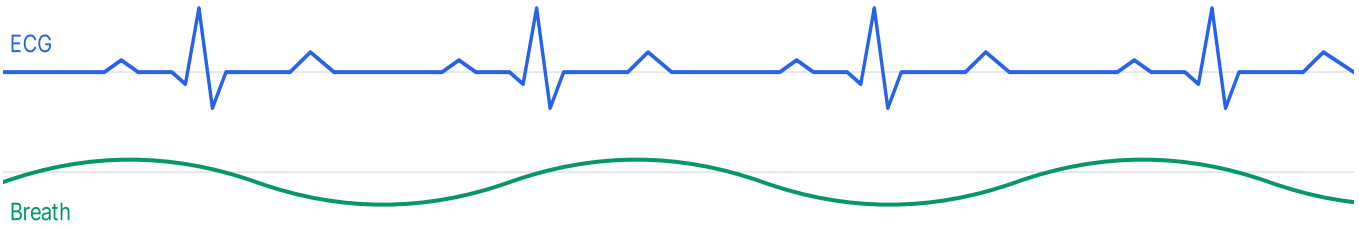
RESEARCH REPORT · NOVEMBER 2026

The Human Twin Report

What has been done to copy a human with AI and sensors, how each piece works, and what could be done next. From cameras and wearables to personality models, memory, voice, robots and brain interfaces.



12
chapters

10
layers of a person

20+
kinds of sensor

1 to 20 yrs
roadmap horizon

HOW TO READ THIS

Chapters 1 to 3 set the goal and the raw data. Chapters 4 to 9 cover each part of a person and how AI copies it. Chapter 10 looks at copying the brain itself. Chapter 11 is a build plan you can follow. Chapter 12 covers what comes next and the hard questions. A companion report, "What Makes Us Human", covers the human science with no AI.

CHAPTER 01

What Does It Mean to Clone a Human?

A clone of a person is not one thing. It is a stack of copies. Each copy captures a different layer of who you are and each layer is far harder than the one below it.

People use the word "clone" loosely. A voice that sounds like your mother is called a clone. So is a chatbot trained on your old emails. So is a 3D avatar that moves like you on a video call. These are real products today. But none of them is a copy of a person in the full sense. Each one copies a surface. The question this chapter asks is simple. What exactly are we trying to copy and how would we know if we got it right?

The honest starting point is this. A human is a living body that has run without a break since birth. It has a brain of about 86 billion neurons. It keeps itself alive by balancing heat, water, sugar and salt every second. It remembers. It forgets. It changes. A digital twin can only copy what we can measure. So the real design question is not "can we copy a human?" It is "which parts can we measure well enough to copy and which parts are still out of reach?"

A ladder of fidelity

The clearest way to think about this is a ladder. Each rung is a layer of the person. Low rungs are easy to see and easy to copy. High rungs are hidden and hard. A good twin climbs as high as the data allows. A weak twin stays near the bottom and fakes the rest.

L1 **Look**

Face shape, skin, hair, body size, clothes. Copied from photos and video. A few minutes of video can drive a talking head today.

L2 **Voice**

Pitch, timbre, accent, pace. Modern voice cloning works from seconds to minutes of clean audio. It is the most mature rung after look.

L3 **Knowledge**

Facts the person knows. Their job, their family, their past. Copied by putting their documents and transcripts in a searchable store that a language model can read.

L4 **Style of talking**

Favorite words, jokes, how long their sentences are, how they open and close a message. Copied by fine tuning or long prompts with many real examples.

L5 **Personality**

Stable traits such as the Big Five. How open, careful, social, warm or anxious the person is across many situations.

L6 **Decisions**

What they would choose when it matters. How they trade money against time, risk against safety, self against others.

L7 **Emotional patterns**

What makes them angry, what calms them, how fast they recover. How mood shifts with sleep, hunger and stress.

L8 **Memories**

Lived episodes with time, place, people and feeling. Not facts about the past but the past as they hold it, with its gaps and errors.

L9 **Body physiology**

Heart rhythm, breathing, temperature, glucose, sleep, hormones. The body that shapes the mind from below.

L10 **Neural level**

The wiring and activity of the brain itself. The only rung that could in theory produce all the others from the inside.

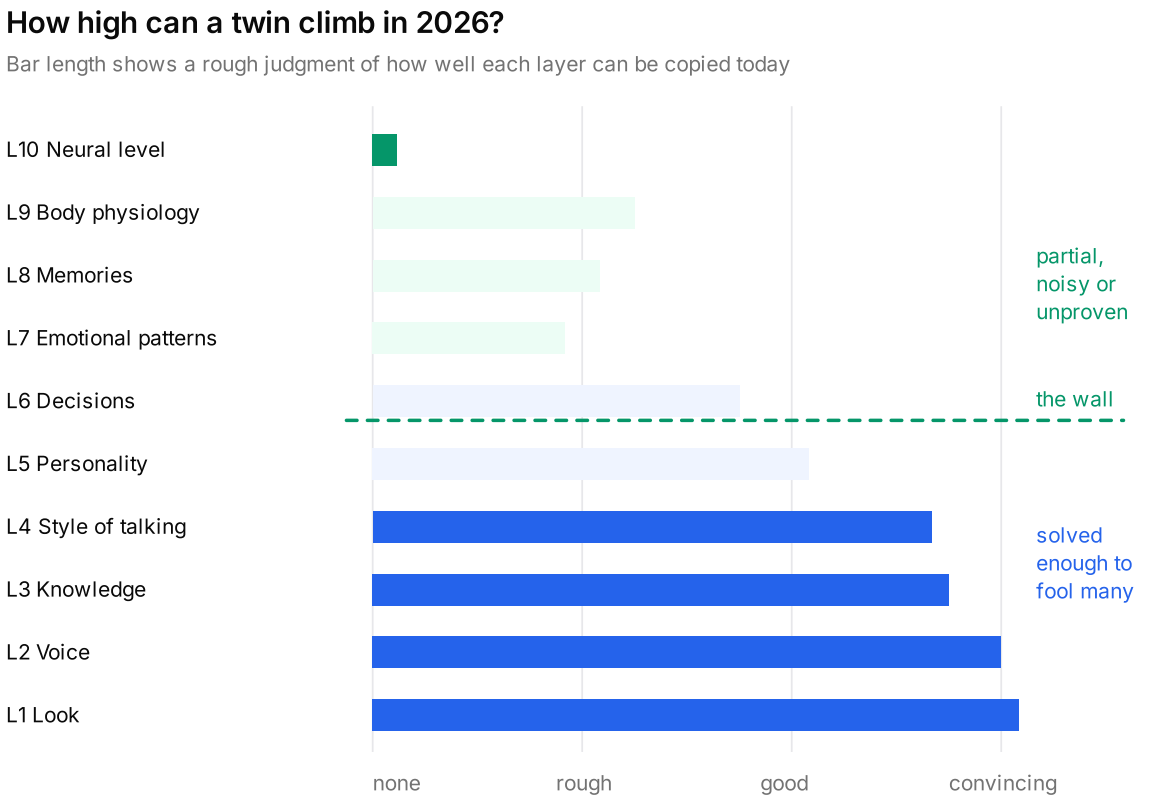


FIG 1.1 The fidelity ladder. Surface layers (L1 to L4) can already be copied well enough to fool strangers. Above personality the evidence thins fast. The dashed line marks where today's systems stop being measured copies and start being guesses. Bar lengths are the authors' judgment from the evidence in Chapters 4 to 10, not a measured score.

Notice the shape. The bottom four rungs are mostly solved for practical use. They are also the rungs that fool people fastest, because humans judge each other first by face, voice and words. That is a trap. A twin that looks and sounds right feels deep even when it is shallow. Most of the work in this report is about the rungs above the line.

Weak copy and strong copy

It helps to name two ends of a spectrum.

A **weak copy** reproduces outputs the person has already produced. Ask it about their wedding and it retells the story they wrote down. Ask it something new and it falls back on a generic answer in their voice. A weak copy is a mirror of the archive. It cannot surprise you in a way that is still true to the person.

A **strong copy** produces new outputs the person would have produced in situations they never faced. Ask it how they would react to a job offer in another country and it gives the answer the real person would give, with the same hesitation and the same reasons. A strong copy has a model of the person that generalizes. It does not just remember. It predicts.

Almost everything built so far is a weak copy with a strong copy's face. That is not an insult. A weak copy can be very useful for memory, for grief, for answering questions on someone's behalf. But it should be sold as what it is.

Three kinds of twin

Researchers in AI and philosophy of mind tend to sort copies by what they share with the original. Three kinds matter for this report.

BEHAVIORAL TWIN

Matches what the person says and does. It is judged only from the outside. Inside it can be anything: a language model, a lookup table, a puppet. Today's persona chatbots and avatar clones are behavioral twins.

FUNCTIONAL TWIN

Matches the inner causes too. It has states that play the same roles as the person's beliefs, goals, moods and memories. When it gets tired or scared, those states shift its choices the way they shift the person's.

EMULATION

Copies the physical machinery, usually the brain, at a level of detail where behavior falls out on its own. Nobody has done this for any animal brain beyond very small ones. Chapter 10 covers the progress.

WHY THE SPLIT MATTERS

A behavioral twin can pass a short test and fail a long one. A functional twin should hold up under new situations. An emulation should hold up everywhere, but may need data we cannot yet collect.

Same input, three kinds of inside

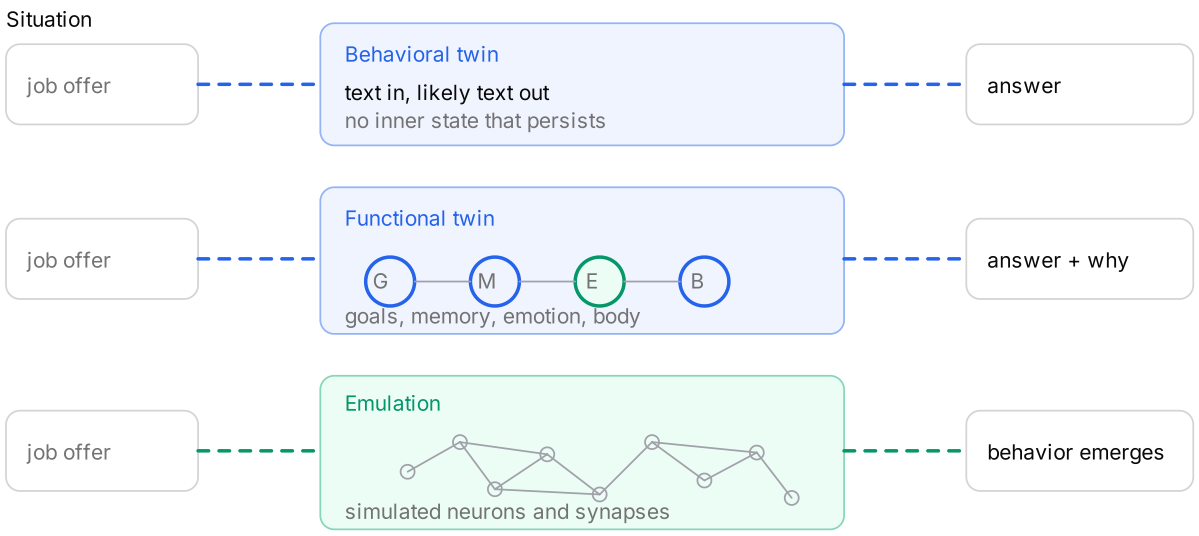


FIG 1.2 A behavioral twin maps input to output with no lasting inner life. A functional twin routes the input through states that stand in for goals (G), memories (M), emotions (E) and body (B). An emulation runs the physical wiring and lets behavior emerge. All three can give the same answer to one question. They differ on the thousandth question.

A practical builder in 2026 should aim for a behavioral twin that slowly becomes functional. You start with a language model that talks like the person. Then you add explicit state: a memory store, a mood variable driven by sensor data, a set of stated values and goals. Each added state is a bet that the person's behavior depends on it. Chapters 5, 6 and 9 show how.

What separates a person from a chatbot today

Language models now write in a person's style well enough to fool friends in short exchanges. So what is still missing? The gap is not mostly about words. It is about the conditions under which words are made.

BODY

EMBODIMENT

You see from one place. You get tired. Your hands shake when nervous. A chatbot has no place, no fatigue and no hands. Many human choices are shaped by what the body can and cannot do right now.

BODY

A BODY TO PROTECT

Life is homeostasis. The brain spends much of its effort keeping blood sugar, heat and fluid in range. Hunger, pain and fear are signals from that work. A chatbot has nothing it must keep alive.

TIME

CONTINUOUS MEMORY

You wake up as the same person who went to sleep. A language model starts each chat from frozen weights plus whatever text you paste in. Memory systems help but they are bolted on.

INNER LIFE

REAL FEELINGS?

Nobody knows if a model feels anything. Most researchers think current models do not, but there is no agreed test. A twin can show emotion. Whether it has emotion is an open question. Chapter 9 goes deeper.

TIME

GROWTH

People change their minds, learn skills, heal and age. A twin frozen at one date drifts away from the living person every day. Updating it without breaking it is unsolved.

SOCIAL

RELATIONSHIPS

Who you are depends on who you are with. You speak one way to your boss and another way to your child. A twin needs a model of each relationship, not just of you.

NOISE

UNPREDICTABILITY

People are inconsistent. Ask the same person the same survey twice and they change about one answer in five. A perfect twin should be just as noisy, in the same places.

Closed loop vs open loop

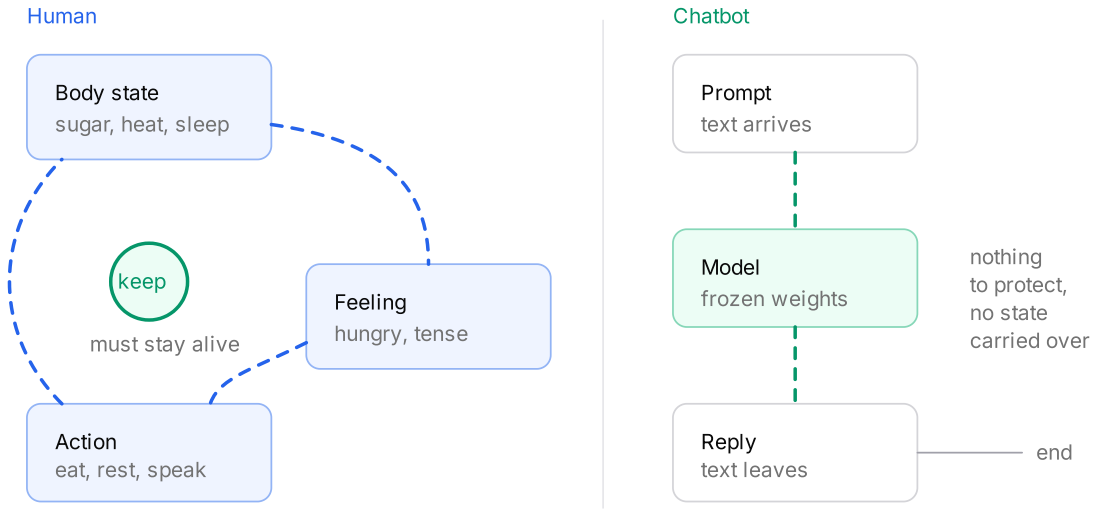


FIG 1.3 A person runs a closed loop. Body state creates feelings. Feelings drive actions. Actions change body state. The loop never stops and the stakes are survival. A plain chatbot runs an open line from prompt to reply. Wearable data (Chapter 2) and internal drives (Chapter 9) are two ways to start closing the loop for a twin.

This is the deep reason the sensor stack matters for a twin. Physiology is not decoration. A person who slept four hours makes different choices. A person whose blood sugar is falling is shorter with others. If the twin knows nothing about the body it cannot copy the patterns that come from the body. It will be a well rested, well fed, calm version of the person all the time. Friends notice that quickly.

How would we test a clone?

You cannot improve what you cannot measure. A twin needs tests that are hard to game and that match what you care about. No single test is enough. Here are the main families.

TURING TEST VARIANTS

The classic Turing test asks if a judge can tell a machine from a human in a text chat. In 2025 Cameron Jones and Benjamin Bergen at UC San Diego ran a careful three party version. A judge chatted with a human and an AI at the same time for five minutes and picked which was human. GPT-4.5 with a prompt telling it to act as a certain kind of young person was picked as the human 73 percent of the time. That is more often than the real humans were picked. Without the persona prompt it won only 36 percent of the time.

This tells us two things. First, passing for "a human" in a short chat is now easy. Second, the persona did most of the work. That is good news for twins and also a warning. Fooling a stranger proves almost nothing about whether the twin is a copy of one specific person. The test a twin needs is harder: can a judge tell *this* human from *this person's twin*?

THE FRIENDS AND FAMILY TEST

The strongest judges are the people who know the person best. A useful protocol: the real person and the twin each answer the same set of new questions. Close friends, a partner and family members see both answers side by side, blind, and guess which is real. If they cannot do better than chance over many rounds, the twin passes at that depth. Run it on text first, then voice, then video. Each channel adds cues judges can use.

Two cautions. Close judges have strong priors and may reject a twin because it is "too polite" even when the real person is polite that day. And the test must use new questions. If the twin has seen the answer in its training data it is reciting, not predicting.

PREDICTION ACCURACY OF CHOICES

The most scientific test is to predict what the person will do before they do it. Ask the twin how the person will answer a survey, play an economic game or rate a product. Then ask the person. Score the match.

The best public example so far is a 2024 Stanford study led by Joon Sung Park. The team ran two hour voice interviews with 1,052 Americans. They gave each full transcript to a language model agent. Then they asked the agents to answer the General Social Survey, a Big Five personality test and several economic games, and compared the answers to the real people's.

The clever part is the ceiling. People do not answer the same survey the same way twice. So the team had each person retake the survey two weeks later. On the General Social Survey the people matched their own earlier answers 81.25 percent of the time. The agents matched 68.85 percent of the time. Divide one by the other and the agents reached 85 percent of human self consistency. Agents built only from demographic facts reached 71 percent.

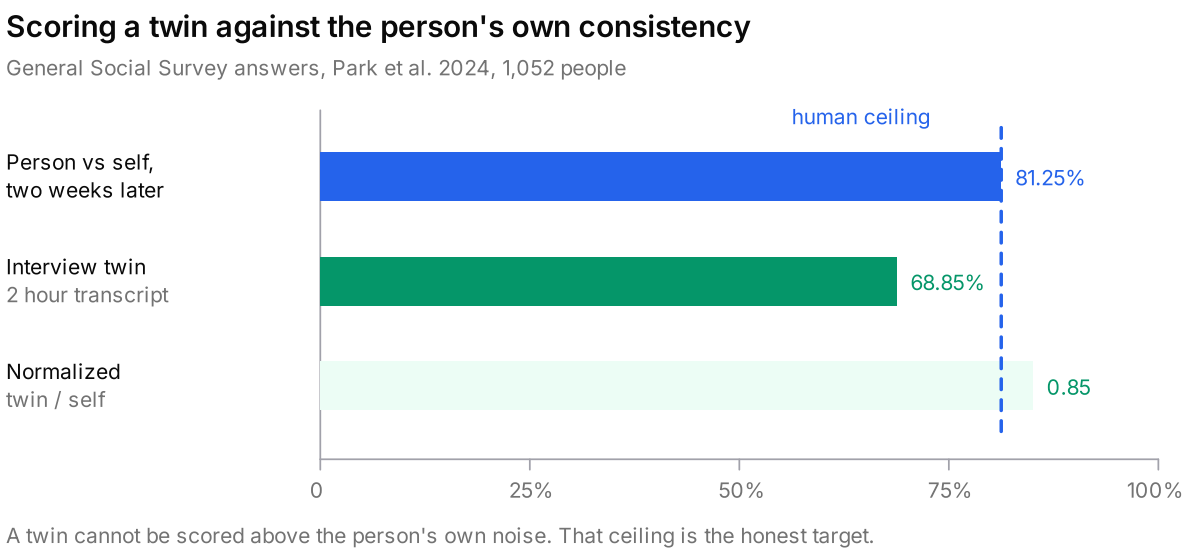


FIG 1.4 Normalized accuracy. The real person sets the ceiling by repeating the test. The twin's raw match is divided by that ceiling. On the same scale the interview agents reached 0.80 for Big Five traits and 0.66 for economic game behavior, so decisions under stakes were the weakest layer. Source: Park et al., arXiv 2411.10109.

