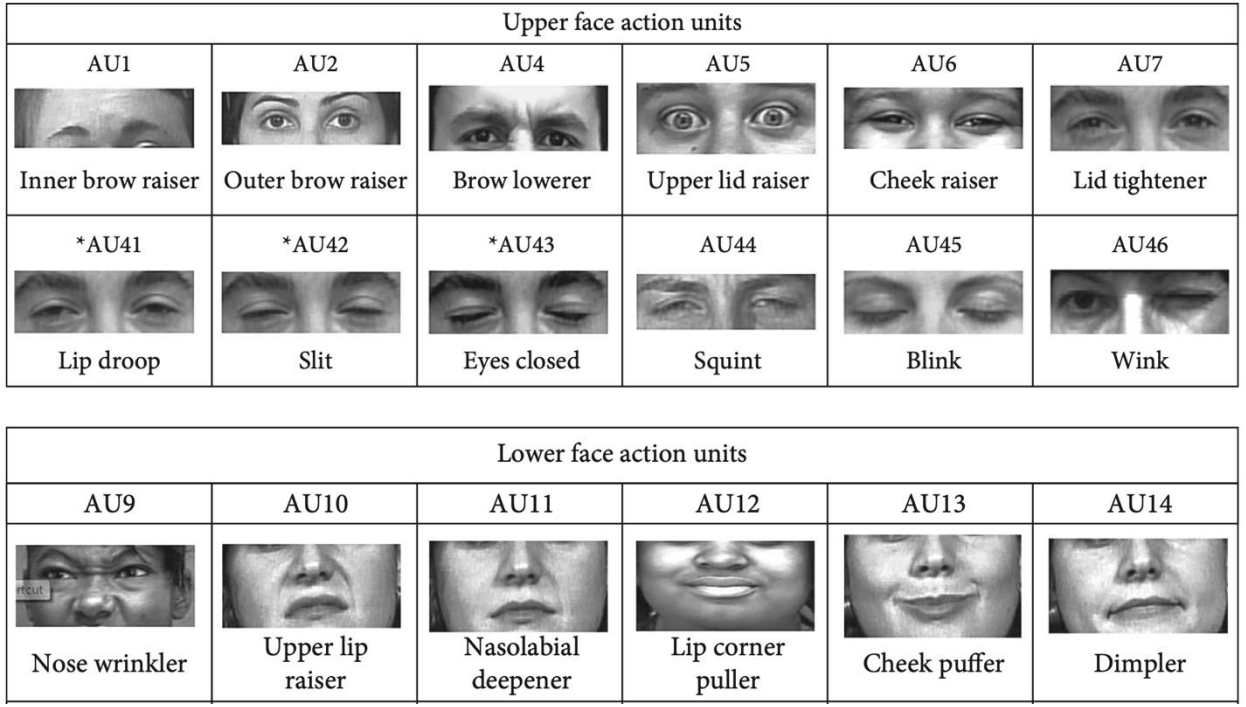
- Using a large dataset trained with different expressions
- Sign-based measurements: a descriptive approach that takes into account the relation between facial signs and emotions

Development of the: Facial Action Coding System (FACS)

FACS

The FACS taxonomy describes 44 Action Units (AU) that can be coded binary (presence/absence) or with a number (corresponding its intensity).

Designed to detect subtle changes in facial features
Viewing videotaped facial behaviour in slow motion, trained observers can manually FACS code all possible facial displays, which are referred to as Action Units (AU)

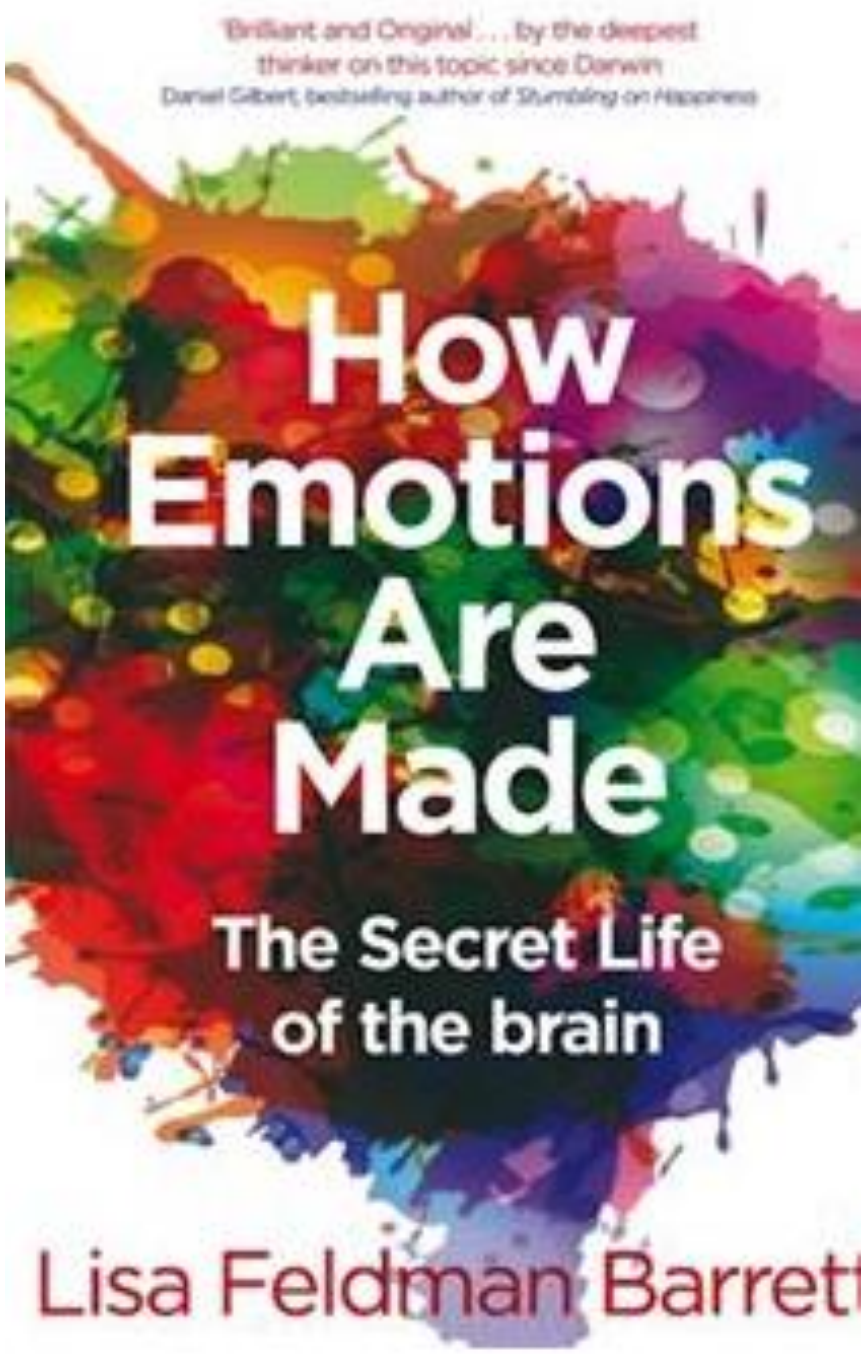


CRITICISMS

Oversimplification of Emotion: Lisa Feldman Barrett (and others) argue that companies are building on a potentially flawed foundation (universal, discrete emotions). Many scientists contest such theory in modern affective science.

Decontextualization: FACS itself, and the software that uses it to analyse facial expressions, do so in a vacuum, ignoring context. However, context (what the person is saying, with whom she is with etc), is crucial for true emotional understanding.

Ethical and Bias Concerns: The AU detection models are trained on data that is often biased (e.g., mostly lighter-skinned faces), and the computational models built will be less accurate for others, leading to perpetuating bias at scale. (problems raise in hiring, security, or policing).