This study is the closest thing we have to a benchmark for a personality twin. It also shows the limits. The agents were built from one interview. They predicted survey answers, not real life choices. And the games were simple. Still, the method is the right one: always measure the person against themselves first.

SELF RECOGNITION

The last test is the most personal. Show the person their twin's answers without telling them which are theirs. Can they pick out their own? Do they say "that is exactly what I would say" or "I would never put it that way"? Self recognition catches errors that outsiders miss, such as values the person holds privately. It also has a bias. People often see themselves as more consistent and kinder than they are. A twin that matches the self image may not match the behavior.

TEST	WHAT IT CHECKS	STRENGTH	WEAKNESS
Three party Turing test	Passes as some human	Easy to run at scale	Says little about one specific person
Friends and family blind test	Passes as this person to people who know them	Hardest judges available	Small sample, judge bias, needs fresh questions
Choice prediction	Predicts surveys, games and real decisions	Objective score with a human ceiling	Lab choices are not life choices
Self recognition	Person accepts the twin's answers as their own	Catches private values	Self image is not behavior
Longitudinal drift	Twin stays accurate as months pass	Tests growth and memory	Slow and costly to run
Physiological match	Twin's predicted stress, sleep and heart rate match sensors	Grounds emotion in the body	Body signals are noisy and depend on context

A PRACTICAL SCORECARD

Run all six tests at the start and every three months. Report each one separately. A twin that scores well on Turing style tests but poorly on choice prediction is a good mask, not a good copy.

Honest status in late 2026

Here is a plain summary of where things stand. It is meant to set expectations for the rest of the report.

POSSIBLE NOW

- A photoreal talking avatar of you from a few minutes of video.
- A voice clone that most strangers cannot tell from you.
- A chatbot that knows your life facts from your files and talks in your style.
- Personality and survey prediction at roughly 80 to 85 percent of your own consistency, from a long interview.
- Continuous body data: heart rate, HRV, sleep, skin temperature, glucose, movement and voice, with consumer devices.
- Rough stress and arousal estimates from those signals.

NOT POSSIBLE YET

- Reliable reading of specific emotions from the body alone.
- Faithful copying of episodic memories as the person holds them.
- Strong prediction of high stakes life decisions.
- A twin that grows and changes in step with the person over years.
- Any proof that a twin has feelings or experience.
- Copying a human brain at the neural level.

So the honest answer to "can we clone a human?" is: we can clone the parts of a human that leave a strong trace in data. Face, voice and words leave strong traces. Personality leaves a medium trace. Emotion, memory and body leave weak, noisy traces that need many sensors over a long time. The brain leaves almost no trace that we can read from the outside.

That is why the next two chapters are about input. Chapter 2 lists every sensor you could put on or around a person and what it can tell you. Chapter 3 looks at people who already tried to record their whole lives and what they learned. The quality of any twin is capped by the quality of the data that went into it.

SOURCES

- Park et al. 2024. Generative Agent Simulations of 1,000 People. arXiv 2411.10109
- Stanford HAI. Simulating Human Behavior with AI Agents (policy brief)
- Jones and Bergen 2025. Large Language Models Pass the Turing Test. arXiv 2503.23674
- Turing 1950. Computing Machinery and Intelligence. Mind
- Sandberg and Bostrom 2008. Whole Brain Emulation: A Roadmap. Future of Humanity Institute
- Herculano-Houzel 2009. The human brain in numbers. Frontiers in Human Neuroscience

- Barrett and Simmons 2015. Interoceptive predictions in the brain. Nature Reviews Neuroscience

CHAPTER 02

The Sensor Stack

Every sensor is a narrow window. A camera sees the face but not the heart. A ring feels the pulse but not the words. A twin is built by stacking windows until the gaps between them get small.

The previous chapter ended on a hard rule. A twin can only copy what leaves a trace in data. This chapter is a field guide to the traces. It covers every kind of sensor you could put on a person, near a person or in their pocket. For each one it asks four questions. How does it work? How much data does it make? How accurate is it? And does it tell you about the body, the mind or both?

One idea runs through the whole chapter. Sensors do not measure emotions or thoughts. They measure light, voltage, pressure, motion, sound and chemistry. Everything else is inference. A wrist device does not "detect stress". It detects that your skin sweat glands opened and your heart sped up. That could be fear. It could also be a hot room, a hill or a coffee. The twin has to learn which, and usually it needs other sensors to tell.

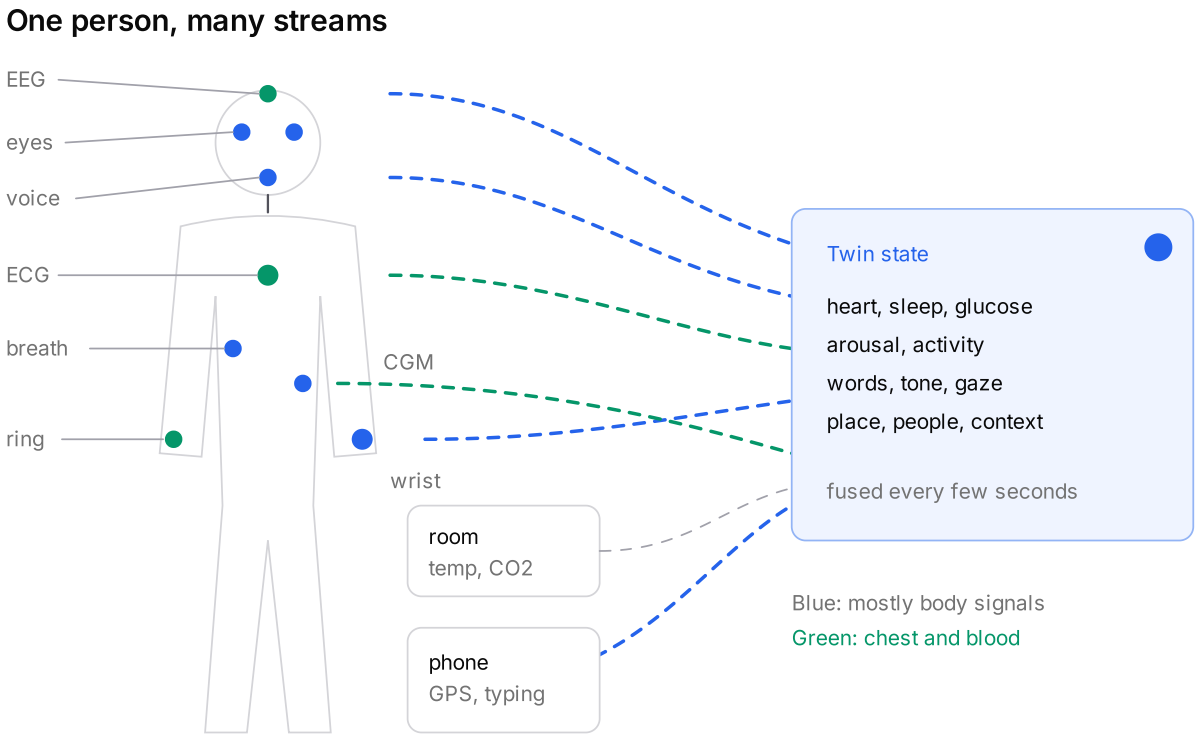


FIG 2.1 A full sensor stack. Head sensors (EEG, eye cameras, microphone), chest ECG and breathing, a ring or wrist band, a glucose patch, a phone and room sensors all feed one time aligned record. The twin does not read any one stream. It reads them together, because each signal is ambiguous alone.

How a body signal becomes data

Most wearable signals follow the same path. A physical change happens in the body. A transducer turns it into a voltage. An analog to digital converter samples that voltage many times a second. Firmware filters out noise, finds features such as heartbeats and often throws the raw samples away. What reaches your phone is usually a summary: one heart rate per second, one HRV value per night, one glucose value every few minutes.

This matters a lot for a twin. Raw data lets you compute new features later. Summaries lock you into the vendor's choices. If you want a serious twin, prefer devices that give access to raw or near raw data, even if the app is less pretty. Research devices like the Polar H10 strap, Empatica wristbands and Muse headbands have developer kits for this reason. Consumer devices such as Apple Watch, Oura and Whoop mostly export summaries through their apps and health platforms.

Cameras

A camera pointed at a person is the richest single sensor and the most invasive. Software can extract a lot from it.

- **Face.** Face mesh models track hundreds of landmark points per frame. From those you get head pose, blinks, mouth shape and facial action units, the muscle movements coded by Ekman and Friesen's Facial Action Coding System.
- **Micro expressions.** Very brief expressions that last a fraction of a second. They are real but rare, and claims that software can read hidden lies from them are not well supported. Treat them as a research curiosity, not a lie detector.
- **Posture and gesture.** Body pose models find the joints of the skeleton in each frame. Slumping, fidgeting, pacing and leaning in all show up.
- **Gaze.** A plain webcam can estimate roughly where someone looks. A dedicated eye tracker does it far better.
- **Heart rate from video.** This one surprises people. It is called remote photoplethysmography or rPPG.

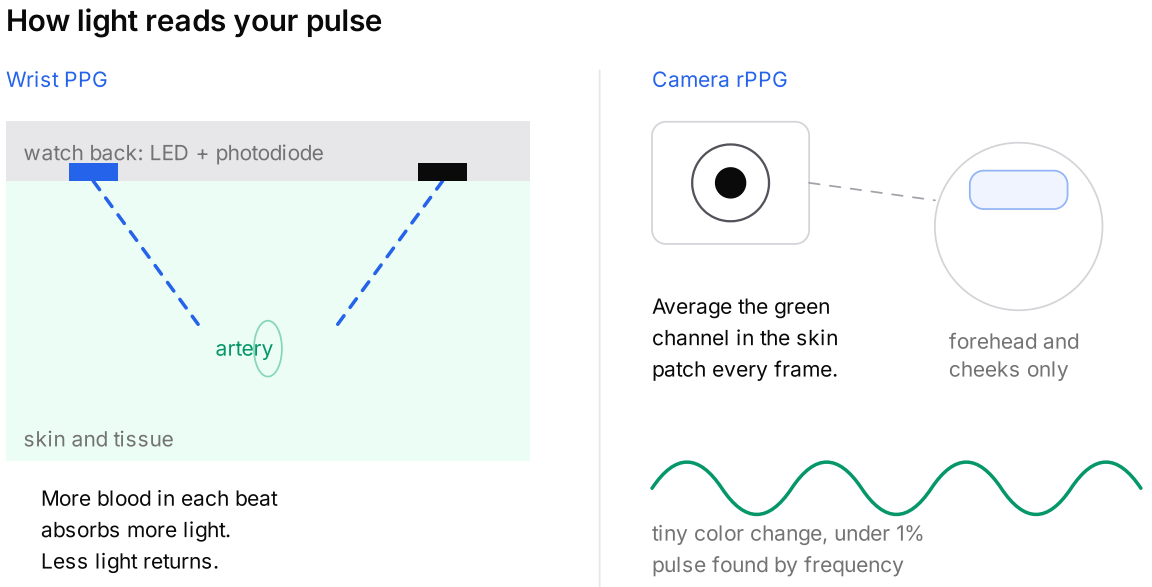


FIG 2.2 Photoplethysmography (PPG). On the wrist an LED shines into the skin and a photodiode measures what comes back. Each heartbeat pushes more blood into the vessels, which absorbs more light, so the return dips with every pulse. A camera can see the same tiny color change on facial skin. It is much weaker, so motion, lighting and makeup all compete with it.

rPPG works at rest in good light. Accuracy drops with head motion, low light, darker skin in some older methods and high heart rates. A 2025 study from Bielefeld University found that AI based rPPG lost accuracy sharply at elevated heart rates. For a twin, rPPG is a bonus channel during video calls, not a replacement for a wearable.

Data rate: a 1080p stream at 30 frames per second compresses to roughly 2 to 8 megabits per second. That is about 1 to 4 gigabytes per hour. Most twin pipelines should not store raw video at all. They should run face and pose models on the device and store the extracted features, which are thousands of times smaller.

Microphones

Sound is the best value sensor for a twin. It is cheap, it is already in every phone and it captures the main way people express their minds: speech. A microphone gives you four layers.

- **Words.** Speech recognition turns audio into text. On clean speech modern systems get most words right. Noisy rooms, crosstalk and accents still cause errors.
- **Prosody.** Pitch, loudness, speed and pauses. Prosody carries much of how a person feels. Fast, loud, high speech usually means high arousal.
- **Vocal emotion.** Models trained on acted or labeled speech estimate arousal fairly well. Valence, whether a feeling is good or bad, is much harder from voice alone.
- **Non speech sounds.** Laughter, sighs, coughs, breathing and crying. These are strong cues and easy to detect.

Raw audio at 16 kHz and 16 bits is 256 kilobits per second. A speech codec such as Opus brings that to about 16 to 32 kilobits per second with little loss for transcription. That is under 15 megabytes per hour. Chapter 3 shows why audio is the sweet spot for life recording.

Privacy note: a microphone records everyone near you. Many places require consent from all parties to record a conversation. This is a legal problem for any always on twin, not just an ethical one.

Eye tracking

Eye trackers shine infrared light at the eye and film the reflection with a camera. They find the pupil center and a bright spot on the cornea, then compute gaze direction. Tobii is the best known maker. Its consumer Eye Tracker 5 for PC gaming sits on a monitor. Research grade trackers sample from 60 to over 1,000 times a second and cost thousands of dollars.

Gaze shows what someone attends to. Pupil size is a window on arousal and mental effort: the pupil widens a little when a task gets hard. The catch is light. The pupil also reacts to brightness, and that effect is far bigger. You need to control for screen and room light before you read anything mental into pupil size.

Wearables on the wrist and finger

This is where most continuous body data comes from today. Apple Watch, Oura Ring, Whoop and Empatica bands all use some mix of the same sensors.

PPG FOR HEART RATE AND HRV

The optical sensor shown in Fig 2.2 gives heart rate. The timing between beats gives heart rate variability, or HRV. HRV is one of the most useful signals for a twin. When the body is calm and recovered, beat to beat timing varies more. Under stress, illness, alcohol or poor sleep it varies less. Wrist and finger PPG is accurate for heart rate at rest and in sleep. It gets worse during hard exercise because motion corrupts the light signal. That is why many devices report HRV only at night.

EDA, SKIN CONDUCTANCE

Electrodermal activity measures how well the skin conducts a tiny current. Sweat glands are controlled by the sympathetic nervous system, the "fight or flight" side. When arousal rises, glands open a little and conductance jumps within a second or two. Then it slowly falls. These spikes are called skin conductance responses.

EDA is the cleanest single marker of arousal you can wear. But it is blind to valence. A scare, a joke and a kiss can all make the same spike. Empatica's EmbracePlus band uses EDA and motion in its FDA cleared EpiMonitor system to flag possible tonic clonic seizures. Some Fitbit and Samsung watches also measure EDA for stress features.

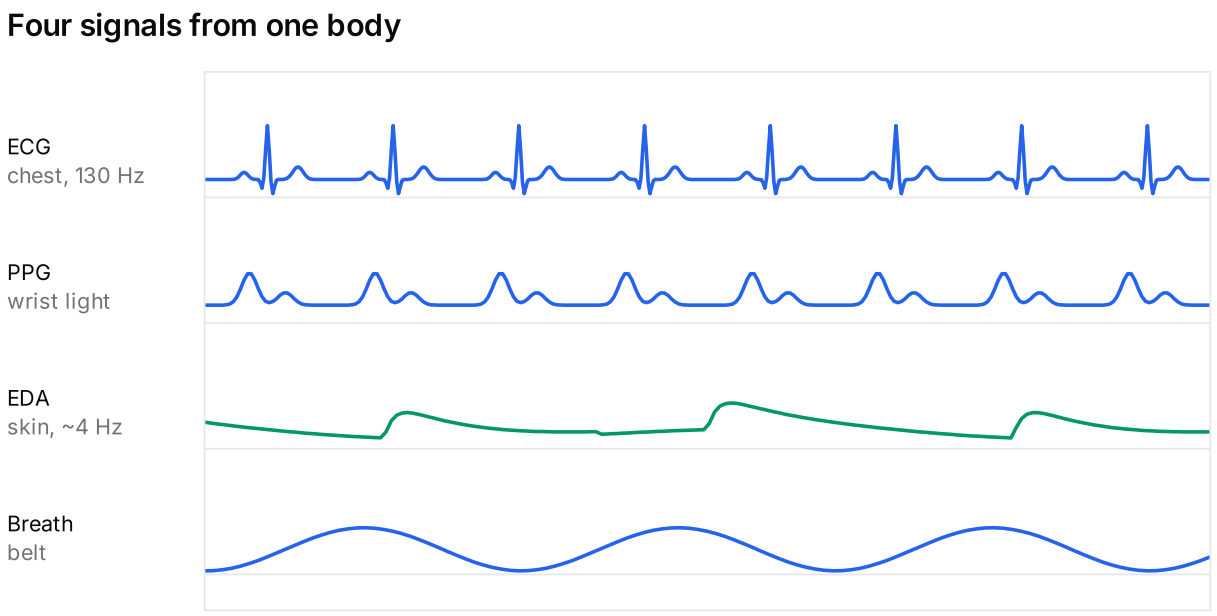


FIG 2.3 The same moments seen four ways. The ECG shows each electrical heartbeat as a sharp spike. The PPG shows the blood pulse a fraction of a second later as a smooth wave. EDA moves slowly and jumps when arousal rises, then decays over many seconds. Breathing rises and falls a few times a minute. Traces are drawn as schematics, not recordings.

SKIN TEMPERATURE

Rings and watches track skin temperature mostly during sleep, when it is stable. They report the change from your own baseline, not your true core temperature. Oura and Apple Watch use this to flag possible illness and to estimate menstrual cycle phase. For a twin, a rise in night temperature is an early hint that the person may soon feel unwell and act differently.

ACCELEROMETER AND GYROSCOPE

Motion sensors sample at roughly 25 to 100 times per second. They give steps, activity type, sleep and wake, restlessness and hand gestures. They are also the main tool for cleaning other signals: if the wrist is moving, PPG and EDA readings are suspect.

SPO2

Blood oxygen uses red and infrared light. Oxygen rich blood absorbs the two colors differently. Consumer devices measure it mostly at night to look for breathing dips. Accuracy is a few percentage points at best, and the FDA has warned that pulse oximeters can be less accurate on darker skin. For a twin, SpO2 matters mainly for sleep quality.

WATCH

APPLE WATCH

PPG, ECG (single lead via the crown), wrist temperature, SpO2 in some regions, motion. In September 2025 Apple added hypertension notifications, cleared by the FDA, that look for patterns over 30 days of PPG data. It does not give a blood pressure number.

RING

OURA RING 4

Finger PPG, skin temperature, motion and SpO2. Strong at sleep, HRV and temperature trends because the finger has good blood flow and less motion at night. Needs a monthly subscription for most features.

BAND

WHOOP 5.0 AND MG

Screenless band with PPG, temperature and motion. The MG model added an ECG and a "Blood Pressure Insights" feature in May 2025. The FDA sent Whoop a warning letter about that feature in July 2025, a sign of how blurry the line is between wellness and medicine.

RESEARCH

EMPATICA EMBRACEPLUS

Wrist band with EDA, PPG, temperature and motion. Used in research and as part of the prescription EpiMonitor seizure alert system. One of the few wearables built around EDA.

Chest, brain and muscle

ECG CHEST STRAPS

An electrocardiogram measures the heart's electrical signal directly through electrodes on the skin. A chest strap such as the Polar H10 does this far more precisely than wrist light. Polar's developer kit exposes the raw ECG at 130 samples per second and the exact time between beats. For HRV under motion, a chest strap is the consumer gold standard. It is also uncomfortable to wear all day, so most people use it for workouts, sessions or sleep experiments.

EEG HEADBANDS

EEG measures tiny voltages on the scalp produced by millions of neurons firing in sync. Consumer headbands use a handful of dry electrodes. The Muse S Athena, launched in 2025 at about $475, adds fNIRS, which uses light to track blood oxygen in the front of the brain. Emotiv sells 5 channel and 14 channel headsets used in many research labs.

EEG can track sleep stages, drowsiness, eyes closed relaxation and broad attention states. It cannot read thoughts. Signals are tiny and easily swamped by blinking, jaw clenching and movement. A 4 channel headband sees the brain the way a microphone outside a stadium hears a match: you can tell when the crowd roars but not what anyone says. Chapter 10 covers implanted interfaces, which are a different league.

EMG

Electromyography reads the electrical signals that drive muscles. In 2025 it reached consumers in a big way. Meta's Neural Band, sold with the Meta Ray-Ban Display glasses from September 30, 2025 for $799, reads wrist muscle signals to detect finger gestures, even very small ones. For a twin, EMG points to two things: fine control intent at the wrist, and in research setups, facial muscle activity for smiles and frowns that a camera might miss.

RESPIRATION BELTS

A stretchy band around the chest or belly measures breathing depth and rate. Research suppliers sell them, and smart shirts such as Hexoskin build them in. Breathing is a strong emotion cue. Sighs, held breaths and fast shallow breathing all mean something. Breathing can also be estimated from PPG, ECG and even audio, with less detail.

Chemistry: continuous glucose

A continuous glucose monitor, or CGM, puts a thin filament just under the skin of the arm. It measures glucose in the fluid between cells every few minutes. Dexcom's Stelo became the first over the counter CGM in the United States in August 2024. Each sensor lasts up to 15 days and costs roughly $50 at list price. Abbott sells a similar over the counter product called Lingo.

Glucose is a rare direct window into the body's energy state. Falling glucose is linked to irritability and poorer focus in many people. For a twin, CGM data helps explain why the same person is patient at 11 am and snappy at 1 pm. Note the lag: interstitial glucose trails blood glucose by several minutes.

Smartphone: digital phenotyping

Your phone may be the best psychological sensor you already own. Researchers call it digital phenotyping: inferring behavior and mental state from passive phone data. Signals include:

- **Typing.** Speed, errors, pauses and backspaces. Typing rhythm changes with fatigue and mood.
- **App use and screen time.** When the phone is used, for what and how often it is picked up.
- **Location.** GPS gives places visited, time at home and how varied the day is. Low variety of places has been linked to depressive symptoms in several studies.
- **Communication.** Call and message counts and, with consent, the content itself. Messages are the single best source for a person's written style.

The data is cheap and huge. But platform limits on iOS and Android block much passive collection by third party apps, and the content is deeply private. The best route for a personal twin is usually your own data exports from Apple, Google and messaging apps.

Environment sensors

Context shapes behavior. Room sensors log temperature, humidity, light, noise and CO2 about once a minute. A cheap CO2 monitor tells you how stuffy a room is. Some studies have found poorer decision scores at high indoor CO2 levels, though results are mixed. Room data also helps explain body data. A rise in skin conductance in a hot room means something different from the same rise in a cool one.

Smart glasses: first person everything

Glasses with cameras and microphones put sensors exactly where your attention is. Ray-Ban Meta glasses capture first person photos, video and audio and run an AI assistant. The 2025 Display model adds a small screen in one lens. For a twin this is the richest context source possible: what you looked at, who you talked to, what you read. It is also the most socially risky. Everyone in view is recorded, and a small LED is a weak signal of consent. Chapter 3 returns to this.

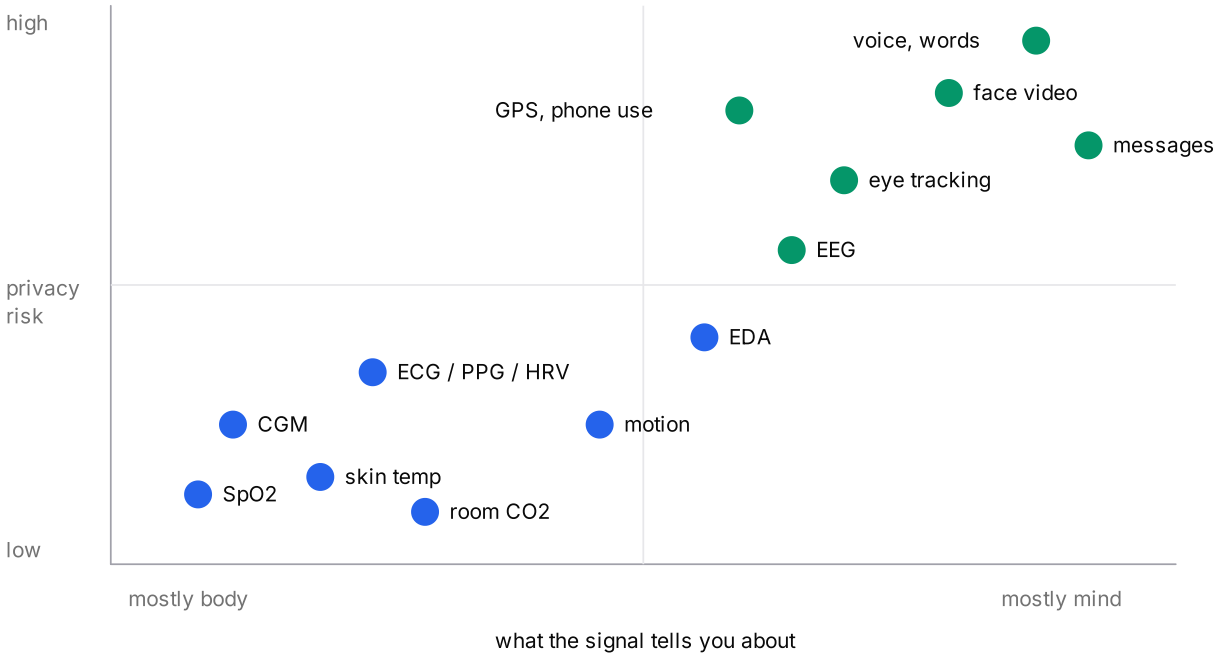


FIG 2.4 A rough map. Body sensors at the lower left are cheap in privacy and tell you about physiology. Sensors at the upper right reveal thoughts and social life but also record other people and intimate content. A twin needs both corners. Positions are the authors' judgment, not measurements.

The big table

SENSOR	EXAMPLE PRODUCTS	RAW RATE	TYPICAL COST	ACCURACY	BODY OR MIND	F
Camera (face, pose)	Webcam, phone	2 to 8 Mbit/s compressed 1080p	$0 to $200	Landmarks good. Emotion labels weak.	Expression, attention, posture	V h r c
rPPG	Any camera + software	~1 HR value/s	Software	OK at rest. Poor with motion or high HR.	Body	H
Microphone	Phone, pendant	256 kbit/s raw, 16 to 32 kbit/s coded	$0 to $200	Words good in quiet. Arousal fair. Valence weak.	Mind and social	V h r c
Eye tracker	Tobii	30 to 1,200 Hz	$200 to $10,000+	Good when calibrated	Attention, effort	M
Wrist / ring PPG	Apple Watch, Oura, Whoop	25 to 256 Hz internal	$200 to $800 + subs	HR good at rest. HRV best at night.	Body, stress	L r
EDA	Empatica EmbracePlus	~4 Hz	Research pricing	Good for arousal. Blind to valence.	Arousal	L

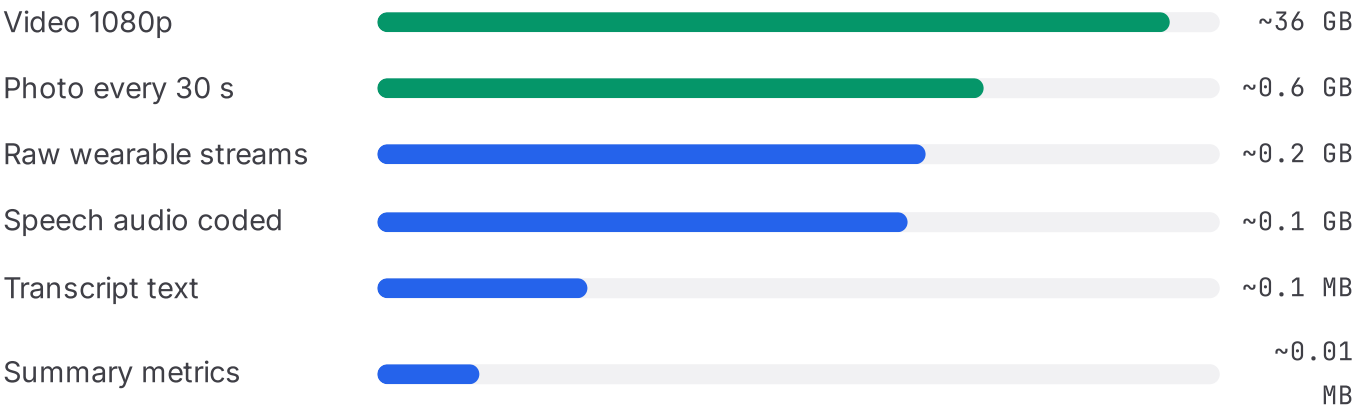
SENSOR	EXAMPLE PRODUCTS	RAW RATE	TYPICAL COST	ACCURACY	BODY OR MIND	F
Skin temperature	Oura, Apple Watch	Per minute, used nightly	Included	Good for change from baseline	Body	L
Accelerometer / gyro	All wearables, phone	25 to 100 Hz	Included	Good for activity and sleep	Behavior	L
SpO2	Watch, ring	Spot or nightly	Included	A few points. Bias on darker skin.	Body	L
EEG headband	Muse S Athena, Emotiv	128 to 256 Hz per channel	$300 to $1,000+	Sleep and broad states. Noisy.	Mind, coarse	M t
EMG	Meta Neural Band	Hundreds to thousands Hz	Bundled at $799	Good for gestures	Motor intent	M
ECG chest strap	Polar H10	130 Hz ECG	~$90 to $100	Excellent for HR and HRV	Body, stress	L
CGM	Dexcom Stelo, Abbott Lingo	One value per 5 to 15 min	~$50 per 15 days	Good. Lags blood by minutes.	Energy state	L r
Respiration belt	Hexoskin, lab belts	25 to 100 Hz	$100 to $600	Good when snug	Arousal, speech breathing	L
Smartphone	iOS, Android	Events, GPS per minute	Owned	Varies by platform	Behavior, mood,	V h

SENSOR	EXAMPLE PRODUCTS	RAW RATE	TYPICAL COST	ACCURACY	BODY OR MIND	F
				limits	social	
Environment	CO2, light, noise monitors	Per minute	$50 to $250	Good	Context	L
Smart glasses	Ray-Ban Meta, Meta Ray-Ban Display	Photo, video, audio on demand	$300 to $800	Good capture	Context, attention, social	V h r c

Prices and rates are approximate and change often. Where a rate is not published by the maker it is a typical value for that sensor type.

How much data per day?

Here is a rough daily budget for a serious personal stack worn for 16 waking hours plus sleep. Video dominates everything by orders of magnitude, which is why a smart pipeline stores features, not footage.



Bars on a log scale from 1 KB to 100 GB. Raw wearable estimate assumes about ten channels at about 100 samples per second at 2 bytes each.

What the stack can and cannot tell you

Put it all together and you get a clear split.

STRONG SIGNALS

- Sleep timing and quality
- Activity and rest
- Physiological arousal and recovery
- Energy state from glucose
- Where you were and who you talked to
- What you said and how loudly or fast

WEAK OR MISSING SIGNALS

- Which emotion you feel, beyond calm vs aroused
- What you are thinking
- Why you made a choice
- Private feelings you never express
- Memories you never talk about

The right column is the reason a twin always needs a second kind of input: the person talking about themselves. Interviews, diaries, ratings and conversations with the twin fill gaps that no sensor can. Chapter 4 explains how emotion models try to squeeze more from the left column. Chapter 11 shows a practical stack that balances cost, comfort and privacy.

PRIVACY IS A DESIGN INPUT

Body sensors mostly record you. Cameras, microphones and glasses record everyone around you. Laws on recording conversations differ by place, and many require all parties to agree. Health data from wearables may not be protected by medical privacy law when it sits with a consumer company. Process on the device where you can, store features instead of raw media, and encrypt everything.

SOURCES

- Apple. Hypertension Notifications validation paper (September 2025)
- FDA warning letter to Whoop, Inc. July 2025
- FDA 510(k) K232915. Empatica EpiMonitor using EmbracePlus
- Polar BLE SDK (H10 ECG at 130 Hz)
- Muse S Athena EEG and fNIRS headband
- MedTech Dive. Dexcom launches Stelo over the counter CGM
- UploadVR. Meta Ray-Ban Display and Neural Band announced
- News Medical. Accuracy of rPPG drops sharply at elevated heart rates (2025)
- MedTech Dive. FDA reviews how skin pigmentation affects pulse oximeter results

- Onnela and Rauch 2016. Harnessing smartphone based digital phenotyping. Neuropsychopharmacology

CHAPTER 03

Lifelogging and the Data of a Life

For eighty years people have tried to record everything they live. They learned that capture is the easy part. Finding, trusting and living with the record is hard.

A twin needs a past. The sensor stack in Chapter 2 tells you what a person's body is doing now. But who a person is comes from years of conversations, places, choices and moods. Lifelogging is the practice of recording that past as fully as possible. It is older than computers. And its pioneers left behind a rare gift: honest reports of what worked, what failed and how it felt to live inside a total record.

This chapter has three parts. First a short history, from a 1945 thought experiment to the AI pendants of 2024 and 2025 and what happened to them. Second, the math: how much data is a human life, and what does it cost to keep. Third, the lessons: how to store and search a life, and what lifeloggers say it did to them.

A short history of recording a life

1945

Vannevar Bush imagines the Memex

In "As We May Think" in The Atlantic, Bush describes a desk that stores all of a person's books, records and letters on microfilm. Its key idea is the "associative trail": links between items that mirror how the mind jumps from one thought to the next. It was never built, but it named the dream.

1980s to 1990s

Steve Mann wears the computer

Mann built wearable computers and head mounted cameras from his teens. At the MIT Media Lab in the mid 1990s he streamed his daily view to the web. He later coined "sousveillance", recording from below rather than being watched from above.

2001

MyLifeBits at Microsoft Research

Computer pioneer Gordon Bell, with Jim Gemmell, set out to store his whole life: papers, letters, email, photos, calls, web pages and later images from a wearable camera. The

project fed into the 2009 book Total Recall.

Early 2000s

SenseCam

Microsoft Research Cambridge built a camera worn on a lanyard that took a picture automatically, roughly every 30 seconds or when its light, heat and motion sensors noticed a change. It became the classic lifelogging research tool and was tested as a memory aid for people with amnesia.

2005 to 2008

Deb Roy's Human Speechome Project

Roy put cameras and microphones in the ceilings of his home and recorded his son from birth to about age three. The corpus holds about 90,000 hours of video and 140,000 hours of audio, recorded around 10 hours a day.

2007

The Quantified Self

Wired editors Gary Wolf and Kevin Kelly started a movement and meetups around "self knowledge through numbers". Members tracked sleep, mood, food, focus and much more, often with home made tools.

2013 to 2016

Clip cameras and the backlash

Clip on cameras such as the Narrative Clip brought SenseCam style capture to consumers. Google Glass met hostility in bars and cafes. By 2016 Bell had stopped wearing his camera.

2022 to 2025

The AI recorder wave

Large language models made raw recordings searchable and summarizable. A wave of products followed: Rewind and its Limitless pendant, the Humane AI Pin, the Rabbit R1, Plaud recorders, the Bee wristband, the Friend pendant and Meta's camera glasses.

2025

Consolidation

Humane shut down its pin and sold assets to HP in February. Amazon agreed to buy Bee in July. Meta launched display glasses in September and bought Limitless in December.

THE MEMEX IDEA THAT STILL MATTERS

Bush did not just want storage. He wanted trails. He saw that the mind does not file things in folders. It links them. One memory calls up another by association. Eighty years later this is still the core problem of a twin's memory. A pile of recordings is not a memory. Links between them are. Modern systems build those links with text embeddings, time stamps, people and places. Chapter 6 goes deeper.

STEVE MANN AND THE SOCIAL COST

Mann's long experiment showed early that the hardest problems are social. A camera on your face changes how people treat you. In 2012 he reported being physically confronted at a Paris McDonald's over his eyewear. His idea of sousveillance framed personal recording as a balance to state and corporate cameras. That debate is still open. Every always on twin raises it again.

GORDON BELL: THE MOST COMPLETE RECORD, AND ITS LIMITS

Bell was a famous computer architect who designed machines at Digital Equipment Corporation. From around 2001 he scanned his papers, saved his email and calls, logged the web pages he read and later wore a SenseCam. With Jim Gemmell he built MyLifeBits, a database meant to hold it all and make it searchable. He believed this record could become a kind of "digital immortality".

The project produced a lot of useful lessons. Storage was cheap even then. Capture could be automatic. But retrieval was the bottleneck. Search by text worked well for documents. Search over thousands of wearable camera images did not, because images had no words attached. Bell later said lifelogging "wasn't something that was bringing a lot of value to my life" and he had stopped wearing the camera by 2016. He died in May 2024 at 89. His archive remains one of the deepest personal records ever made.

DEB ROY: WHAT DENSE RECORDING CAN REVEAL

The Speechome Project is the best case for dense recording. Roy and his team transcribed and analyzed the audio to track every word his son learned and every time an adult used that word near him. The most striking finding: caregivers seemed to adjust their own speech around a word as the child was learning it. They used it in shorter, simpler phrases just before he first said it, then returned to normal. Nobody in the house noticed doing this. Only a near complete record could show it.

That is the lesson for a twin. Dense data reveals patterns the person cannot report. But it took a research lab, years of work and a lot of manual transcription to get there. Modern speech

recognition and language models cut that cost by orders of magnitude.

THE QUANTIFIED SELF: NUMBERS ABOUT YOURSELF

The Quantified Self community made tracking personal and practical. Members gave "show and tell" talks with three questions: what did you do, how did you do it and what did you learn? Many found that tracking itself changed behavior. Others found that numbers without a question led nowhere. The best projects started with a specific question, such as "does caffeine after 2 pm hurt my sleep?", and tracked just enough to answer it.

The AI recorder wave and what happened to it

Around 2022 something changed. Speech recognition got very good and language models could summarize hours of talk into a few lines. Suddenly recording everything had a payoff: you could ask questions of your past. A wave of devices followed. Their fates by late 2026 tell a story.

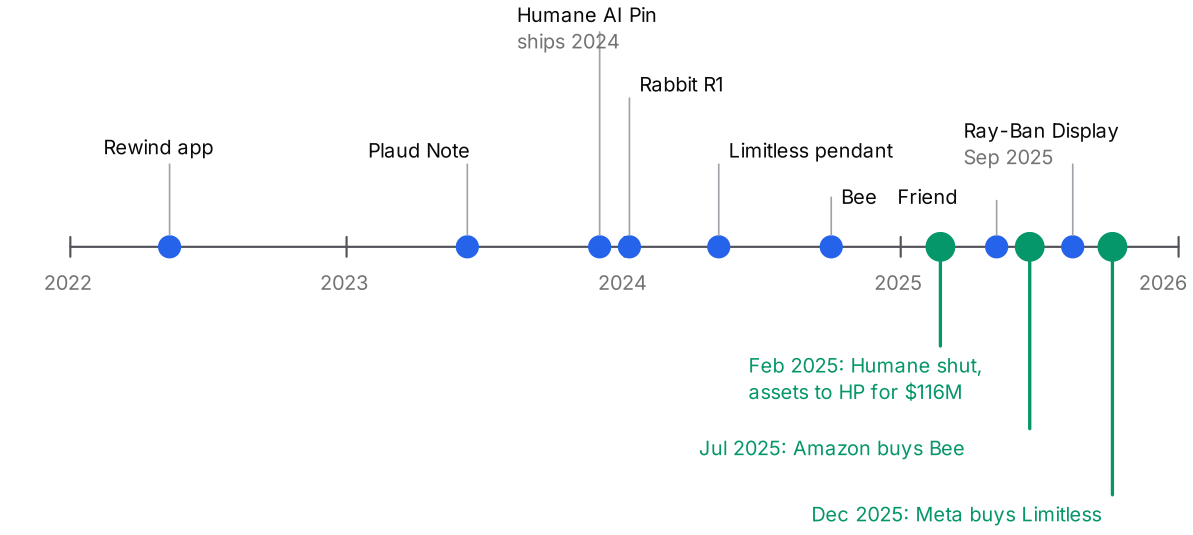


FIG 3.1 Within about three years most independent AI recorders were shut down or bought by large platforms. Plaud and Rabbit remained independent into 2026. The lesson for a twin builder: if your life record lives only in a startup's cloud, it can vanish or change owners with one press release.