RESTRICTIONS

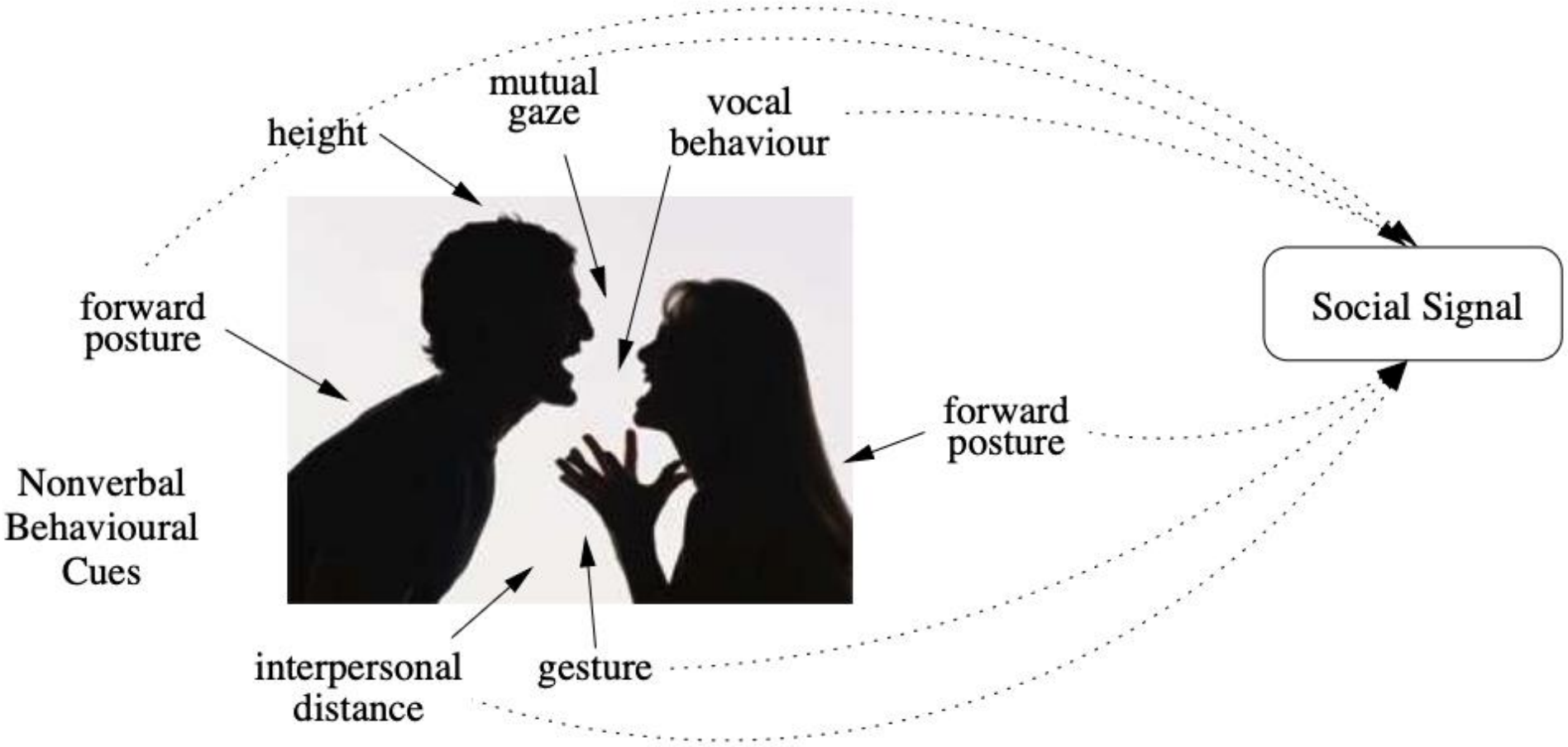
The AI act puts limitations on emotion recognition systems.

In sensitive contexts like the **workplace and education, where they are broadly prohibited**.

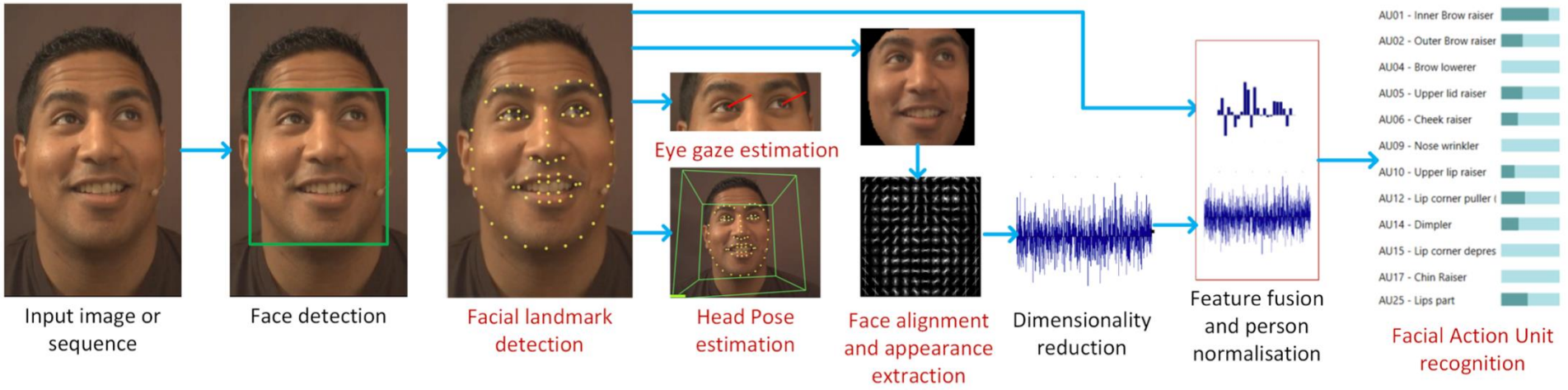
Workplaces: Prohibited use includes monitoring employees' emotional tone in call centers, during virtual meetings, or for performance assessment and recruitment purposes.
Educational Institutions: Banned uses include monitoring student engagement, attention levels, or emotional responses in classrooms or e-learning environments.

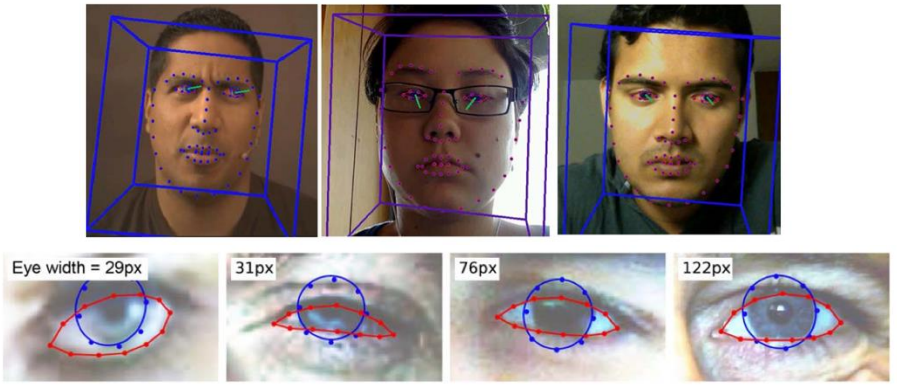
There are exceptions and in other contexts, these systems are generally classified as high-risk and subject to stringent regulations.

AND GAZE? AND POSTURE?



GAZE TRACKING – EXAMPLE: OPENFACE





Baltrusaitis, Tadas, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. "Openface 2.0: Facial behavior analysis toolkit." In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pp. 59-66. IEEE, 2018.