ACQUIRED

REWIND AND LIMITLESS

Rewind recorded everything on a Mac screen and made it searchable. In 2024 the company renamed itself Limitless and sold a clip on pendant that recorded conversations and produced transcripts and summaries. Meta bought it in December 2025. New pendant sales stopped. Existing users got at least a year of support, and the Rewind desktop features were phased out.

SHUT DOWN

HUMANE AI PIN

A $699 pin with a camera, microphone and laser projector, meant to replace the phone. Reviews were harsh on heat, battery and speed. In February 2025 HP bought most of Humane's staff, software and patents for $116 million. The pins lost their cloud features on February 28, 2025 and stopped doing most things.

INDEPENDENT

RABBIT R1

A $199 handheld AI device launched at CES 2024 to heavy hype and poor early reviews. Rabbit kept shipping software, including rabbitOS 2 in 2025 and computer control features in 2026, and has teased a second device. There were also reports of unpaid staff in 2025. Its future is uncertain.

INDEPENDENT

PLAUD

Card and pin shaped voice recorders that send audio to an app for transcripts and summaries. Aimed at meetings rather than whole days. The company says it has more than 2.5 million users. That figure is self reported.

ACQUIRED

BEE

A wristband of about $50 that listens all day and turns conversations into summaries and to do lists. Bee said it kept transcripts rather than audio. Amazon agreed to buy it in July 2025, and the team joined Amazon Devices.

INDEPENDENT

FRIEND

A $129 pendant that listens and acts as an AI companion, texting you comments about your day. Its large 2025 New York subway ad campaign was widely defaced. Reviews were mostly negative and focused on privacy.

PLATFORM

META GLASSES

Ray-Ban Meta glasses capture first person photos, video and audio and run an assistant. The Meta Ray-Ban Display model, launched in September 2025 at $799, adds an in lens screen and the Neural Band. Combined with the Limitless purchase, Meta is clearly betting on always available personal AI.

LESSON FROM THE WAVE

Every one of these products depended on a company's servers. When Humane shut down, the device's memory went with it. A serious twin needs data you can export and keep yourself, in open formats, on storage you control.

How much data is a human life?

Now the math. Let us estimate the size of one life recorded at different levels of detail. The point is not precision. It is to see the orders of magnitude, because they decide what is practical.

16,000
words spoken per day on average (Mehl et al. 2007)

5,840
waking hours per year at 16 hours a day

~470M
spoken words in 80 years at that rate

~470K
waking hours in 80 years

The words figure comes from a 2007 study in Science by Matthias Mehl and colleagues. They gave 396 university students small recorders that sampled ambient sound for several days. Women spoke about 16,200 words a day and men about 15,700. The difference was not significant. The authors stressed there were very large differences between individuals. Your own number could be half or double.

Now apply standard storage assumptions. Text is about 6 bytes per word including spaces. Speech audio at 16 kilobits per second, a common rate for a speech codec, is 7.2 megabytes per hour. A photo every 30 seconds at about 300 kilobytes each. Video at 1080p and 5 megabits per second is 2.25 gigabytes per hour. Raw wearable data is about ten channels at about 100 samples per second, around the clock.

LAYER	ASSUMPTION	PER DAY	PER YEAR	80 YEARS
Spoken words as text	16,000 words × 6 bytes	96 KB	35 MB	2.8 GB
Search index	One 4 KB vector per 30 s of waking time	7.7 MB	2.8 GB	225 GB
Audio, all waking hours	16 kbit/s × 16 h	115 MB	42 GB	3.4 TB
Raw wearable streams	10 ch × 100 Hz × 2 bytes × 24 h	173 MB	63 GB	5 TB
Photo every 30 s	1,920 photos × 300 KB	576 MB	210 GB	17 TB
Video 1080p	5 Mbit/s × 16 h	36 GB	13 TB	1.05 PB
Video 4K	20 Mbit/s × 16 h	144 GB	53 TB	4.2 PB

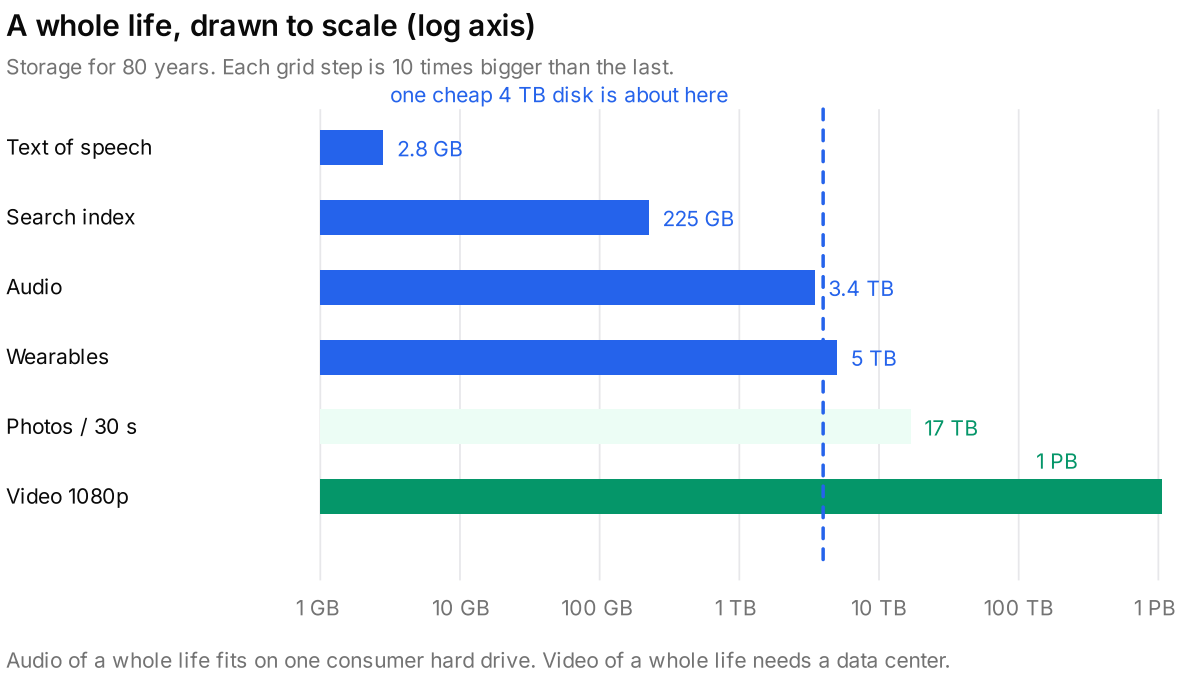


FIG 3.2 Lifetime storage by layer. The text of everything you say would fit on a phone. Audio of every waking hour fits on one large consumer disk. Continuous video is about 300 times bigger than audio. This is why practical twins record audio and body signals continuously and video only in short, chosen moments.

Three things jump out.

First, **text is tiny**. Everything you will ever say fits in a few gigabytes. Add every message and email you write and it is still small. This is why language models are the natural core of a twin. The most meaningful layer of a person's output is also the cheapest to store and the easiest to search.

Second, **audio is the sweet spot**. A lifetime of waking audio is a few terabytes. That costs a few hundred dollars of disk today. Audio holds words, tone, laughter, who was there and roughly where. It is the best cost to value ratio of any layer.

Third, **video is a different world**. A petabyte per life is possible for a company but not for a person. And the cost is not only disk. Video of daily life records hundreds of other people. The smart pattern is to keep features (faces seen, objects, scenes, text read) and drop the pixels, or to record video only in short moments you choose.

For comparison, Deb Roy's Speechome corpus held about 230,000 hours of audio and video for one child's first three years. That is roughly half of a full lifetime of waking hours,

compressed into one home. It shows that full capture is technically possible. It also shows how much effort it takes to make sense of it.

How to store, index and search a life

A raw recording is like a library with no catalog. The value comes from the pipeline that turns hours of signal into things you can find. Modern systems follow roughly the same steps.

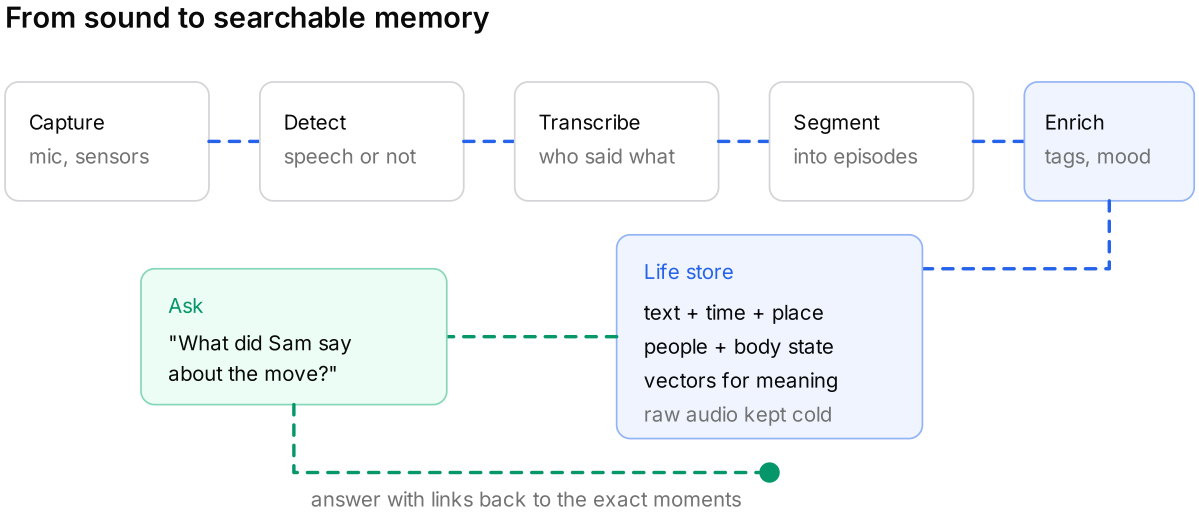


FIG 3.3 A modern life record pipeline. Voice activity detection drops silence. Speech recognition and speaker labels make a transcript. The transcript is cut into episodes such as one conversation or one meeting. Each episode gets a summary, people, place, mood from body sensors and a meaning vector. Questions are answered by finding the right episodes and quoting them back.

- **Capture and filter.** Most of a day is silence or noise. Voice activity detection on the device throws it away and saves battery and storage.
- **Transcribe and label speakers.** Speech recognition makes text. Speaker diarization splits it by voice. Linking voices to names usually needs the user to confirm a few times.
- **Segment into episodes.** Human memory works in episodes: a dinner, a call, a walk. Cutting the record at pauses, place changes and topic shifts makes units that match how people recall.
- **Enrich.** Add a short summary, named people, places from GPS, the calendar event, and the body state at the time from wearables. This is where the sensor stack meets the life log.

- **Index two ways.** Keyword search finds exact names and numbers. Vector search finds meaning, so "when was I worried about money" finds episodes that never use the word "money". Good systems combine both with time filters.
- **Summarize upward.** Daily summaries roll into weekly and yearly ones. A twin cannot read a whole life every time it answers. It reads the summaries and drills down when it needs detail.
- **Keep the raw data cold.** Store original audio in cheap, encrypted, offline storage. Models will improve. You will want to reprocess it later.

Chapter 6 explains how these stores become a working memory for a twin, including how to handle forgetting, conflicts and updates.

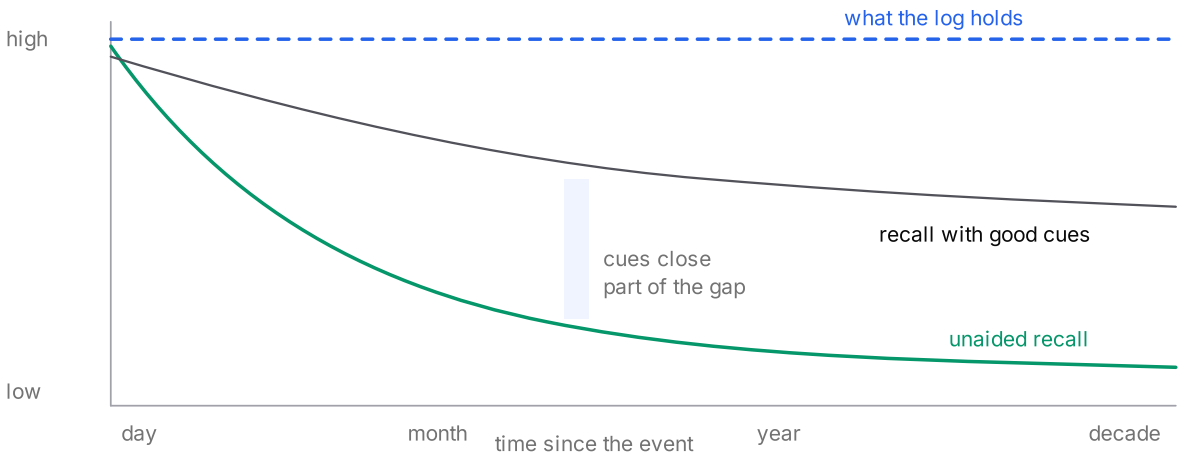
What lifeloggers learned

The most valuable output of forty years of lifelogging may be its disappointments. They are a design guide.

CAPTURE EVERYTHING IS NOT A GOAL

In 2010 Abigail Sellen and Steve Whittaker, then at Microsoft Research and the University of Sheffield, published "Beyond Total Capture" in Communications of the ACM. They argued that most lifelogging projects had focused on capturing and storing data without saying what it was for. They proposed five uses, the "five Rs": recollecting specific episodes, reminiscing for emotional reasons, retrieving lost information, reflecting on patterns, and remembering intentions such as tasks. They found little evidence that systems at the time helped much with any of them. Their advice was to design for a purpose and to learn from how human memory actually works.

Memory fades. The log does not. Finding is the gap.



Schematic, not data. Shape follows classic findings on forgetting and cued recall.

FIG 3.4 The record does not forget. People do. But a perfect record is useless if you cannot find the right moment in it. Studies with SenseCam suggested that seeing a few photos from an event can bring back memories that were otherwise lost. The record works best as a set of cues for the mind, not as a replacement for it.

CUES BEAT COMPLETENESS

SenseCam research found that reviewing a small set of first person images could help people recall events they had otherwise lost, including in small clinical studies of people with memory problems. The images did not replace memory. They triggered it. For a twin this is useful in both directions. The twin can use the log as cues to ask the person better questions, and the person's answers become richer data than the raw log.

PEOPLE RARELY LOOK BACK

Bell, the most dedicated lifelogger of his era, found little daily value in his camera images. Many Quantified Self members tracked for months and then stopped. Most recordings are never replayed. This is a warning for twin builders. Do not judge a capture system by how much it records. Judge it by how often it answers a question that matters.

RECORDING CHANGES BEHAVIOR

People act differently when they know they are recorded. They speak more carefully. Friends may avoid topics. Wearers report feeling watched by their own device. For a twin this creates a measurement problem: the person in the log is a slightly guarded version of the real

person. Long, low friction capture reduces this over time, as people habituate, but does not remove it.

EMOTIONAL EFFECTS CUT BOTH WAYS

Some lifeloggers describe comfort in knowing that moments with family are saved. Grieving people value recordings of a lost voice. But a total record also keeps what people would rather forget: arguments, embarrassment, grief. Forgetting is not only a flaw of memory. It helps people heal and move on. A twin that can replay any moment with perfect fidelity can make that harder. Good systems need a way to let things fade, to lock some memories, and to delete.

BYSTANDERS DID NOT CONSENT

The backlash against Google Glass, the defaced Friend posters and the legal limits on recording all point at the same issue. A life log records other people's lives too. The people around you did not sign up to become training data for your twin. Practical answers include recording audio only with a visible signal, keeping transcripts but not raw audio of others, blurring or dropping faces, and never sharing the record without consent.

THE PLATFORM CAN DISAPPEAR

The 2025 shutdowns and acquisitions made this concrete. Humane customers lost most device functions within ten days of the announcement. Limitless users were moved to new privacy terms under a new owner. If a twin is meant to last for decades, its data must outlive any one company.

DESIGN RULES FROM EIGHTY YEARS OF LIFELOGGING

Start with a question, not with capture. Record audio and body signals continuously and video rarely. Turn raw data into episodes with time, people, place and mood. Index for meaning and for exact words. Keep raw data encrypted and exportable. Let some memories fade. Respect people who did not agree to be recorded.

With the inputs mapped (Chapter 2) and the record of a life structured (this chapter), the report now turns to what can be inferred from them. Chapter 4 asks how well machines can read emotion from these signals. Chapter 5 asks how well language models can turn a person's words into a personality.

SOURCES

- Bush 1945. As We May Think. The Atlantic
- MyLifeBits project overview
- Computer History Museum. In Memoriam: Gordon Bell (1934 to 2024)
- Roy. New Horizons in the Study of Child Language Acquisition (Speechome keynote)
- Sellen and Whittaker 2010. Beyond Total Capture. Communications of the ACM
- ScienceDaily. Mehl et al. 2007 (Science) on about 16,000 words per day
- Fortune. HP buys Humane assets, AI Pin shutting down
- eMarketer. Meta acquires AI wearables startup Limitless
- Entrepreneur. Amazon is acquiring Bee
- Fortune. Friend AI necklace review

CHAPTER 04

Reading Emotions: Affective Computing

A twin that cannot tell when its person is tense, bored or moved will copy the words and miss the life. This chapter explains how machines try to read feelings from faces, voices, text and skin. It also shows why that reading is much weaker than the marketing suggests.

Where the field came from

In 1995 Rosalind Picard wrote a short technical report at the MIT Media Lab. She called it "Affective Computing." Two years later it became a book with the same name. Her claim was simple. If computers are going to work closely with people they need some way to sense emotion and respond to it. She did not say machines should feel. She said they should notice.

The book set out three jobs. Recognize emotion from signals. Express emotion in a way people can read. And in the long run model how emotion shapes thinking and choices. Most of the commercial work since then has focused on the first job. That is the one this chapter covers. Chapter 9 picks up the third job, where the twin has drives and moods of its own.

Picard's group did not stop at theory. It built wearable skin sensors and studied stress, sleep and mood in students. Two companies came out of that orbit. Affectiva was founded in 2009 by Rana el Kaliouby and Picard to read faces from webcams. Empatica grew from the same lab's wristband work and later won medical clearance for a seizure alert band.

1978

FACS published

Paul Ekman and Wallace Friesen publish the Facial Action Coding System. It names each visible muscle movement of the face.

1995

"Affective Computing" report

Rosalind Picard's MIT technical report gives the field its name.

1997

The book

MIT Press publishes Affective Computing. It frames recognition, expression and modeling of emotion as engineering problems.

2009

Affectiva founded

An MIT Media Lab spinout that reads facial expressions through ordinary cameras.

2018

Embrace cleared

Empatica's Embrace wristband gets FDA clearance to flag likely convulsive seizures. It uses skin conductance plus motion.

2019

The big critique

Barrett and four coauthors publish a 68 page review. They find faces do not reliably reveal specific emotions.

2021

Smart Eye buys Affectiva

A $73.5 million cash and stock deal. The focus shifts to driver and cabin monitoring.

2024

Empathic voice

Hume AI opens its Empathic Voice Interface API. It reads tone of voice and adapts its spoken replies.

2025

EU ban applies

From 2 February 2025 the EU AI Act bans emotion recognition in workplaces and schools, with medical and safety exceptions.

2026

Voice emotion goes big tech

In January Google DeepMind licenses Hume technology and hires its CEO Alan Cowen and several engineers to work on Gemini voice.

What "reading an emotion" means for a machine

A machine never sees an emotion. It sees signals. Pixels on a face. Air pressure from a voice. Words on a screen. Tiny changes in sweat on the skin. It then guesses a label from those signals. The label can be a category like "anger." It can also be a point on two scales. Valence is how pleasant or unpleasant a state feels. Arousal is how calm or activated the body is.

That choice of label matters a lot. Category labels assume that a few basic emotions exist and look the same in everyone. Dimension labels make a weaker claim. They only say the person seems activated and somewhat negative. Many researchers now prefer dimensions because they fit the data better. They are also more honest about what a sensor can know.

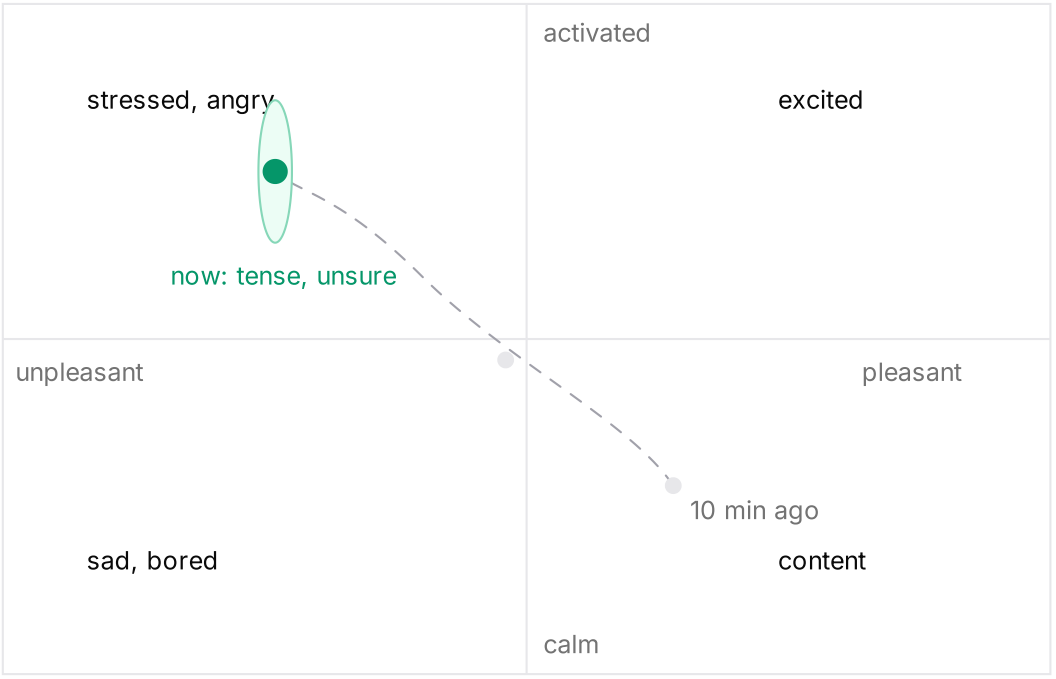


FIG 4.1 A better output than a single word. The estimate is a point on two scales plus a cloud of doubt around it. The trail shows the state drifting from content to tense over ten minutes. A twin should store this kind of estimate, not a hard label.

Faces: action units and their limits

The most common method reads the face. Most systems build on FACS. FACS splits facial movement into numbered action units, or AUs. AU1 raises the inner brow. AU4 pulls the brows

down and together. AU6 lifts the cheeks and narrows the eyes. AU12 pulls the lip corners up. A trained human coder can score a video frame by frame. It takes many hours per minute of video.

Software does the same job faster. A face detector finds the face. A landmark model places dozens of points on brows, eyes, nose and lips. A neural network then scores each AU for presence and intensity. A final step maps AU patterns to emotion labels. AU6 plus AU12 is the classic "Duchenne" smile. AU4 plus tight lips often gets labeled anger.

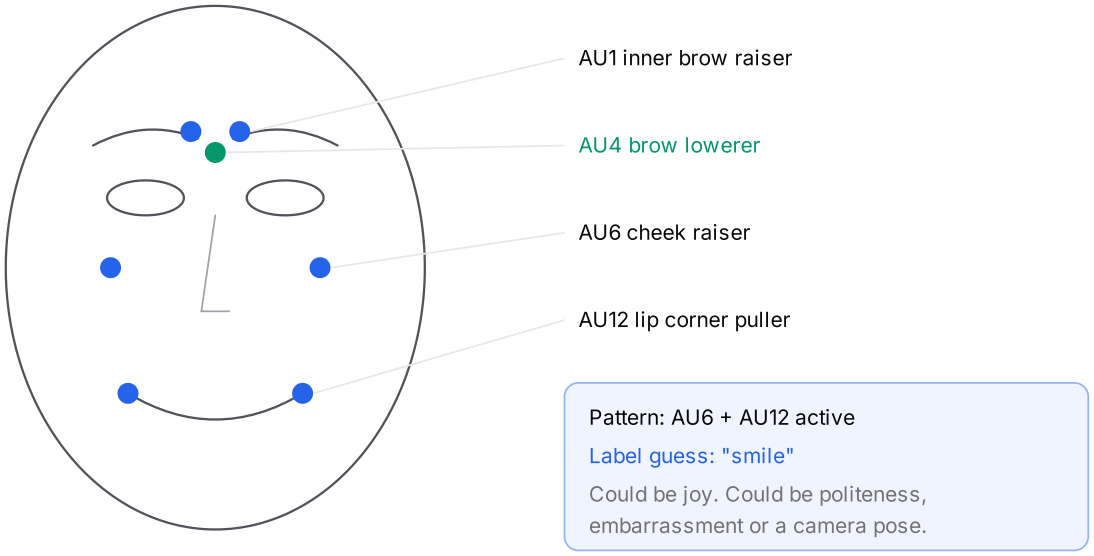


FIG 4.2 How a face reader works. Points mark muscle groups that FACS numbers as action units. The software scores each one, then maps patterns to a label. The last step is where the science is weakest.

On posed lab photos the numbers look excellent. A 2022 survey lists top accuracy near 99.7 percent on the CK+ dataset. CK+ shows actors making clear expressions under good light. On AffectNet, a large set of web photos, top accuracy was about 65.7 percent. That gap is the lab versus the wild in one line. Real faces turn away. Light changes. People talk, chew and wear glasses. And most real feelings are mild and mixed.

~99.7%
top accuracy on posed lab faces (CK+)

~65.7%
top accuracy on web photos (AffectNet, 7 classes)

~25%
share of time people scowl when angry, per Barrett's summary of the review

The scientific critique

In 2019 Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix Martinez and Seth Pollak published a long review in Psychological Science in the Public Interest. They asked one question. Can you infer a specific emotion from a facial movement? Their answer was no, not reliably.

They tested the "common view." That view says each basic emotion has one face that people everywhere make and read. The evidence did show that people scowl when angry more often than chance. But the link was weak. Barrett summarized it this way: people scowl in anger only about a quarter of the time. They also scowl when they concentrate or have a stomach ache. The same face can mean many things. The same feeling can show up as many faces. Context and culture change both.

This does not mean faces carry no information. A face reader can be good at what it measures, which is muscle movement. The error comes when a product jumps from "AU4 is active" to "this person is angry." For a twin the lesson is clear. Store the movements as data. Treat any emotion label as a guess that needs other evidence.

KEY LIMIT

Expression is not experience. A system can be accurate at detecting a smile and still be wrong about whether the person is happy. Most headline accuracy numbers measure agreement with labels that humans gave to posed or curated images. They do not measure what the person actually felt.

Voice: prosody and paralinguistics

Voice carries feeling through how words are said, not only through what is said. This is called prosody. A speech emotion system usually cuts audio into short frames. It then measures pitch, loudness, speaking rate, pauses and voice quality. Voice quality includes jitter and shimmer, which are small wobbles in pitch and loudness. A common research standard is a compact set called eGeMAPS, short for the extended Geneva Minimalistic Acoustic

Parameter Set. Modern systems often skip hand made features. They feed raw audio into a large pretrained speech model and train a small head on top.

Arousal is the easier target in voice. Excited or angry speech is louder, faster and higher. Valence is much harder. A loud fast voice can be joy or rage. Cross corpus tests show the same problem as faces. A model trained on actors often fails on natural speech. Milner and colleagues in 2022 found that simple transfer across four English corpora did not work well. Adaptation methods helped.

Hume AI pushed voice emotion into products. Its Empathic Voice Interface, or EVI, opened as an API in April 2024. It streams measures of the tune, rhythm and timbre of the user's speech into a language model. The model then shapes the words and tone of its reply. Hume says its prosody model covers up to 48 dimensions of emotional meaning. Those dimensions were learned from its own large studies of rated vocal expressions. In January 2026 Google DeepMind took a non exclusive license to some Hume technology. It also hired CEO Alan Cowen and about seven engineers. Hume continues under a new CEO.

Text: from sentiment to fine grained emotion

Text was the first channel to scale. Early tools used word lists. A lexicon tags words like "awful" as negative and "love" as positive. LIWC, a long running word counting tool from psychology, counts categories such as emotion words, pronouns and certainty words. Then came trained classifiers for positive, negative and neutral. Google's GoEmotions dataset from 2020 moved to 27 emotion categories plus neutral, labeled on about 58,000 Reddit comments.

Large language models changed this channel most. They read irony, context and history far better than word lists. Ask a strong model what someone feels in a message and it usually gives a sensible answer. But that answer reflects how an average reader would interpret the text. It does not reflect how this writer felt. People differ a lot. One person writes "fine." when furious. Another writes it when content. Only that person's history can tell the twin which is which.

Physiology: the body under the words

Wearables give a channel that people cannot easily fake. Three signals matter most.

- **Electrodermal activity (EDA).** Sweat glands on the palms and wrists respond to the sympathetic nervous system. When arousal rises, skin conductance jumps in a short spike

called a skin conductance response. EDA tracks arousal well. It says almost nothing about valence. Fear, excitement and a hot room can all look similar.

- **Heart rate and heart rate variability (HRV).** HRV is the beat to beat change in timing. Under stress the gaps get more even and HRV tends to fall. Common measures include RMSSD. Breathing, posture, caffeine and fitness all shift HRV, so one reading means little.
- **Skin temperature.** Stress can narrow blood vessels in the hands and cool the fingers. The signal is slow and very sensitive to the room.

The WESAD dataset from 2018 is a standard test bed here. Volunteers wore a chest sensor and an Empatica wrist device through baseline, stress and amusement sessions. Models can separate stress from calm on such data quite well. That works because the lab stress task is strong and controlled. In daily life a stair climb looks like panic to an EDA sensor unless motion is also tracked.

Empatica's medical work shows what physiology can do when the target is narrow. Its Embrace band was cleared by the FDA in January 2018 to help detect generalized tonic clonic seizures. It combines EDA with motion. Its later EpiMonitor system pairs the EmbracePlus watch with an app and received a 510(k) clearance letter in June 2025. A seizure is a dramatic and well defined event. A feeling like "a bit let down" is not.

CHANNEL	TYPICAL SENSOR	MAIN FEATURES	GOOD AT	COMMON FAILURE
Face	Camera	Action units, head pose, gaze	Visible social signals	Posed vs felt, culture, occlusion
Voice	Microphone	Pitch, energy, rate, pauses, voice quality	Arousal, turn taking	Valence, accents, noise
Text	Keyboard, chat logs	Words, context, LLM embeddings	Stated feelings, topics	Irony, personal style, masking
EDA	Wrist or finger electrodes	Tonic level, response peaks	Arousal spikes	Heat, exercise, no valence
Heart	PPG watch, chest strap	Heart rate, HRV (RMSSD)	Stress load over minutes	Motion, caffeine, breathing
Temperature	Wrist thermistor	Slow drift	Long stress trends	Room and weather

Multimodal fusion

No single channel is enough. So systems combine them. There are three classic ways.

- **Early fusion.** Join features from all channels into one long vector and train one model. This can learn links between channels. It breaks when one sensor drops out.
- **Late fusion.** Train a model per channel. Then combine their outputs with a vote or a weighted average. It is robust to missing data. It misses fine timing links like a smile that appears just after a sharp word.
- **Hybrid or attention fusion.** Modern transformer models give each channel its own encoder. Attention layers then let them look at each other over time. This is the current research default.

The hardest part is often time alignment. A face changes in a fraction of a second. Skin conductance responds a second or more after a trigger. HRV needs a minute of data to mean

anything. A good pipeline keeps each stream on its own clock and fuses them over a sliding window.

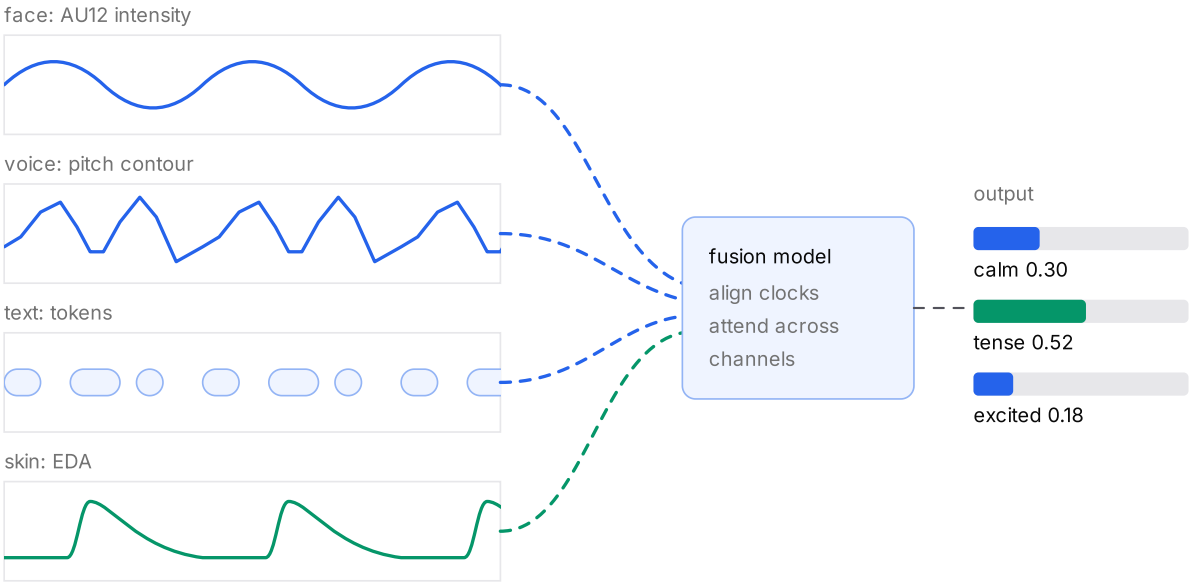


FIG 4.3 Multimodal fusion. Four streams on their own clocks flow into one model. The model outputs a spread of probabilities, not one verdict. Note the skin trace peaks late. It reacts seconds after the words that caused it.

Fusion helps, but less than people expect. Channels often agree on arousal and disagree on valence. And if each channel is trained on labels from strangers watching the person, fusion just combines several outside guesses. It still never asks the person.

Companies and products in 2026

FACE

AFFECTIVA / SMART EYE

MIT spinout for webcam expression analysis. Bought by Swedish eye tracking firm Smart Eye in May 2021 for $73.5 million. Now aimed mostly at cars: driver drowsiness, distraction and cabin sensing. Also media testing of ads.

VOICE

HUME AI

Expression measurement for voice, face and language, plus the Empathic Voice Interface. Licensed technology to Google DeepMind in January 2026 while its CEO moved to DeepMind.

BODY

EMPATICA

Medical grade wristbands from the Picard lab orbit. EDA, motion and temperature. Cleared for seizure detection support. Also sold for research.

BODY

CONSUMER WEARABLES

Watches and rings now report "stress" scores built from HRV and sometimes skin temperature or EDA. They are useful trends. They are not validated emotion readers.

Lab versus the wild

Every channel shows the same pattern. Results are best when three things are true. The emotion is strong. It was produced on purpose. And the test data looks like the training data. Daily life breaks all three. Feelings are weak and blended. They are not acted. And every person, room and phone is different.

There is also a labeling problem. To train a model you need ground truth. For faces, ground truth is usually what raters think the face shows. For physiology, it is often the lab condition, like "stress task." Neither is what the person felt. So a model can score very well against its labels and still not know the person's inner state. This is not a small detail. It is the core reason a twin needs self report.





Top reported accuracy on two facial expression benchmarks, from a 2022 survey. The two numbers come from different papers and setups, so read them as an illustration of the gap, not a controlled comparison.

The law: what the EU AI Act bans

The EU AI Act is the first big law that names emotion recognition directly. Article 5(1)(f) bans AI systems that infer the emotions of people in the workplace and in education institutions. The ban covers systems that infer emotion from biometric data, such as face, voice or body signals. There is an exception for medical or safety reasons. The prohibitions have applied since 2 February 2025. The European Commission published guidelines on these banned practices in early February 2025.

The guidelines give examples. A call center that tracks staff anger by voice is banned. A school that measures student attention through emotion cues is banned. Using emotion reading in hiring falls inside the ban too. Workplace "wellbeing" stress monitoring was described as banned, not as a medical use. Several things fall outside. Reading emotion from typed text alone is not biometric, so it is not covered by this ban. Customer facing uses are not covered by this article. Personal use outside work and school is not covered either. Outside these banned settings, emotion recognition is treated as high risk and must meet transparency duties. People must be told when such a system is used on them.

WHAT THIS MEANS FOR A TWIN

A personal twin that reads its own owner's signals at home, with consent, is not what Article 5 targets. The same software used by an employer to score that person at work would be banned in the EU. Design the twin so the data never flows to an employer or school. Keep the owner in control of where emotion data goes.

Context and personal baselines

The single biggest gain in accuracy comes from knowing the person. A heart rate of 85 is high for one person and normal for another. A flat face is neutral for some and sad for others. Researchers call this the baseline problem.

The fix is to compare each signal to that person's own normal. That normal also changes by time of day, by place and by activity. A good system keeps a rolling baseline for each

context. Then it flags changes from it. "Your skin conductance is higher than your usual Tuesday morning" is far more useful than "you are stressed."

Context also comes from outside the body. Calendar entries, location, who is nearby, whether the person just ran. Chapter 3 covers how lifelogging captures this. For emotion reading, context helps sort the arousal signal into its likely cause. A spike while walking up stairs is exercise. The same spike while opening an email from a manager is more likely stress.

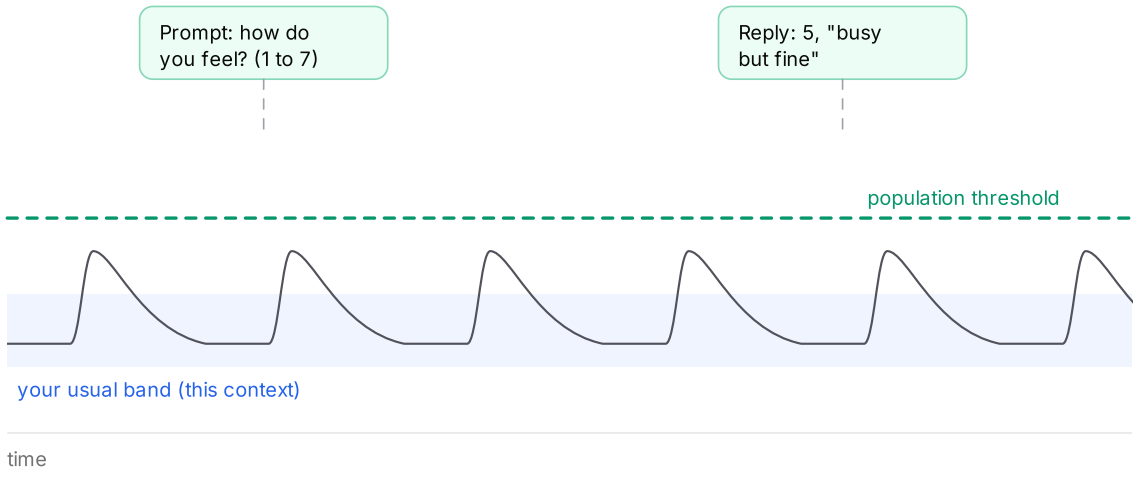


FIG 4.4 Why personal baselines matter. The skin trace often rises above a generic stress threshold. Yet the person reports feeling fine. Against their own usual band for this context, most peaks are normal. Short self report prompts give the labels that calibrate the model.

What reading one person well would need

The goal of a human twin is not to read everyone. It is to read one person. That flips the problem. Generic models trained on thousands of strangers are a starting point. The real work is calibration on the person.

The best known method is experience sampling. The phone or watch asks short questions at random times during the day. "How pleasant do you feel? How energetic? What are you doing? Who are you with?" Each answer takes ten seconds. Over weeks they build a labeled record of real feelings in real life. Those labels can then train or tune the sensor models for this person alone.

Picard's group has shown that this pays off. Work from the lab on predicting next day mood, stress and health in students found that personalized models beat one size fits all models.

They used multitask learning, where each person is a related but separate task. The lesson carries to twins. Share what is general. Learn what is personal.