GESTURES: TYPES

Emblems

Gestures directly translated to words

E.g. peace sign

Different meanings across cultures

E.g. giving the finger

Illustrators/Iconic gestures

Gestures or facial expressions that accompany speech to make it vivid, visual, or empathic

E.g. illustrating the size of something, throwing a ball, finger pointing, raised eyebrows

Regulators

Nonverbal behaviours used to coordinate conversation

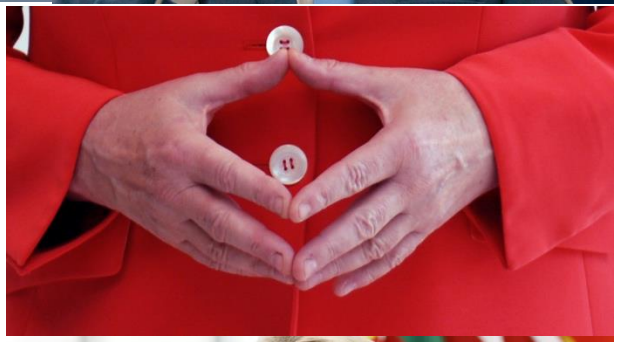
E.g. head nodes

E.g. looking at/orienting the body towards someone

Self-adaptor

Unconscious behaviours that release nervous energy

E.g. touching face, tug hair, bite lips



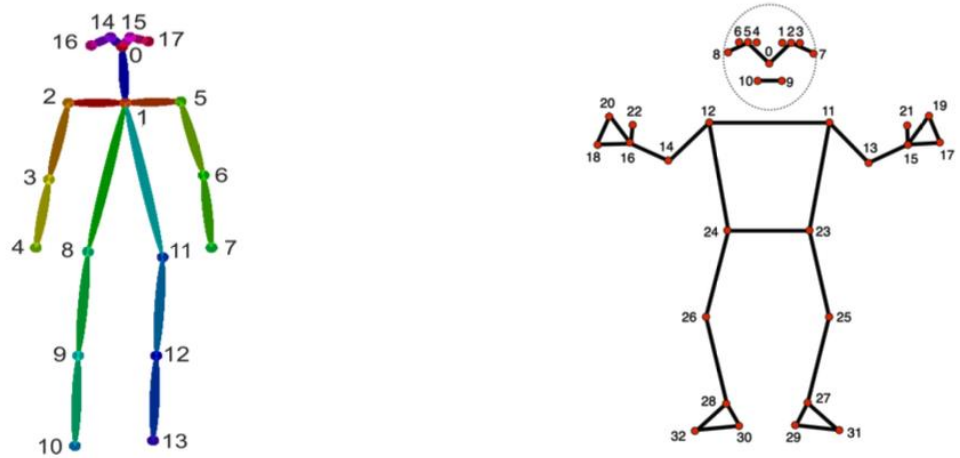
- Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

HUMAN BODY ANALISYS

Focused on detecting and localizing key body features, from the nose to the toes, in images or video inputs.

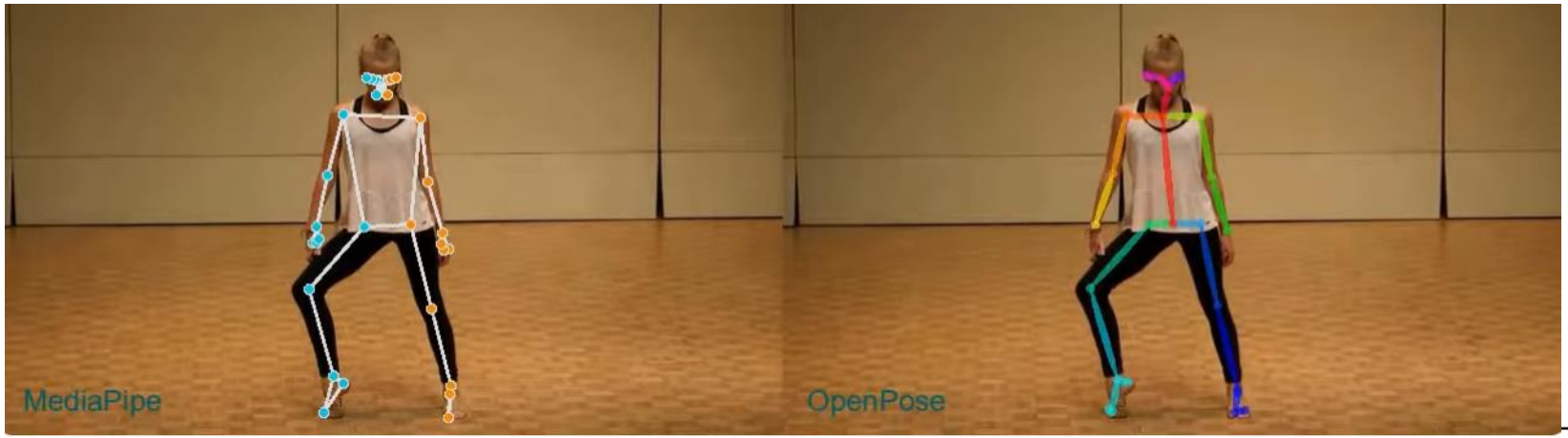
The goal is to accurately identify the position and orientation of a person's body within a scene.

Applied in diverse fields, such as video surveillance, medical assistance , and sports motion analysis, etc.



(a) Openpose Skeleton **(b)** MediaPipe Skeleton

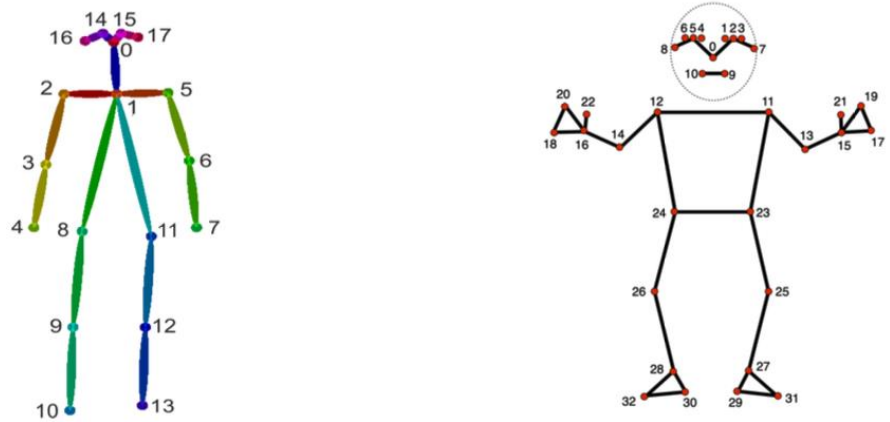
Figure 3.1: Comparison of skeleton representations: OpenPose and MediaPipe.



TÉCNICO LISBOA

TOOLS FOR HUMAN POSE ESTIMATION

Tools available:
OpenPose, PoseNet, MoveNet, MediaPipe Pose rely on either top-down or bottom-up methods to detect keypoints in human bodies.



(a) Openpose Skeleton

(b) MediaPipe Skeleton

Figure 3.1: Comparison of skeleton representations: OpenPose and MediaPipe.

Table 3.1: Specifications of each HPE library.

HPE Libraries	Released Year	Max Number Keypoints	Keypoints Position in Body Parts	Type of Pose	Method	Underlying Network	Pose Output
OpenPose [50]	2017	135	Face, hand, head, upper body, lower body	Single and multi person	Bottom-up	ImageNet with VGG-19	2D
PoseNet [51]	2017	17	Head, upper body, lower body	Single and multiperson	Top-down	ResNet and MobileNet	2D
MediaPipe Pose [52]	2020	33	Head, upper body, lower body	Single and multi person	Top-down	CNN	3D
MoveNet [53]	2021	17	Head, upper body, lower body	Single and multi person	Bottom-up	MobileNetV2	2D

IDENTIFYING HUMAN ACTIONS

Human actions usually involve human-object interactions, where we can see articulated motions along complex temporal structures.

Actions are spatio-temporal patterns!