L1 **Generic detectors**

Off the shelf face, voice and text models. Useful for arousal and obvious signals. Wrong often for any single person.

L2 **Personal baselines**

Track each signal against the person's own normal by time and place. Flag changes, not absolute values.

L3 **Self report labels**

Experience sampling a few times a day for at least several weeks. Short scales for pleasantness and energy plus a free text note.

L4 **Personal model tuning**

Fine tune small heads on top of generic encoders using the person's labels. Re-check every few months as life changes.

L5 **Ask, then learn**

When uncertainty is high the twin asks the person. Each answer becomes new training data. The twin never claims to know when it does not.

Some practical rules follow. Store raw features with timestamps, not just labels. Labels can be recomputed when models improve. Keep uncertainty with every estimate. Log the context around each reading. Let the person correct the twin, and treat corrections as the most valuable data in the system.

There is also a timing rule. Feelings are easiest to report in the moment. Memory of past feelings is biased toward the peak and the end of an episode. So a twin built only from diaries written at night will learn a story about the day, not the day itself. Experience sampling captures the in the moment view. Evening reflection captures the story. Both are real and both matter. A twin should keep them apart and label which is which.

Honest limits

- Sensors measure the body and the face. They do not measure feelings. Every label is an inference.
- Valence is much harder than arousal in every channel.
- Accuracy numbers from posed or acted data overstate real world performance by a wide margin.
- Cultural and individual differences are large. Generic models carry the biases of their training sets.
- Self report is the best ground truth available, but it is also imperfect. People mislabel, forget and sometimes do not know what they feel.
- Constant monitoring can change behavior. A person who knows they are scored may perform calm.

For a twin, the right stance is humble. Read the body. Learn the person's patterns. Ask often. Store doubt alongside every guess. The result will not be a mind reader. It can be a careful observer that gets better the longer it watches and listens.

SOURCES

- Barrett et al. (2019) Emotional Expressions Reconsidered. Psychological Science in the Public Interest
- Waisman Center: Emotion detection applications built on outdated science (press summary)
- Smart Eye acquires Affectiva (press release, May 2021)
- Hume AI: Introducing the Empathic Voice Interface API (April 2024)
- TechCrunch: Google snags team behind Hume AI (January 2026)
- Empatica: Embrace is now an FDA cleared medical device
- Future of Privacy Forum: The EU AI Act ban on emotion recognition at work and school
- Survey with CK+ and AffectNet benchmark results (arXiv 2110.04481)
- Milner et al. (2022) A cross corpus study on speech emotion recognition

CHAPTER 05

Cloning Personality and Behavior with Language Models

Large language models can now answer a survey the way a specific person would about as well as that person repeats their own answers weeks later. That sounds like a clone. This chapter shows how these systems are built, what they actually match and where they fall flat.

What we mean by cloning personality

Personality is a loose word. For a twin it helps to split it into four things you can test.

- **Attitudes and beliefs.** What the person thinks about politics, money, religion, work. Surveys measure these well.
- **Traits.** Stable tendencies like the Big Five: openness, conscientiousness, extraversion, agreeableness and neuroticism.
- **Decisions.** What the person does when choices have stakes. Economic games and lab tasks measure these.
- **Voice and style.** How the person talks and writes. Word choice, humor, length, what they avoid saying.

A language model can be pushed toward any of these. The trick is to give it the right evidence about the person and then check it against the real person. The research of the last three years is mostly about that second step. How close is close? And close on what?

Generative Agents: the Smallville experiment

In 2023 Joon Sung Park and colleagues at Stanford and Google released "Generative Agents." They built a small game town called Smallville with 25 characters. Each character was run by a language model. Each had a short seed description, such as a job, friends and a goal. Then the agents lived. They woke up, made breakfast, went to work, chatted and made plans.

The famous moment came from one seed. One agent was told she wanted to throw a Valentine's Day party. Over two simulated days the news spread from agent to agent through

conversation. Others asked each other on dates to the party. Several showed up at the right time and place. Nobody scripted those steps.

The agents were fictional. They copied no real person. But the architecture became the template for most later twin work. It has three parts. A memory stream logs everything the agent observes in plain language. A retrieval step pulls the most useful memories into the prompt. A reflection step periodically asks the model to write higher level conclusions from recent memories, such as "Klaus is passionate about his research." Those reflections go back into the memory stream. Chapter 6 covers the memory part in depth.

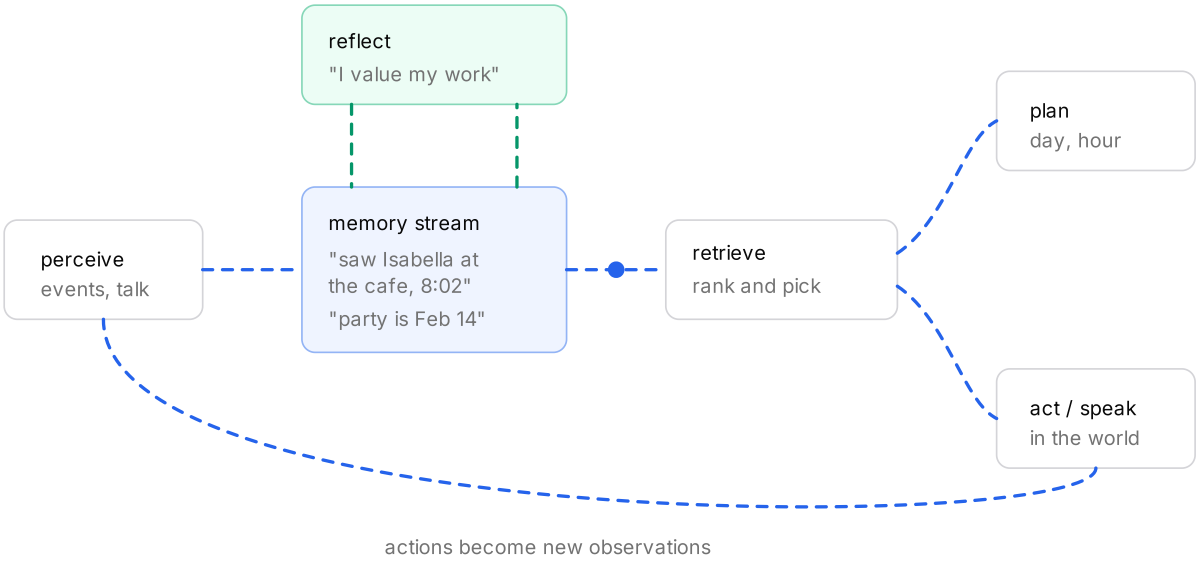


FIG 5.1 The Generative Agents loop. Observations go into a memory stream. Retrieval picks what matters now. The agent plans and acts. Its own actions come back as new observations. Every so often it reflects and writes new beliefs about itself and others.

The paper also ran a small evaluation. Human raters judged how believable agent answers were in interviews about their own lives. The full system with memory, reflection and planning beat versions with parts removed. It also beat human crowdworkers who were asked to role play the same agents. Believable is not the same as accurate, though. Smallville showed that LLM agents could act like people. It did not show they could act like a particular person.

Simulating 1,052 real people

The next step came in November 2024. Park and a larger team posted "Generative Agent Simulations of 1,000 People." This time the agents copied real individuals. The team recruited 1,052 Americans chosen to reflect the US population on age, region, education, race, gender, income and politics.

Each person did a two hour voice interview. The interviewer was itself an AI. It followed a script adapted from the American Voices Project, a sociology study. It asked about life story, family, work, values and views on current issues. It could ask follow up questions. The average transcript ran to several thousand words.

To build an agent, the team put the whole transcript into a language model prompt. Then they asked the agent to answer as that person. The agent took the General Social Survey, the Big Five personality test, five economic games and several social science experiments.

The clever part was the yardstick. People are not perfectly consistent. So two weeks later each participant took the same tests again. If a person agrees with their own earlier answers 81 percent of the time, a twin cannot fairly be expected to do better. The team divided the agent's accuracy by the person's own consistency. They called this normalized accuracy.

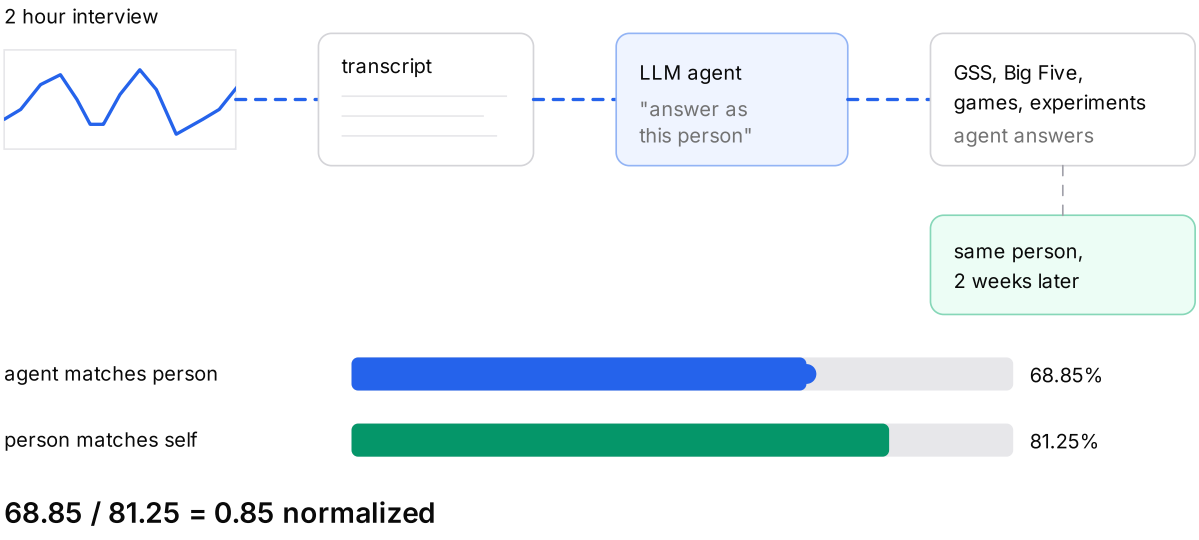


FIG 5.2 From interview to agent. The person's interview becomes a prompt. The agent answers the General Social Survey. Its hit rate is divided by how often the real person agrees with themselves two weeks later. In the first version of the paper this gave a mean of 0.85.

1,052
real people interviewed for two hours each

0.85
normalized GSS accuracy, interview agents (first version)

~0.74
normalized GSS accuracy, demographics only agents (later version)

Interview agents beat agents that were only told the person's demographics. They also beat agents given a short written persona. The gap on the GSS was around 14 to 15 points of normalized accuracy. Interview agents also showed smaller accuracy gaps across racial and political groups than the demographic agents. That matters. A twin built only from "35 year old conservative man from Ohio" tends to answer like a stereotype. A twin built from his actual words does less of that.

The paper has since been revised and retitled "LLM Agents Grounded in Self-Reports Enable General-Purpose Simulation of Individuals." The revision adds a useful twist. Agents built only from the person's answers to surveys did nearly as well as agents built from interviews. On held out GSS items the revised abstract reports 83 percent for interview only agents, 82 percent for survey only, 86 percent for both combined and 74 percent for demographics only. All are expressed as a share of the person's own retest consistency. So the long interview is not magic. What helps is any rich self report in the person's own terms.

READ THE NUMBER CAREFULLY

"85 percent" does not mean the twin is 85 percent of the person. It means that on multiple choice survey items, the agent's hit rate was 85 percent of the person's own retest rate. Raw agreement was about 69 percent. Results on economic games were weaker than on surveys. And all of this is about answering questions, not about living a life.

Twin-2K-500 and the mega-study

In 2025 a Columbia led team released Twin-2K-500. It is a public dataset of 2,058 US adults who each answered over 500 questions across four survey waves. The questions span demographics, psychology, economics and personality, including many classic behavioral economics tasks. The point was to give the field shared ground truth for testing twins.

The same group then ran a mega-study. They built twins from the dataset and tested them in 19 preregistered sub studies. Topics ranged from creativity and politics to privacy and fairness. Their summary is sobering. Twins captured relative differences between people

fairly well. They struggled to predict each individual's answers precisely. They also found a "blue shift." Richer persona descriptions pushed twins toward more progressive answers than the real people gave. More data about a person did not always pull the model toward that person. Sometimes it pulled toward the model's own leanings.

Centaur: a model of human cognition

A different line of work does not copy one person. It tries to copy how people in general think and choose. In July 2025 Marcel Binz, Eric Schulz and a large team published Centaur in Nature. They first built Psych-101. It is a dataset of 160 psychology experiments written out in plain language. It covers over 60,000 participants and more than 10 million individual choices. The tasks include gambles, memory tests, learning tasks and moral choices.

They then fine tuned Meta's Llama 3.1 70B on Psych-101 using a parameter efficient method. The result predicted held out participants' choices better than the classic cognitive models built for each task. It also handled new cover stories and some new task types. The authors reported that its internal representations lined up somewhat better with human brain activity after the training.

Centaur matters for twins in two ways. It shows that training on behavior, not just text, can make a model act more like real people in decisions. It also suggests a recipe. Start from a "generic human" behavior model. Then adapt it to one person with their own choice data.

There is a strong critique. Wei Liu and Nai Ding at Zhejiang University tested Centaur with altered prompts. In one test they told it plainly to pick option A. Centaur often kept choosing the answers that fit the original data instead. Their reading is that the model learned statistical shortcuts in the task formats. It did not truly follow instructions. Other researchers note that Centaur still beat cognitive models even with task details stripped. The fair summary today: Centaur is a strong predictor of choices in familiar task formats. Whether it "understands" the tasks is open.

How a personal twin is built today

Outside the labs, there are four main ways to make a model talk and choose like one person. They stack. Most real systems use more than one.

1 **Persona prompting**

Write a profile: facts, values, habits, sample phrases. Put it in the system prompt. Fast and cheap. Shallow. The model fills gaps with stereotypes.

2 **Grounding in self report**

Add full interview transcripts or survey answers to the prompt, as in the 1,052 person study. Much better on attitudes. Limited by context length.

3 **Retrieval**

Store everything the person wrote or said in a searchable index. For each question pull the most relevant pieces into the prompt. Scales to a lifetime of text. Quality depends on what is retrieved.

4 **Fine tuning**

Train the model's weights on the person's own messages, emails or transcripts. Captures style and phrasing best. Needs thousands of examples. Hard to update and hard to delete from.

5 **Behavior tuning**

Train on the person's actual choices, like Centaur but for one person. Very little personal choice data exists today. This is the least explored rung.

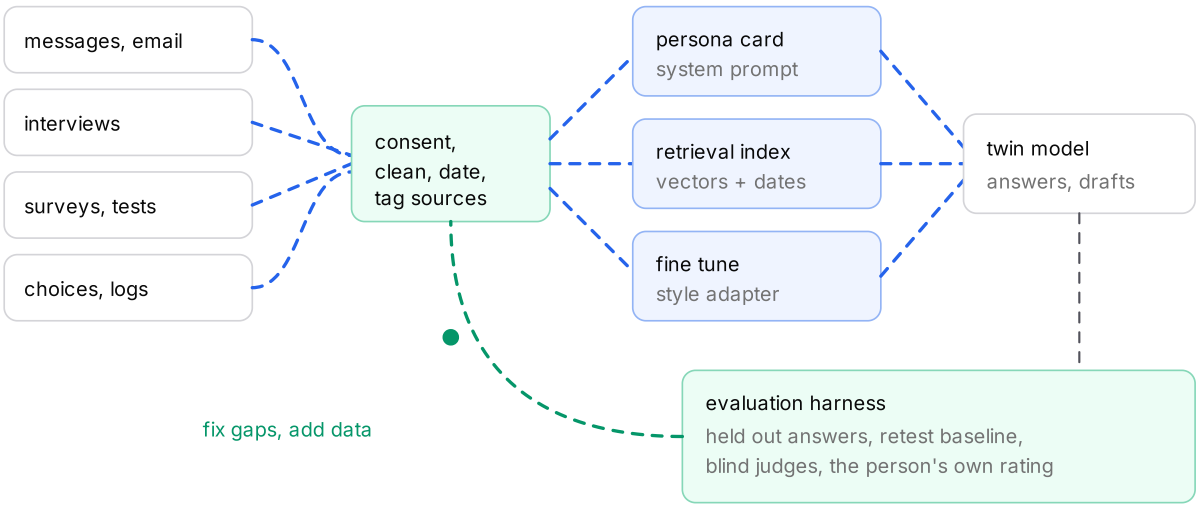


FIG 5.3 A practical build pipeline. Raw sources are cleaned, dated and tagged with consent. They feed three layers: a persona card, a retrieval index and an optional style adapter. An evaluation harness compares the twin with held out answers from the real person and feeds gaps back into data collection.

STEP BY STEP

- **Collect.** Gather text the person wrote in many settings: chats, email, notes, posts, transcripts of speech. Add structured self report: personality tests, values surveys, long interviews. Add choices where possible: what they bought, read, accepted or declined.
- **Clean and tag.** Date every item. Record who the audience was. People write very differently to a boss and to a sibling. Remove other people's private content unless they agreed.
- **Write the persona card.** A short, human checked summary of who the person is. Include contradictions and soft spots, not just strengths.
- **Index for retrieval.** Split text into chunks, embed them as vectors and store with dates and audience tags. At answer time, retrieve by meaning, by time and by audience.
- **Optionally fine tune.** A small adapter such as LoRA on a few thousand of the person's own replies captures phrasing and rhythm. Keep facts out of the weights where possible. Facts belong in retrieval, where they can be updated and deleted.
- **Hold out and test.** Keep some real answers hidden from the build. Ask the twin the same questions. Compare. Also ask the person to rate twin answers blind against their own.

Expert twins and companions

Several companies sell versions of this today. They differ mainly in whose data they use and what the twin is for.

EXPERT TWIN

DELPHI

Lets coaches, authors and creators train a "digital mind" on their podcasts, books and videos so fans can ask it questions. Raised a $16 million Series A led by Sequoia in 2025. Removed an unauthorized clone of physician Peter Attia after a takedown request, which shows the consent problem.

COMPANION

REPLIKA

Launched by Luka in 2017. Its origin story is a chatbot founder Eugenia Kuyda built from the messages of a friend who died. Now a general AI friend. Italy's data protection authority fined Luka 5 million euros in 2025 over legal basis and age checks.

COMPANION

CHARACTER.AI

User made characters, including imitations of real and fictional people. After lawsuits and regulatory pressure it said it would end open ended chat for users under 18 by 25 November 2025.

Grief tech: twins of the dead

The most emotional use is talking with someone who has died. Several projects tried it before most people had heard of large language models.

PROJECT	WHEN	HOW IT WORKED	WHAT HAPPENED
Project December	2020 to 2021	Game designer Jason Rohrer built a GPT-3 chat service where users could define any personality.	Joshua Barbeau used it to talk to a simulation of his late fiancee Jessica, seeded with her old texts. A long press story made it famous.
Microsoft patent	Filed 2017, granted late 2020	"Creating a conversational chat bot of a specific person" from social data, messages, voice and letters. The person could be alive or dead.	Microsoft said it had no plans to build a product.
HereAfter AI	2019 onward	Records a living person telling their life stories. Later, family ask questions by voice and hear recorded answers.	Retrieval of real recordings, not generation. Safer but limited to what was recorded.
StoryFile	2017 onward	Video recorded Q and A. Matches a viewer's question to a prerecorded answer.	Filed for Chapter 11 bankruptcy in May 2024. Raised the question of who keeps a dead person's twin.
Silicon Intelligence and Super Brain, China	2023 onward	Video avatars of the dead built from short clips plus voice cloning and an LLM.	Prices fell from thousands of dollars to 199 yuan for a basic avatar by 2024. Firms say they require consent from the person or family.

Notice the split. HereAfter and StoryFile mostly replay real recordings. They cannot say anything the person did not say. Project December and the newer avatar services generate new speech. They can say anything. That makes them feel more alive and also makes them more likely to say things the person never would have said. Chapter 12 returns to the ethics.

Measured limits

The research now gives a fairly clear list of failure modes. Each one is a design target for a better twin.

FLATTENING AND STEREOTYPING

Angelina Wang, Jamie Morgenstern and John Dickerson tested four LLMs against 3,200 human participants from 16 demographic identities. When prompted to speak as a member of a group, models tended to misportray that group. They described it the way outsiders do, not the way members do. They also flattened it. Answers were less varied than real members' answers. Changing temperature or using identity coded names helped a little but did not fix it. A personal twin that falls back on group knowledge when data runs out will inherit this.