FROM PERCEPTIONS TO ACTIONS

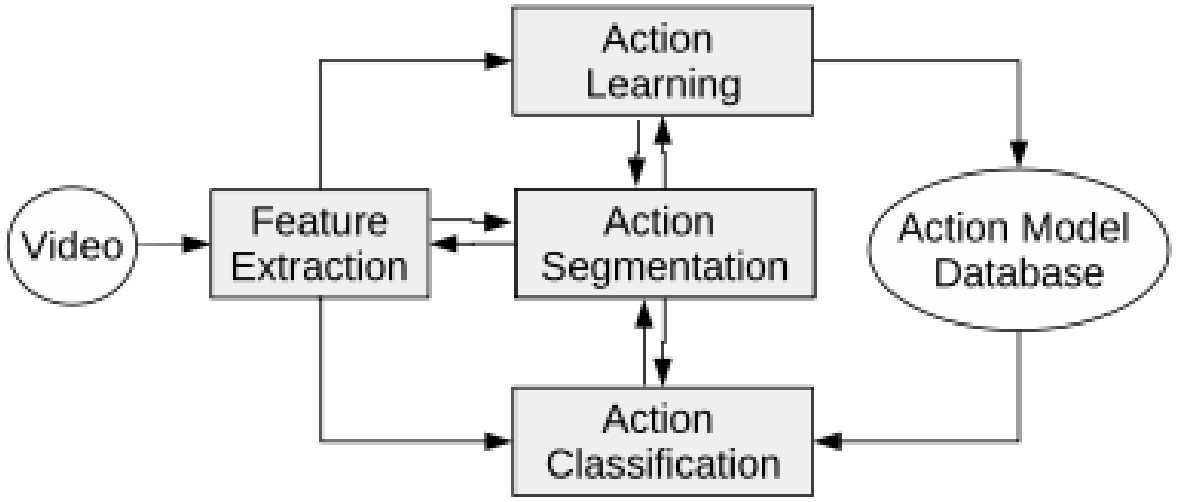


Figure 1: A typical data-flow for generic action recognition system comprises inter-dependent stages of feature extraction, learning, segmentation and classification.

IDENTIFYING HUMAN ACTIONS

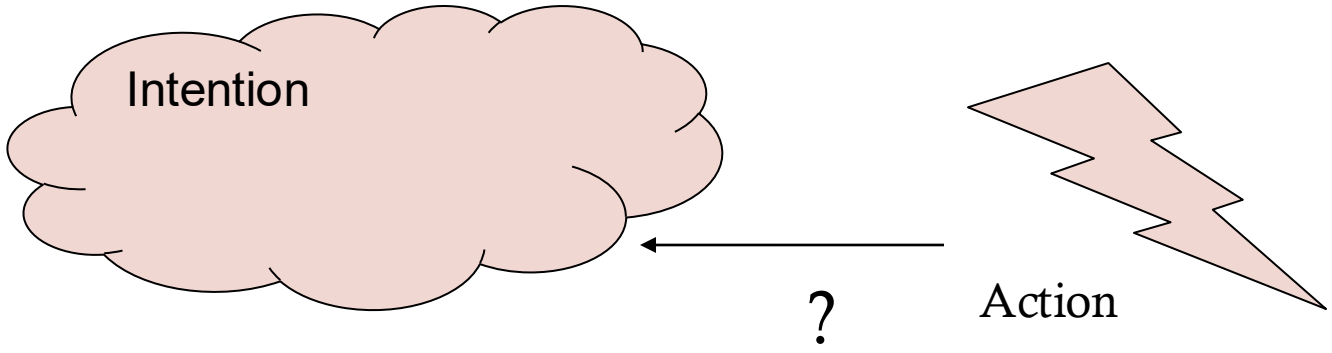
Issues in action recognition:

- Extraction and representation of suitable spatio-temporal features
- Modeling and learning of dynamical patterns

PROBLEM: Limited datasets

INFERING INTENTIONS FROM ACTIONS

Goal: not only to figure out the action of the user but also, "why" the user is performing that action



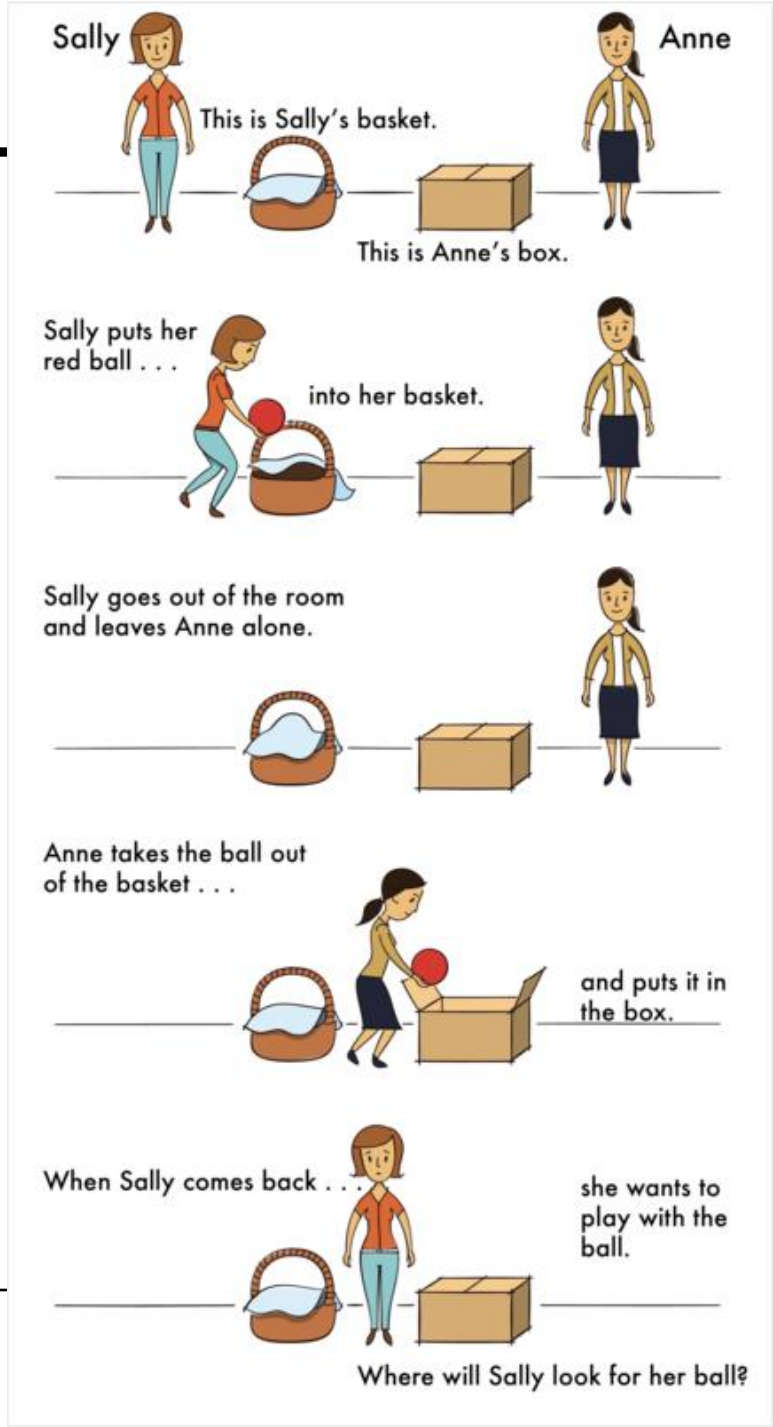
- Possible approaches:
  - Using object affordances to anticipate the human's next activity in order to enable the robot to plan ahead for a reactive response
  - Using **Theory of Mind** for intention recognition

WHAT IS THEORY OF MIND?

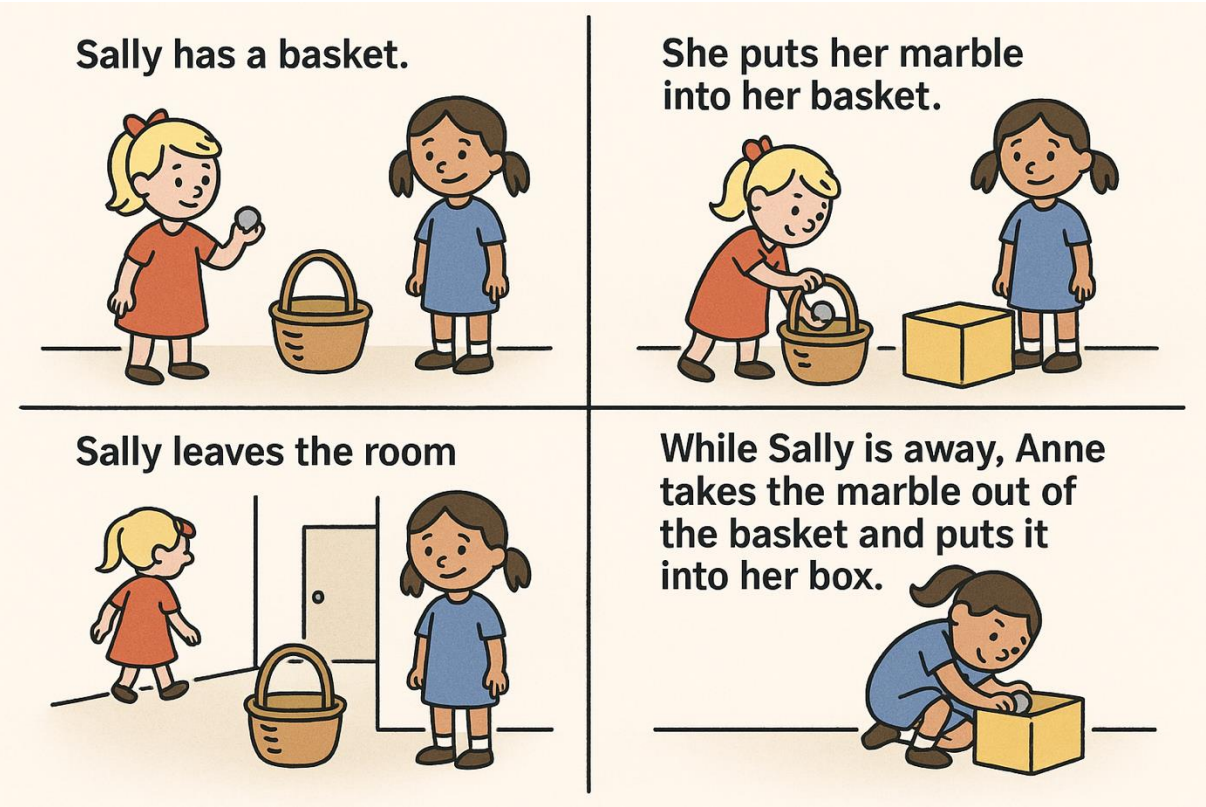
- Definition: Theory of mind is the **ability to understand that other people have their own thoughts**, beliefs, and emotions, which may differ from your own. It is central to social interaction and empathy.
- Note: Children typically begin developing theory of mind between ages 3 and 5, often demonstrated through tasks that test whether they can recognize false beliefs.

WHY IS TOM IMPORTANT?

- **Predicting behavior:** to anticipate what someone will do based on what we think they know or believe.
- **Perspective-taking:** Theory of mind helps us recognize that others don't share all of our knowledge or experiences, and can act differently, especially if they see the situation from a different angle.
- **Inferring intentions:** to interpret the motives behind people's actions.



False belief experiment

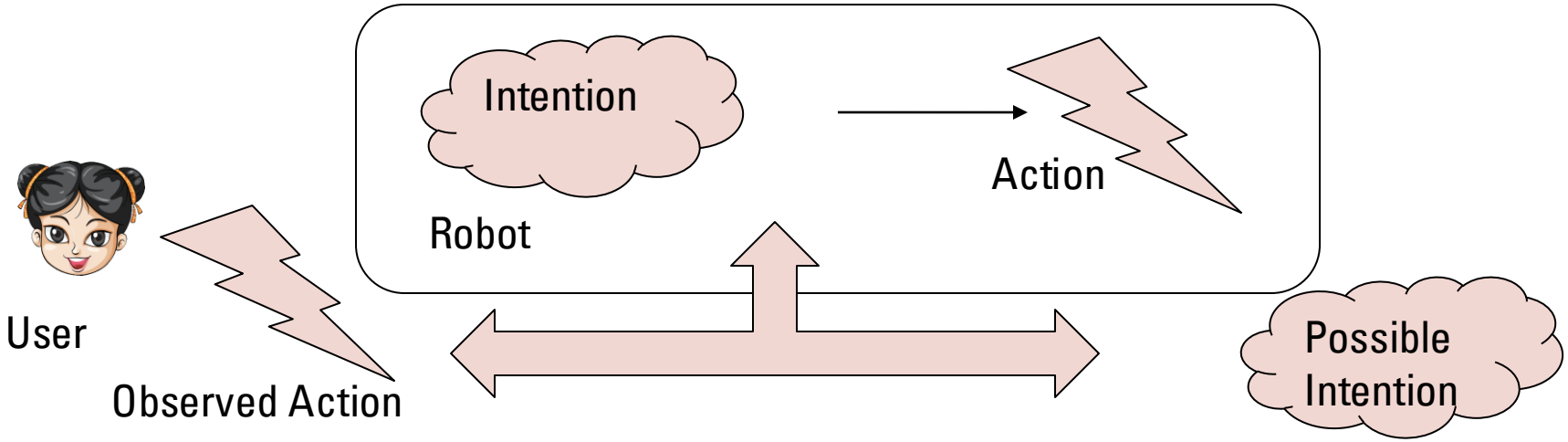


ChatGPT5.1

Gemini

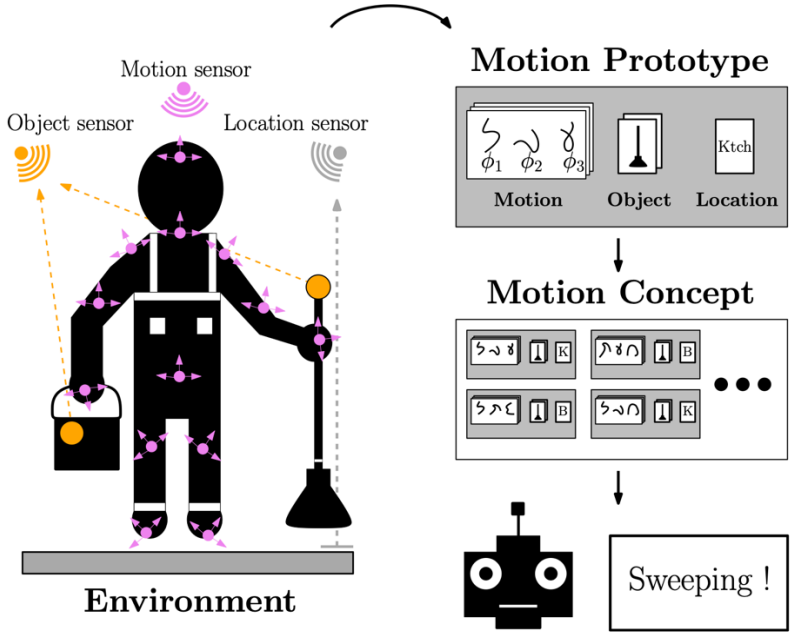
THEORY OF MIND

USING TOM FOR INTENTION RECOGNITION



- Simulation theory - people attribute mental states using their own mental processes (create a model of the other)
- Using simulation with states derived from taking the perspective of another person make inference about the other's actions and intentions

HUMAN ACTIONS – MULTIMODAL APPROACH



Vasco, Miguel, Francisco S. Melo, David Martins de Matos, Ana Paiva, and Tetsunari Inamura. "Learning multimodal representations for sample-efficient recognition of human actions." In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4288-4293. IEEE, 2019.

LMMS- LARGE MULTIMODAL MODELS

Gemini 3 Pro presents true visual and spatial reasoning with state-of-the-art performance across document, spatial, screen and video understanding (see Google Gemini 3 Pro site)

Human Actions

- The primary, cooperative action being performed is making a bed, specifically tossing or spreading a duvet/comforter over the mattress.
- Person on the Left (Male): He is holding one side of the large blanket with both hands, lifting it up and forward over the bed.
- Person on the Right (Female): She is mirrored, holding the opposite side of the blanket with both hands, coordinating her lift and spread motion with the other person.
- Cooperation: The action clearly demonstrates a cooperative task requiring synchronized timing and effort.