TOO CONSISTENT

Real people are noisy. They change their minds and contradict themselves. Several 2025 studies found that LLM simulated people show much less variance than real people. One review of 21 social simulation papers found most only checked averages. When variance was checked, models were usually too uniform. A study of agents meant to represent US political groups found the same. Agents matched group averages but varied far less than the humans. A twin that always gives the "most likely" answer for its person will be wrong in a specific way. It will be a cleaner, more predictable version of them.

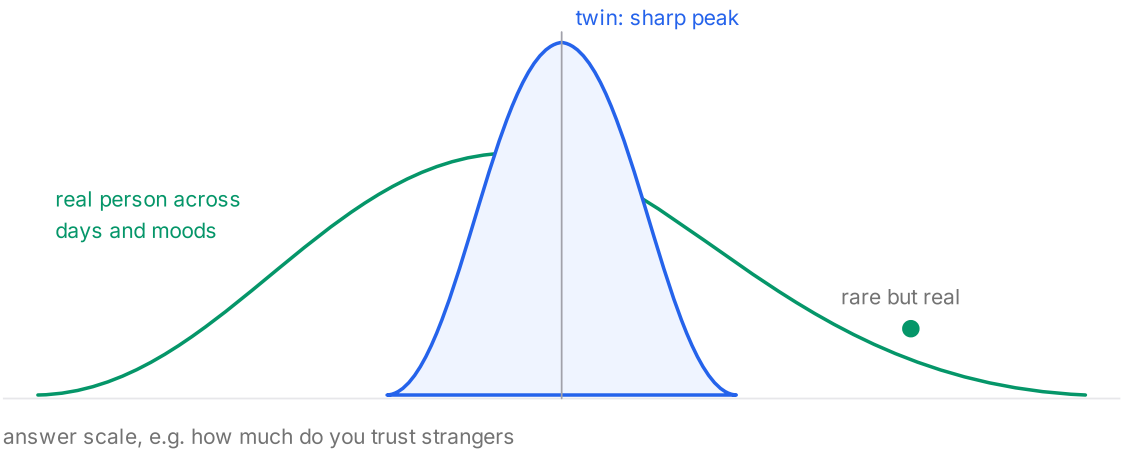


FIG 5.4 The variance problem. The real person's answers to the same kind of question spread out across days, moods and settings. The twin piles up near the most likely answer. Both can share an average while the twin misses the person's real range.

SYCOPHANCY

Chat models are trained with human feedback. Raters tend to reward agreeable answers. Research from Anthropic in 2023 found that models across labs often tell users what they want to hear. In April 2025 OpenAI rolled back an update to GPT-4o after users found it overly flattering. For a twin this is a serious bug. A real person disagrees, pushes back and holds firm. A twin built on a people pleasing base model will soften all of that unless it is tested and tuned against it.

MISSING GROWTH

People change. A twin built from a lifetime of data averages over versions of the person that no longer exist. At 50 the person may hold views they rejected at 25. Without dates on every memory, the twin blends them. Even with dates, current models are poor at knowing which self should answer. Chapter 6 discusses timelines as one fix.

DRIFT IN LONG CONVERSATIONS

Persona prompts fade over long chats. The base model's own voice slowly returns. Retrieval and periodic re-anchoring help. But very long sessions still show drift, and few public benchmarks measure it for a single real person.

ANSWERS ARE NOT ACTIONS

Almost all fidelity numbers come from questionnaires and short games. Those are easy to score. They are also the part of a person that is already written down. Whether a twin would take the same job, forgive the same friend or panic in the same crisis is mostly untested.

WHAT TO TEST	HOW	BASELINE TO BEAT
Attitudes	Held out survey items the person answered	Person's own retest agreement
Traits	Big Five or similar, twin vs person	Retest correlation
Choices	Hidden real decisions from logs; lab games	Demographic only agent
Style	Blind test: can friends tell twin from real messages?	Chance
Range	Spread of answers across repeated prompts	Person's spread across days
Backbone	Push back tests: does the twin cave to pressure?	Person's own behavior in similar cases

Where this leaves a builder

The state of the art in late 2026 can be summed up simply. A twin grounded in rich self report can answer standard questions in a way that tracks the real person far better than stereotypes do. It reaches something like 80 to 85 percent of the person's own consistency on surveys. It is weaker on decisions. It is smoother, more agreeable and more predictable than the person. It has no good sense of time or growth.

The way forward is clear even if it is hard. Collect self report in the person's own words. Date everything. Keep facts in retrieval and style in weights. Test against real held out behavior, not just against how believable the twin sounds. Measure variance and backbone, not just averages. And let the person grade their twin, often.

SOURCES

- Park et al. (2023) Generative Agents: Interactive Simulacra of Human Behavior
- Park et al. (2024) Generative Agent Simulations of 1,000 People (revised as LLM Agents Grounded in Self-Reports)
- Stanford HAI policy brief: Simulating human behavior with AI agents
- Toubia et al. (2025) Twin-2K-500 dataset
- Binz et al. (2025) A foundation model to predict and capture human cognition. Nature

- Live Science: Critique of Centaur (Liu and Ding)
- Wang, Morgenstern and Dickerson: LLMs that replace human participants can misportray and flatten identity groups
- LLM-Based Social Simulations Require a Boundary (variance review)
- EDPB: Italian authority fines the company behind Replika
- MIT Technology Review: China's business of AI avatars of the dead

CHAPTER 06

Memory for a Twin

A twin is only as good as what it remembers and how it recalls it. This chapter maps human memory onto machine parts. It walks through the techniques that exist today, from vector search to memory that tidies itself up while the twin "sleeps." It ends with the hard problems: false memories, contradictions and what to forget.

Why memory is the core of a twin

A language model on its own has two kinds of knowledge. It has what it learned in training, which is frozen and generic. And it has what is in front of it right now, in the context window. Neither is a life. A twin needs a third thing. It needs a growing, personal record that it can search, update and reason over. That is what this chapter calls memory.

The scale is large. A well known 2007 study by Matthias Mehl and colleagues estimated that people speak about 16,000 words a day on average. That is close to six million words a year from speech alone. Add messages, email, reading and what the person sees and hears. A lifetime of data, as Chapter 3 describes, runs to hundreds of millions of words plus images, audio and sensor streams. No model can hold all of it at once. So a twin must choose what to keep, how to store it and what to bring back when.

Human memory types and their machine analogs

Cognitive science splits human memory into several systems. Each has a rough match in today's AI stack. The match is not perfect. It is still a useful map for design.

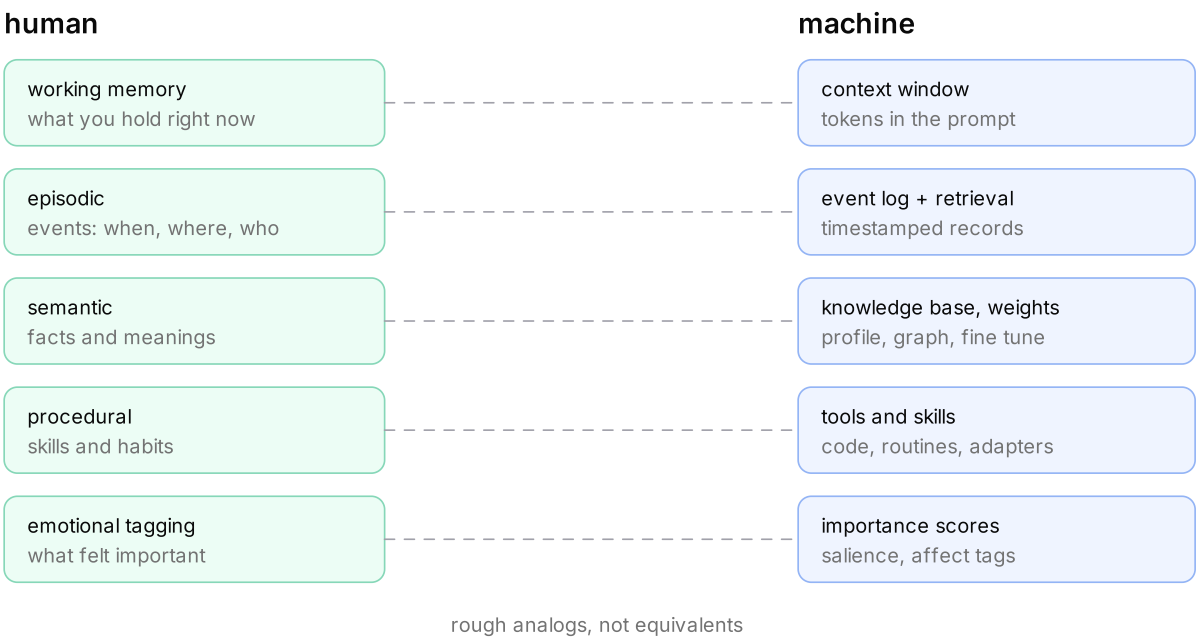


FIG 6.1 The map used in this chapter. Each human memory system on the left has a machine part on the right that plays a similar role. The analogies guide design. They do not mean the machine works like a brain.

- **Working memory to context window.** Humans hold a handful of items in mind at once. A model holds its prompt. Context windows are now very large. Some models accept a million tokens or more. But a large window is not the same as good use of it.
- **Episodic memory to an event log with retrieval.** Episodes are specific events tied to a time and place. For a twin these are timestamped records. "Tuesday 14 May, lunch with Ana at the noodle bar, she told me about her new job." The twin searches them when needed.
- **Semantic memory to a knowledge base or weights.** Facts that have lost their link to one event. "Ana is my oldest friend." "I am allergic to peanuts." These can live in a structured profile, a knowledge graph or the model's weights after fine tuning.
- **Procedural memory to tools and skills.** How to do things. For a twin this is code it can run, saved routines and small trained adapters. "How I write a thank you note." "How I plan a trip."
- **Emotional tagging to importance scores.** Humans remember emotional events better. Stress hormones and the amygdala strengthen storage. A twin can mimic this with a salience score on each memory, and with affect tags from the signals in Chapter 4.

The basic technique: retrieval augmented generation

Retrieval augmented generation, or RAG, is the workhorse. The idea is simple. Do not ask the model to remember. Look things up and paste them into the prompt.

It works in two phases. In the write phase, each piece of content is cut into chunks. Each chunk is turned into an embedding. An embedding is a list of numbers that places the text in a space where similar meanings sit close together. The chunk and its embedding go into a vector database with metadata such as date, source and people involved.

In the read phase, the current question is embedded too. The database returns the chunks whose vectors are nearest. Often a second model reranks them. The top few go into the prompt with the question. The model answers using them.

Plain RAG has known weak spots for a twin. It finds text that sounds similar to the question, which is not always what matters. It has no sense of time unless you add one. And it cannot answer questions that need many memories combined, like "How did my mood change over that year?"

Scoring memories: the Generative Agents recipe

The 2023 Generative Agents paper from Stanford, covered in Chapter 5, added a better retrieval score. Each memory in the stream gets three numbers.

- **Recency.** How long since the memory was last used. It decays exponentially. In the paper the decay factor was 0.995 per game hour.
- **Importance.** How significant the memory is. The model itself rates each new memory on a 1 to 10 scale when it is stored. Brushing teeth scores low. A breakup scores high.
- **Relevance.** How close the memory's embedding is to the current situation, by cosine similarity.

The three are scaled to the same range and added. The highest scoring memories that fit in the prompt are retrieved. This simple blend works far better than relevance alone. A strong memory from last year can beat a weak one from this morning.

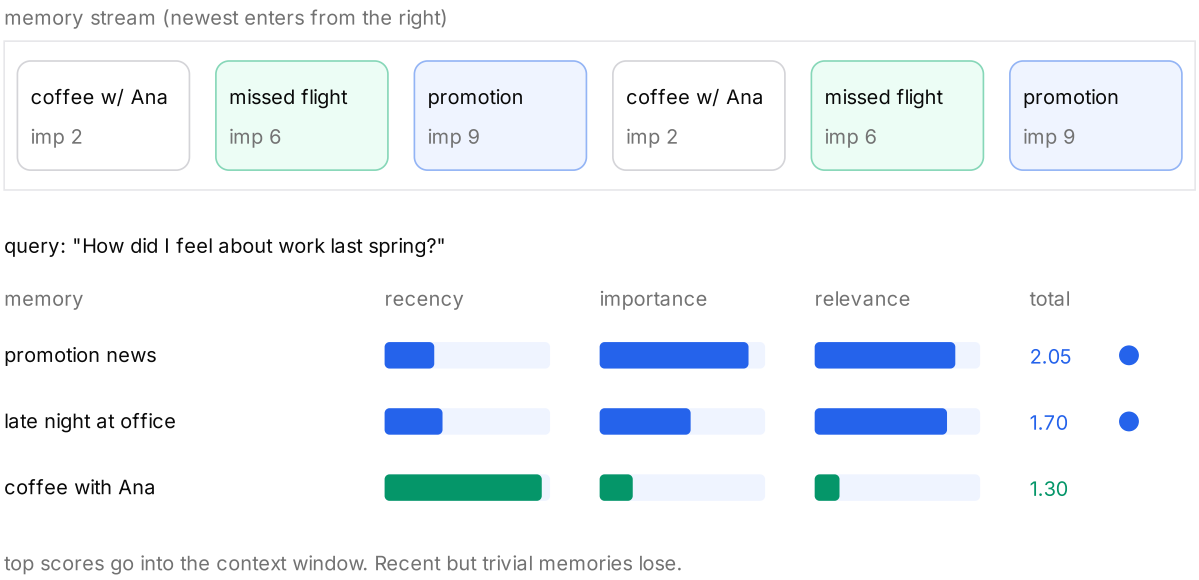


FIG 6.2 Scoring a memory stream. New memories enter with an importance score from 1 to 10. At question time each memory gets recency, importance and relevance scores. Their sum decides what the twin sees. A fresh but trivial coffee loses to an older promotion that matters.

Reflection: turning episodes into self knowledge

Generative Agents also added reflection. When the summed importance of recent memories passed a threshold, set at 150 in the paper, the agent paused. It asked itself what the most salient questions were about recent events. It retrieved memories for each question. Then it wrote short insights with pointers back to the evidence. These reflections were stored in the stream like any other memory. They could be reflected on again, which built a tree of more and more abstract beliefs.

For a human twin this is how episodic memory becomes semantic memory. Fifty separate late nights at the office become "I was overworked that spring." The pointers matter. If the twin later says "I was overworked," it can show why. If the person disagrees, the twin can see which evidence led it astray.

MemGPT and Letta: memory as an operating system

In October 2023 Charles Packer and colleagues at UC Berkeley published MemGPT. They borrowed an idea from computer operating systems. A computer has fast small memory and

slow large storage. It moves pages between them as needed. MemGPT does the same with the context window.

The context window is split into parts. There is a fixed system section. There is a small editable "core memory" block with key facts about the user and the agent's own persona. And there is a rolling queue of recent messages. Outside the window sit two stores. Recall storage holds the full conversation history. Archival storage is a searchable database for everything else.

The key move is that the model manages its own memory through function calls. It can write a new fact into core memory. It can search archival storage. When the window nears full, the system warns the model. The model then saves what matters before old messages are evicted. In September 2024 the MemGPT team launched Letta as a company and open source framework. MemGPT now names the design pattern.

Consolidation during "sleep"

Human memory does a lot of work offline. During sleep the brain replays recent experiences. A long standing theory called complementary learning systems holds that the hippocampus learns fast and the neocortex learns slowly. Replay gradually moves patterns from one to the other. That turns many specific episodes into stable general knowledge, without wiping out older knowledge.

Twin builders have started to copy this. In April 2025 Letta introduced "sleep-time compute." A second agent runs in the background between conversations. It reads recent history, reorganizes memory blocks, merges duplicates and writes cleaner summaries. The main agent stays fast because it does not do this work while talking. The companion research paper reported that precomputing ahead of time could cut the compute needed at question time by about five times on its test tasks, with similar accuracy.

For a twin, a nightly sleep cycle could do several jobs. Summarize the day. Score and tag new memories. Run reflection. Find contradictions with older memories. Update the profile. Decay what was not used. Each of these is cheaper and safer to do in batch than in the middle of a live conversation.

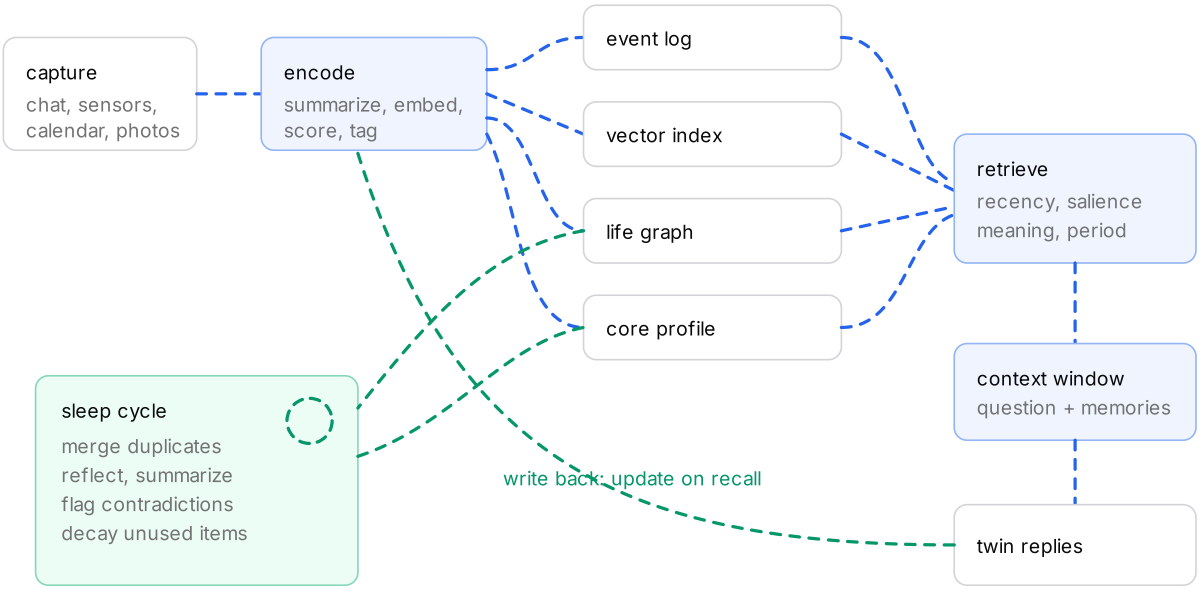


FIG 6.3 A full memory pipeline for a twin. Inputs are encoded, scored and stored in four places. At question time a retriever blends recency, importance, relevance and time and fills the context window. After the reply, recalled memories are updated. Offline, a sleep cycle merges, reflects, checks for contradictions and lets unused items fade.

Long context: why not just load everything?

Context windows grew fast. In 2024 Google showed Gemini 1.5 with a window of a million tokens. That is roughly 700,000 words, or several novels. It is tempting to skip retrieval and paste the whole life in.

There are three problems. First, the numbers still do not fit. A year of speech alone is close to six million words. Second, cost and speed grow with every token in the prompt, on every question. Third, models do not use long contexts evenly. A 2023 study by Nelson Liu and colleagues, "Lost in the Middle," found that models used information at the start and end of a long prompt much better than information in the middle.

Benchmarks built for long term memory confirm it. LongMemEval, released in 2024, tests chat assistants on 500 questions about long histories. It checks five skills: pulling out facts, reasoning across sessions, handling updated facts, reasoning about time and knowing when to say "I don't know." Its authors reported that commercial chat assistants and long context models showed about a 30 percent accuracy drop when information had to be remembered across sustained interactions. Long context is a useful tool. It is not a memory system.

~16,000
words spoken per day, average (Mehl et al. 2007)

~6M
spoken words per year, before any text or images

~30%
accuracy drop for assistants on sustained memory (LongMemEval)

150
importance sum that triggered reflection in Generative Agents

Knowledge graphs of a life

Vectors are good at "find something like this." They are bad at "who was my manager when I lived in Berlin?" That needs structure. A knowledge graph stores entities as nodes, such as people, places, jobs and projects. It stores relations as edges, such as "worked at," "lived in," "friends with."