Human Posture

- Both individuals display engaged and open postures, focused entirely on the task and each other.
- Person on the Left:
- Torso/Hips: Facing the bed.
- Arms: Extended forward and slightly upward, holding the blanket taut.
- Stance: Standing close to the edge of the bed.

LMMS- LARGE MULTIMODAL MODELS

Expression Observed	Inferred Emotion	Detail/Interpretation
Wide, Open Smile	**Joy / Happiness**	The smile is clear, genuine (suggesting Duchenne quality, though not fully confirmable from the photo), and indicates a strong, positive internal state.
Direct Gaze	**Engagement / Affection**	His eyes are focused entirely on the woman on the right, suggesting he is fully present and positively interacting with his partner.
Body Posture	**Contentment**	The relaxed but engaged posture suggests ease and comfort in the activity and the shared space.

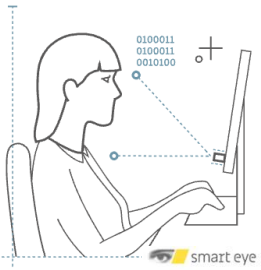
Gaze Analysis

Individual	Gaze Direction	Interpretation
Person on the Left	Directed fully toward the **Person on the Right**.	This indicates focused attention and active engagement with the other individual. The gaze is part of a feedback loop reinforcing the shared moment and positive emotion.
Person on the Right	Directed fully toward the **Person on the Left**.	This mirrors the other person's attention. She is receiving his gaze and reciprocating it, confirming the mutual nature of the interaction.

Gemini 3 Pro presents true visual and spatial reasoning with state-of-the-art performance across document, spatial, screen and video understanding (see Google Gemini 3 Pro site)

Timestamp	Observed Expression/Action	Inferred Emotion	Analysis
00:00 - 00:03	Small, Closed-Mouth Smile. Head slightly tilted. Eyes wide and engaging the camera.	Pleasure / Amusement	The child initiates the video with a gentle, self-conscious smile, suggesting they are aware of being filmed and are pleased or amused by the attention.
00:03 - 00:05	Smile Broadens. The head tilts more noticeably. The smile becomes more intense, showing teeth.	Joy / Playfulness	The emotion escalates. The broader smile and the exaggerated head tilt indicate a peak of positive, playful emotion, likely in response to something heard or encouraged by the person filming.
00:05 - 00:09	Arms Raised, Posture Changes. The smile disappears entirely. The eyes shift downward, and the lips purse or are drawn slightly inward.	Self-Consciousness / Confusion / Neutrality	This is a distinct shift. The child performs an action (pulling down their shirt) and their expression becomes neutral, or possibly slightly confused or self-monitoring, suggesting a brief internal shift away from the playful interaction.
00:09 - 00:15	Expression Returns to Smile/Grin. The mouth slowly curls into a pronounced smile, showing teeth, and the eyes brighten again.	Happiness / Re-engagement	The child re-engages with the camera, returning to the earlier state of positive emotion, indicating the brief moment of distraction or confusion has passed.
00:15 - 00:17	Laughter / Open-Mouth Smile. The mouth opens wide, often associated with a spontaneous vocalization (laughter, although the video is silent).	Peak Joy / Excitement	This is the highest intensity positive emotion in the clip, suggesting something highly stimulating or funny occurred just off-camera.
00:17 - End	The sequence repeats the shifts between smiling/joy, and brief moments of self-monitoring/neutrality.	Varying Joy and Engagement	The later segment (starting around 00:31) shows strong, broad smiles interspersed with playful actions like pursing the lips and looking down, characteristic of a child actively engaging with an adult.

Affectiva Emotion Sensing AI

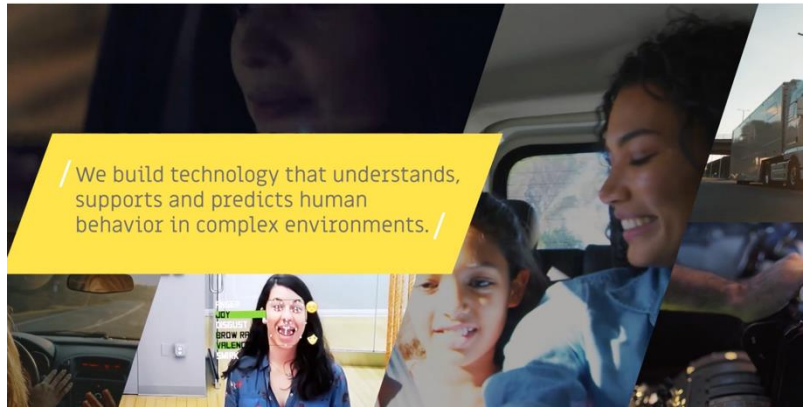


/ Welcome to the world of Affectiva emotion sensing AI

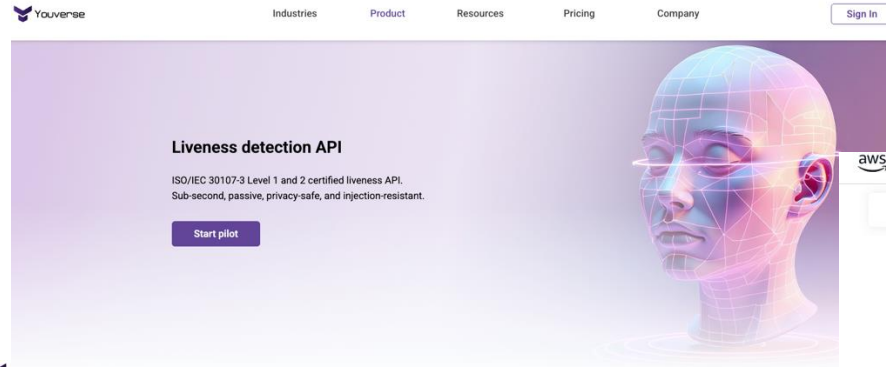
Facial coding is a key technology in automotive interior sensing, as it can give nuanced insight into what's happening with people in a vehicle. The data enables car manufacturers to improve safety and deliver personalized mobility experiences that enhance comfort, wellness, and entertainment.

A pioneer of emotion sensing AI, Affectiva's core technology offers a robust and diverse database of 15M+ faces and 8B+ facial frames from 90 different countries. Our emotion sensing AI solutions offer 20+ facial classifiers to accurately detect emotions and cognitive states from the face. For CES 2026, we have simplified our demo to show attendees a selection of our metrics. As you interact with the camera and make

Human Insight AI Solutions Technology News & Buzz About us Investors Smart Eye at CES 2026



Youverse Industries Product Resources Pricing Company Sign In



Why most liveness checks fail

Most vendors stop at Level 1 spoofing tests, which only catch printed photos or simple masks. Today's attackers use deepfakes and injection techniques that bypass weak solutions.

Youverse achieves ISO 27001 certification, strengthening its commitment to world-class information security

Lisbon, Portugal   3 December 2025 Youverse proudly announces that it has achieved ISO/IEC 27001 certification, the globally recognized standard for information security ma...

Written on 02 Dec 2025

Youverse launches Age Estimation and Liveness Detection APIs for secure, frictionless digital trust

Lisbon, November 3rd, 2025 Youverse announced the launch of two groundbreaking solutions: its Age Estimation API and Liveness Detection API. Designed to meet the rising demands of regulators, busines...

Written on 31 Oct 2025

Youverse named finalist for Finnovate Awards 2025 in Best ID Management/KYC Solution category

Youverse has been named a finalist for the 2025 Finnovate Awards in the Best ID Management/KYC Solution category. This recognition highlights Youverse’s pioneering work in transforming digital...

Written on 01 Aug 2025

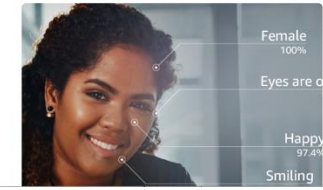
aws re:Invent Discover AWS Products Solutions Pricing Resources Search Sign in to console Create account

Amazon Rekognition Overview Use cases Features Pricing Resources FAQs Customers

Why Amazon Rekognition?

Learn how Amazon Rekognition can help your business and development teams to solve your most pressing computer vision needs—with no ML skills required and at a lower cost.

Features



Our Technologies

Facial Recognition | Human Body Recognition | Image Beautify | Image Recognition

Face Detection

Detect and locate human faces within an image, and returns high-precision face bounding boxes.Face** also allows you to store metadata of each detected face for future use.

Face Comparing

Check the likelihood that two faces belong to the same person. You will get a confidence score and thresholds to evaluate the similarity.

Face Searching

Find similar-looking faces to a new face, from a given collection of faces. Face***s fast and accurate search returns a collection of similar faces, along with confidence score and thresholds to evaluate the similarity.

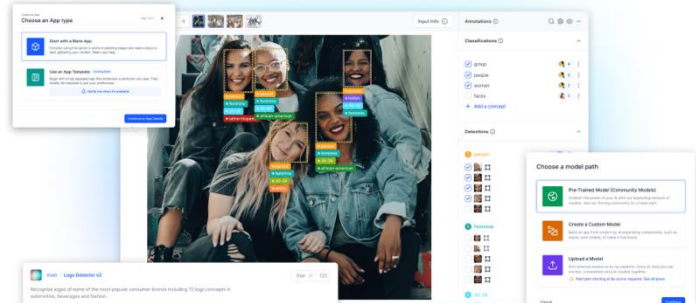
LISBOA

Why Platform Solutions Company Developers Pricing

Computer Vision

Deep learning AI models that provide human-like interpretation of video, image, text, and audio data

Request a Demo



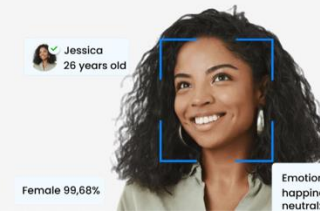
LUXAND Cloud API FaceSDK Industries API Pricing Docs Contact Us ES Sign in Try for f

Try the Free Plan. No Credit Card Required. Sign Up Now

iBeta GDPR COMPLIANT NIST

AI-Powered Face Recognition API

Integrate **Face Recognition** using our cloud API. Detect and compare human faces. Identify previously tagged people in images. Recognize age, gender, and emotions in the photo.




CURRENT CHALLENGE...DATA

Exponential growth of datapoints used to train notable AI systems

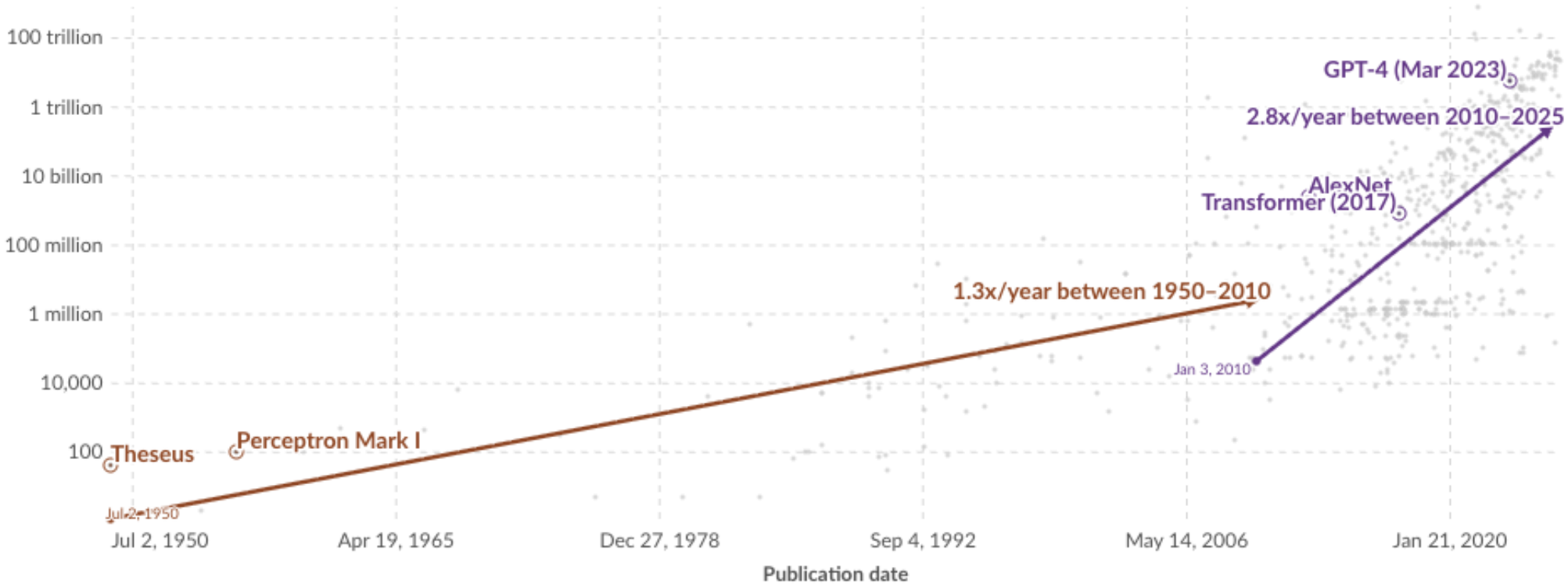
Our World in Data

The number of unique data points used to train the model. Each domain has a specific data point unit; for example, for vision it is images, for language it is words, and for games it is timesteps. This means systems can only be compared directly within the same domain.

Table | Chart

Settings

Training datapoints (unique datapoints; plotted on a logarithmic axis)



Play time-lapse | Jul 2, 1950 | Oct 10, 2025

Data source: Epoch AI (2025) – Learn more about this data

OurWorldinData.org/artificial-intelligence | CC BY

Note: The regression lines show a sharp rise in data used to train AI systems since 2010, driven by the success of deep learning methods that leverage neural networks and massive datasets.

TÉCNICO LISBOA

DATA..
DATA…
DATA…
…THE CASE OF NEO

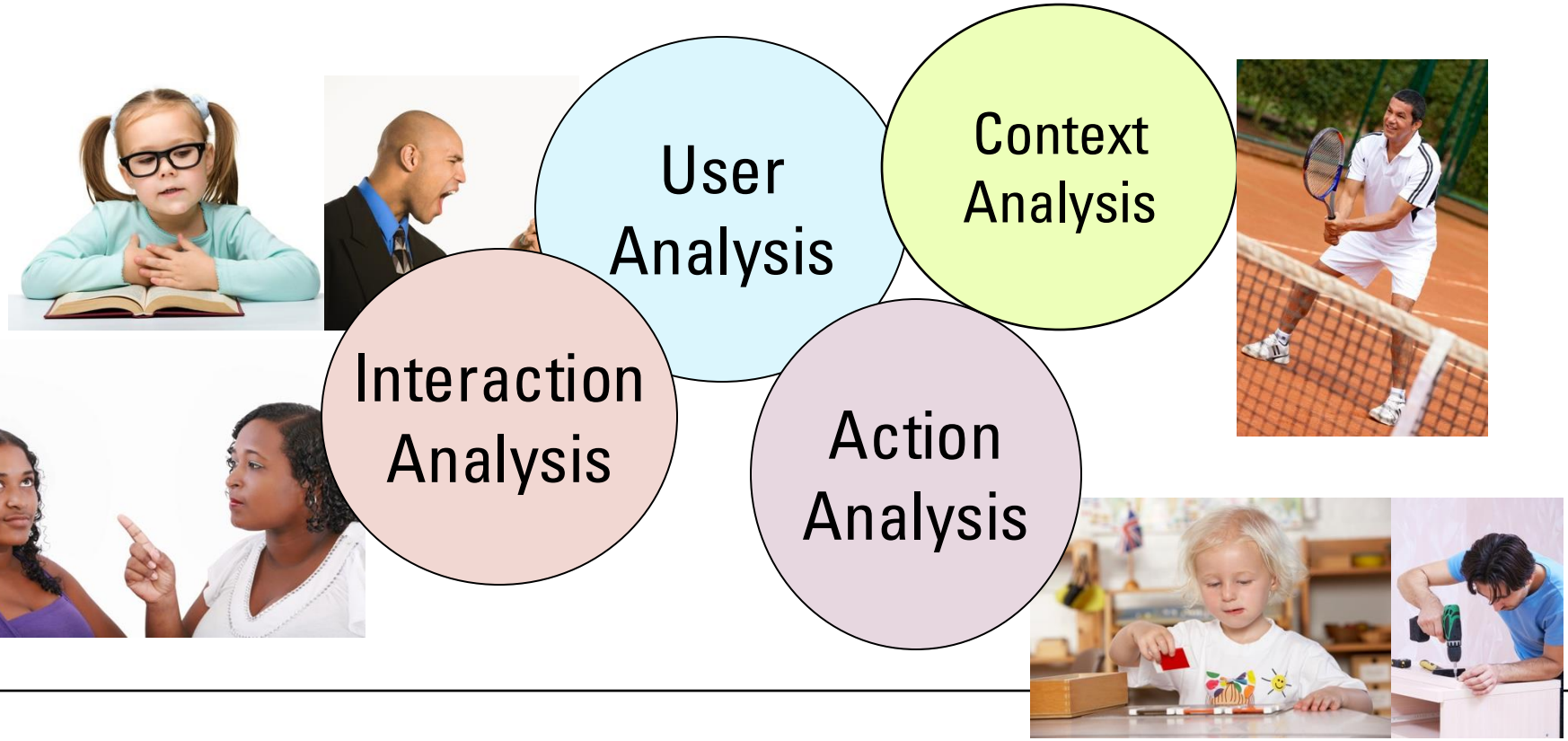
Current approaches: Build a large enough dataset that will be used for robots to recognise and able to do actions in the world

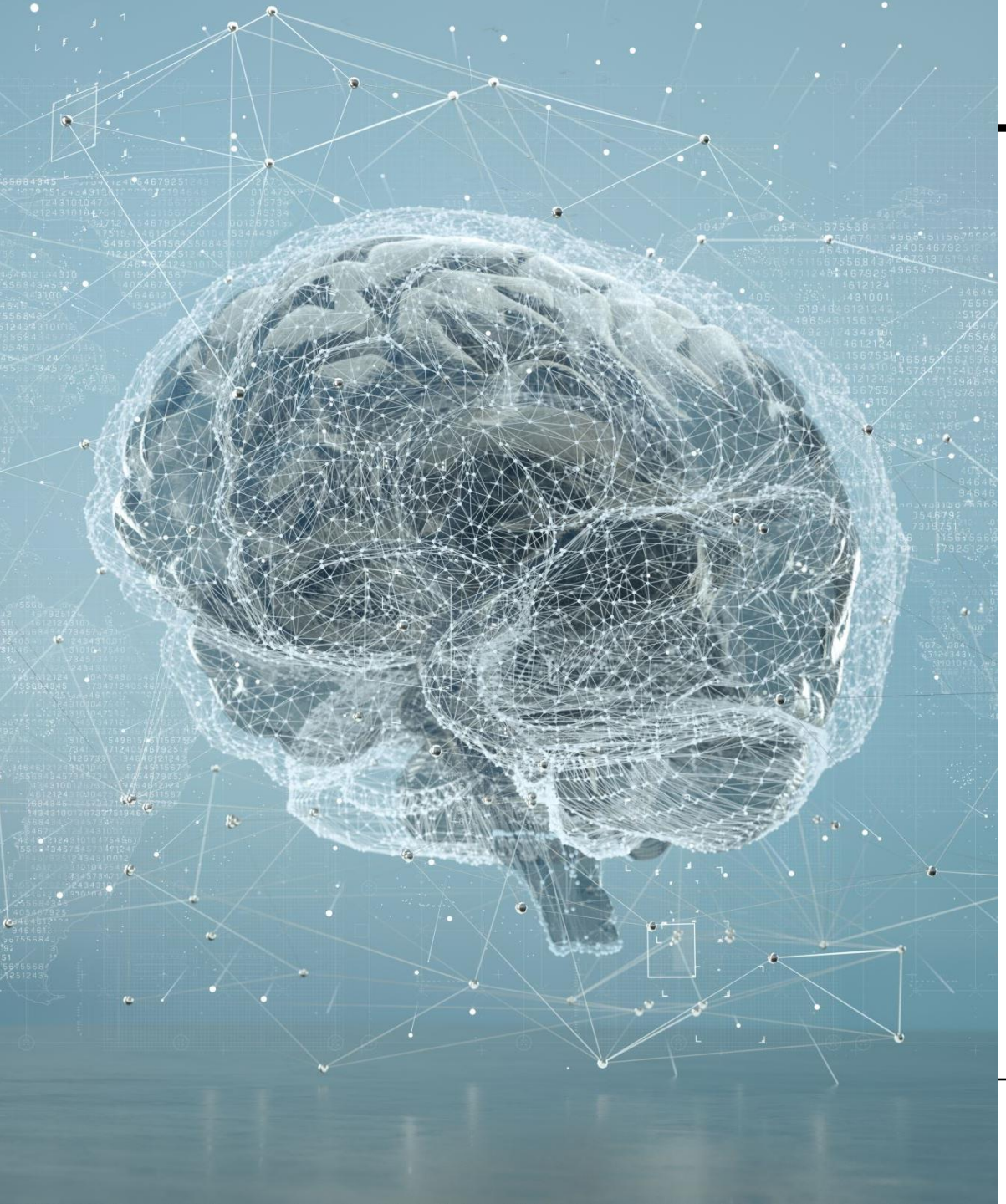
- https://www.youtube.com/watch?v=f3c4mQty_so

DATA..
DATA...
DATA...

SOCIAL PERCEPTION

HOW TO GO FROM SENSORS TO UNDERSTANDING OF INDIVIDUALS (HUMANS) AND GROUPS IN A SOCIAL CONTEXT





THE PROBLEM STILL TO BE SOLVED

Based on the limited perceptual capabilities of a robot, how to build technology to understand the social situation and the user's (and other agents') affective, social, motivational and informational states, in order to respond in a socially appropriate manner.

PAPERS/BOOKS TO READ

- *Vasco, Miguel, Francisco S. Melo, David Martins de Matos, Ana Paiva, and Tetsunari Inamura. "Learning multimodal representations for sample-efficient recognition of human actions." In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4288-4293. IEEE, 2019.*
- *Mataric, M. J. (2007). The robotics primer. MIT press.*
- *Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and vision computing, 27(12), 1743-1759.*
- *Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.*
- *Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.*
- *Estrela, Margarida, H2R: Robotic Failures and Social Exclusion in Human-Human-Robot Settings, MSc thesis, MEIC, 2025.*

2016

Towards Multi-Modal Intention Interfaces for Human-Robot Co-Manipulation

Luka Peternel, Nikos Tsagarakis and Arash Ajoudani

HRI^2 lab of Advanced Robotics department
Istituto Italiano di Tecnologia, Genoa, Italy

https://www.youtube.com/watch?v=e3t5odKe6_c

EMBODIMENT: OTHER SENSORS

- **Proximity and Motion (LiDAR, Ultrasonic, IR sensors)**
  - **Cues:** Distance, approaching/speeding away, personal space intrusion.
  - **Techniques:** Tracking movement patterns, social distance (Hall's proxemics).
  - **Example:** A robot adjusts its distance to maintain a comfortable personal space.
- **Other Sensors**
  - **Thermal cameras:** Detect stress or attention (via skin temperature).
  - **Galvanic Skin Response (GSR), Heart Rate monitors:** For affective computing (but less common in robots)