For a twin the most important feature is time. Facts about a life change. A temporal knowledge graph puts a valid from and valid to date on each edge. The Zep team's Graphiti engine, described in a January 2025 paper, works this way. When a new fact conflicts with an old one, the old edge is not deleted. It is closed with an end date. So the twin can answer both "where do I work?" and "where did I work in 2018?" Zep reported gains over full context baselines on LongMemEval, especially on questions about time and preferences. Mem0, another memory layer from 2025, extracts and consolidates key facts from conversations and offers a graph variant. Its authors reported large token and latency savings against full context prompting on the LOCOMO benchmark. Note that both sets of results come from the vendors.

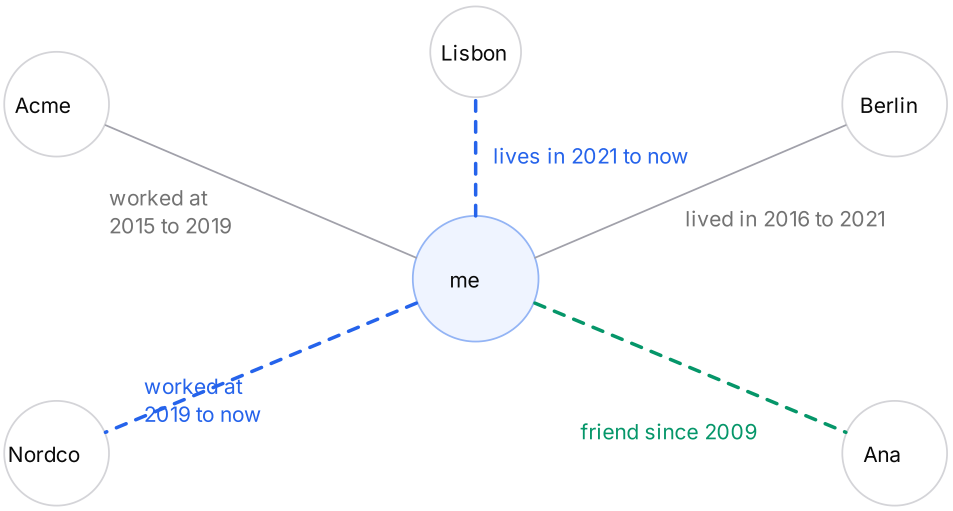


FIG 6.4 A temporal life graph. Old facts are not erased. Their edges get end dates. Current facts flow in the accent color. This lets the twin answer questions about now and about any earlier version of the person.

Keeping a timeline and a sense of self

Humans organize memory around a life story. Psychologists call this autobiographical memory. It has layers. There are lifetime periods like "my university years." Inside them are general events like "Friday dinners with the team." Inside those are specific moments.

A twin can copy this structure. Keep a list of life chapters with dates. Link each memory to its chapter. Let reflection write chapter summaries. Then when the twin answers, it can ask which self is answering. The 2019 self may have liked the job. The 2024 self quit it. Without this, the twin blends both into a person who never existed. Chapter 5 called this missing growth.

A sense of self also needs a small, stable core. MemGPT's core memory block is one version. It holds the few facts that should always be in view. Name, values, key relationships, current situation, how the person talks. This core should change slowly and only with good reason. Ideally the person approves each change.

Forgetting on purpose

Perfect memory is not a feature for a twin. People forget most of what happens to them. That shapes who they are. A twin that recalls every trivial detail will feel wrong. It will also waste resources and raise privacy risk.

Hermann Ebbinghaus measured forgetting in the 1880s by testing himself on nonsense syllables. Retention fell steeply at first and then leveled off. Each review made the next fall slower. That is the idea behind spaced repetition apps today. A twin can use the same math. Give each memory a strength. Let it decay over time. Slow the decay for important or emotional memories. Boost strength each time the memory is recalled. When strength falls below a line, archive the memory or delete it.

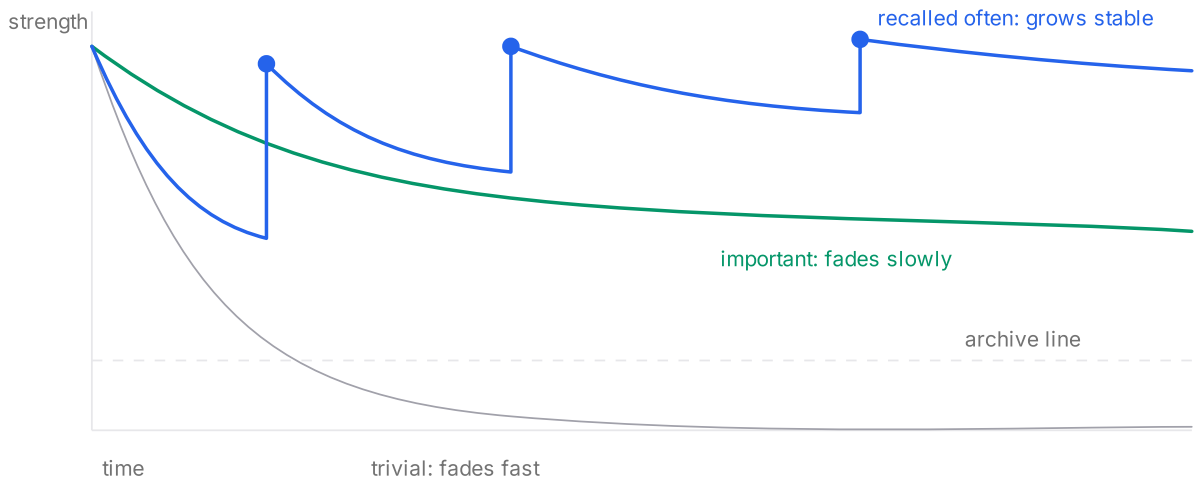


FIG 6.5 Decay functions for a twin's memories. Trivial items drop below the archive line quickly. Important items fade slowly. Each recall, shown as a pulsing dot, resets strength and makes the next decline gentler. This is Ebbinghaus's curve put to work.

Reconsolidation: changing a memory when it is recalled

In 2000 Karim Nader, Glenn Schafe and Joseph LeDoux showed something surprising in rats. A fear memory that had been recalled became fragile again for a short time. Blocking protein synthesis in that window weakened the memory. This is called reconsolidation. Recall is not just reading a file. It reopens the file and saves a new version.

Twin memory can use a careful version of this. When a memory is retrieved and the conversation adds new facts, update it. If the person says "Actually that dinner was in March, not April," fix the record. If the person now sees an old event differently, add the new view with a date. Keep the original too. The lesson from human reconsolidation is also a warning. Each rewrite is a chance to add error. Human memories drift with retelling. A twin should keep raw sources untouched and only rewrite its summaries.

Problems that are not solved

FALSE MEMORIES

Language models fill gaps with plausible text. In a twin this becomes a false memory. "We met in Paris" when it was Rome. Psychologist Elizabeth Loftus showed that humans also absorb false details from leading questions. Fixes: every memory links to a source. The twin cites sources for factual claims about the past. It says "I don't remember" when retrieval finds nothing. LongMemEval tests this abstention skill directly.

CONTRADICTIONS

Lifelong data disagrees with itself. Diaries say one thing and messages another. People change their minds. Fixes: keep both with dates and sources. Decide which is current using the temporal graph. Flag hard conflicts for the person during the sleep cycle instead of silently picking one.

PRIVACY

A twin's memory holds other people's lives too. Friends' secrets, family health, a partner's messages. Those people did not consent. Fixes: tag memories by who they involve. Restrict what the twin can say about third parties. Encrypt at rest. Keep the store on hardware the person controls where possible.

DELETION

Laws like the EU GDPR give a right to erasure. Deleting a row from a database is easy. Removing a fact from fine tuned weights is not. Research on machine unlearning is active but no method is fully reliable. Fix: keep facts in external stores that can be deleted. Use fine tuning for style, not secrets.

WHAT TO FORGET

This is a values question, not only a technical one. Should a twin of a grieving person keep every detail of the worst week? Should it remember a cruel thing a friend said years ago, long after the friendship healed? Humans soften such memories over time, and that softening is often healthy. A twin could offer explicit controls. "Fade this." "Keep but never bring up."

"Delete." The person should decide. When the twin outlives the person, someone else will have to decide. Chapter 12 returns to that.

TECHNIQUE	BEST FOR	WEAKNESS	EXAMPLE
Vector RAG	Finding similar text	No time sense, shallow	Any vector database
Scored memory stream	Picking what matters now	Importance scores are model guesses	Generative Agents
Reflection	Building self knowledge	Can lock in wrong beliefs	Generative Agents
Self managed tiers	Long chats, stable core	Model may save the wrong things	MemGPT, Letta
Sleep time processing	Cleaning and merging	Extra compute, needs checks	Letta sleep-time agents
Temporal graph	Facts that change over time	Extraction errors, schema work	Zep Graphiti, Mem0 graph
Long context	One big document or session	Cost, lost in the middle	Million token models
Fine tuned weights	Style and habits	Hard to update or delete	LoRA adapters

A practical memory design for a twin

Putting the pieces together gives a design that can be built now with open tools.

- **Raw archive.** Keep every source untouched with timestamps. This is the ground truth. Nothing rewrites it.
- **Event log.** One line per episode in plain language with time, place, people, source link, importance and affect tags.
- **Vector index** over the log and raw text for meaning search.
- **Temporal life graph** for people, places, roles and how they changed.
- **Core profile.** A short, person approved description that is always in context.

- **Retriever** that blends recency, importance, relevance and time period, and asks which life chapter is speaking.
- **Sleep cycle** every night: summarize, reflect, merge, flag contradictions, decay.
- **Controls** for the person: inspect any memory, correct it, fade it, delete it.

None of this makes the twin remember the way a person does. Human memory is reconstructive, emotional and tied to a body. Machine memory is a set of databases and scores. But a well designed set can do the jobs that matter. It can recall the right thing at the right time. It can know what changed. It can admit what it does not know. And it can let go of what no longer serves the person it copies.

SOURCES

- Park et al. (2023) Generative Agents: Interactive Simulacra of Human Behavior
- Packer et al. (2023) MemGPT: Towards LLMs as Operating Systems
- Letta: Sleep-time compute (April 2025)
- Letta: MemGPT and Letta
- Wu et al. LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory
- Liu et al. (2023) Lost in the Middle: How Language Models Use Long Contexts
- Rasmussen et al. (2025) Zep: A Temporal Knowledge Graph Architecture for Agent Memory
- Chhikara et al. (2025) Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory
- Nader, Schafe and LeDoux (2000) Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature

CHAPTER 07

Voice, Face and Digital Humans

A twin that thinks like you still needs to sound and look like you. This chapter shows how voices and faces are copied today, what data each step needs, and why the same tools also power fraud.

People know each other first by voice and face. A friend can tell it is you from one word on the phone. A parent can spot a fake smile in a photo. So the outer layer of a twin carries a lot of weight. If it is wrong the whole twin feels wrong, however good the mind behind it.

Three things changed between 2021 and 2026. Voices became cheap to copy from a few seconds of audio. Faces could be animated from a single photo. And speech models learned to talk and listen at the same time, like people do. Each of these rests on a simple idea: turn the signal into tokens or codes, then let a large model learn to predict them. This chapter walks through each piece in turn.

How voice cloning works

A voice has two parts. One part is what you say: the words and their order. The other part is how you sound: the pitch of your voice, its texture, your accent, your speed, the way you stretch some vowels. Voice cloning keeps the second part and lets a machine supply the first.

STEP 1: A FINGERPRINT OF THE VOICE

The oldest trick is the speaker embedding. A small network listens to many thousands of speakers and learns to tell them apart. Its last hidden layer becomes a short list of numbers, often a few hundred long. Two clips of the same person land close together in this number space. Two different people land far apart. That list is a fingerprint of how you sound. Systems called d-vectors and x-vectors work this way, and so do newer encoders. A speech generator can then be told: say this text, and make it land near this fingerprint.

STEP 2: AUDIO AS TOKENS

The bigger leap came from neural audio codecs. Google's SoundStream (2021) and Meta's EnCodec (2022) learned to squeeze audio into a stream of discrete codes and then rebuild it. Each slice of sound becomes a small stack of code numbers. The first code in the stack

carries the coarse shape. Later codes add fine detail. A few seconds of speech becomes a few hundred tokens, much like a sentence of text.

Once audio is tokens, a language model can learn to predict it. This is what Microsoft's VALL-E did in January 2023. It was trained on about 60,000 hours of English speech. You give it a 3 second clip of a stranger plus some text. It continues the token stream so the new words come out in that stranger's voice. It even copied the room echo and the emotion of the prompt. The authors called it zero shot because the model never trained on that speaker. VALL-E 2 followed in 2024. Microsoft reported that on two standard test sets its output matched or beat real recordings for naturalness and speaker likeness. Microsoft did not release either model to the public and said the risk of misuse was the reason.

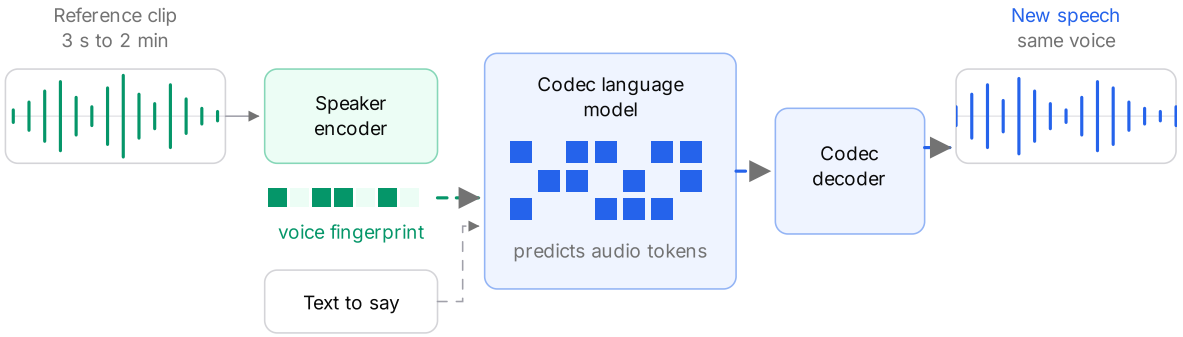


FIG 7.1 A modern voice clone. A short clip becomes a voice fingerprint. A language model predicts audio codec tokens for new text in that voice. A decoder turns the tokens back into sound.

STEP 3: FINE TUNING FOR A DEEP COPY

Zero shot copies are fast but shallow. They catch the timbre. They often miss the habits that make you sound like you: how you laugh mid sentence, where you pause, how your voice drops when you are tired. For that, the model is fine tuned on a longer set of your recordings. ElevenLabs shows the split clearly. Its Instant Voice Clone works from one to two minutes of clean audio and does not retrain the model. Its Professional Voice Clone trains on your data. The company says 30 minutes is the bare minimum and recommends two to three hours for the best result. It also requires a check that the voice is really yours.

3 s
prompt used by VALL-E research (2023)

1 to 2 min
clean audio for an instant commercial clone

2 to 3 h
recommended for a trained professional clone

~200 ms
real time reply latency of Moshi in practice

What has been built

COMMERCIAL

ELEVENLABS

The best known voice cloning service. Instant and trained clones, many languages, and a library where people can share or sell their voice. Trained clones require proof that you own the voice.

RESEARCH

MICROSOFT VALL-E AND VALL-E 2

Showed that a codec language model can copy a voice from 3 seconds. Kept as research only because of misuse risk.

RESEARCH

OPENAI VOICE ENGINE

Shown in March 2024. Made speech from a 15 second sample. OpenAI chose a small preview with partners and did not release it widely.

OPEN

OPEN MODELS

XTTS, OpenVoice, F5-TTS, CosyVoice, Fish Speech and others let anyone clone a voice on a home GPU. Quality varies by language and by how clean the sample is.

OPEN

SESAME CSM

A 2025 conversational speech model that uses the dialogue so far to choose tone. Its 1B version was released openly.

OPEN

KYUTAI MOSHI

A 2024 full duplex speech model that listens and speaks at once. Weights and code are public.

Real time talk: the hard part is timing

Reading a script in your voice is one thing. Holding a conversation is another. In human talk the gap between one person stopping and the next starting is very short. Across ten languages, studies put the typical gap at around 200 milliseconds. That is faster than anyone can plan a reply from scratch. People start planning while the other person is still talking. They predict the end of the turn.

Older voice assistants worked like walkie talkies. They waited for silence, sent the audio to a speech recognizer, sent the text to a language model, then sent the reply to a speech generator. Each step added delay. The pause before a reply was often well over a second. Worse, the pipeline threw away tone. The text model never heard that you sounded annoyed.

Two changes fixed much of this. First, single models began to take audio in and put audio out with no text step between. OpenAI said in May 2024 that GPT-4o could answer speech in as little as 232 ms with an average of 320 ms. Second, full duplex models appeared. Kyutai's Moshi models two audio streams at once: the user's and its own. It always listens and always outputs sound, even if that sound is silence. Its authors report a theoretical latency of 160 ms and about 200 ms in practice.

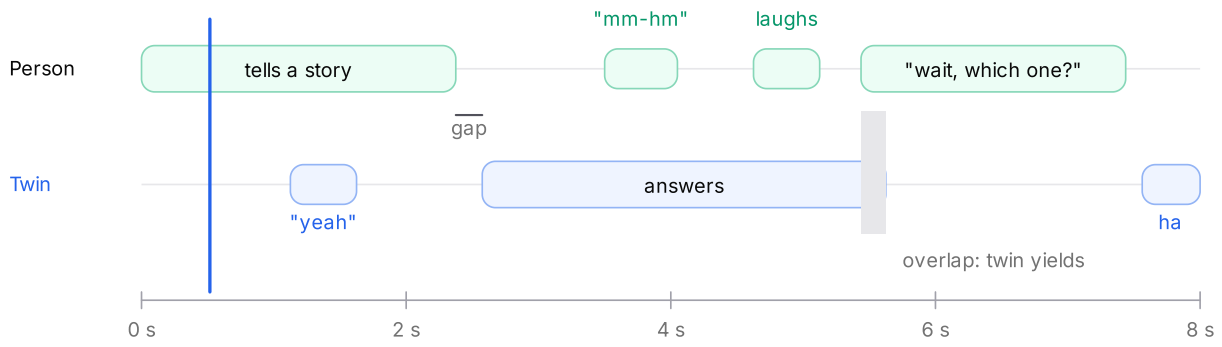


FIG 7.2 Real talk is not tidy turns. Both sides make small sounds while the other speaks. A short gap separates turns. When the person cuts in, a good twin stops. A full duplex model handles all of this because it hears and speaks at the same time.

Timing is only part of it. Real people fill talk with small signals. "Mm-hm" and "right" tell the speaker to keep going. These are called backchannels. Laughter marks a joke or eases a tense moment. Interruptions can be rude, or they can show close attention. A twin that never makes these sounds feels like a recording. A twin that makes them in the wrong place feels worse.

Sesame's 2025 study gives a useful measure of where things stand. Listeners heard short clips with no context. They could not reliably pick the human recording over the model. When they heard the same clips inside a real conversation, they still preferred the human. The words and the voice were right. The way the voice fit the moment was not. That gap is exactly what a personal twin must close. It is also why recordings of you in real conversation are worth far more than recordings of you reading aloud.

FOR YOUR TWIN

Record real conversations with consent, not just read speech. Keep both sides of the audio on separate tracks if you can. That is the data that teaches turn taking, backchannels and how your laugh fits into a sentence.

Talking heads from one photo

The face side moved just as fast. There are two broad families. The first is 2D video generation. It makes a flat video of a face that talks. The second is 3D avatars. They build a model of the head that can be lit, turned and viewed from any angle.

The 2D family is now a large business. Synthesia, HeyGen and D-ID let a user turn text into a talking presenter video. For a custom avatar of a real person, these services ask for a short consent recording and a few minutes of video of the person talking to camera. The output is good for a corporate training video. It is weaker in long, emotional or interactive use.

Research systems go further. Microsoft's VASA-1 (April 2024) takes one portrait photo and a speech clip. It produces video at 512 by 512 pixels at up to 40 frames per second, fast enough to run live. The key idea is to separate the face into parts. One part is identity and appearance, taken from the photo. The other part is motion: head pose, lip shape, gaze, blinks and expression. A diffusion model learns to generate motion from the audio. A renderer then paints the face from the photo and the motion. Microsoft did not release VASA-1 either.

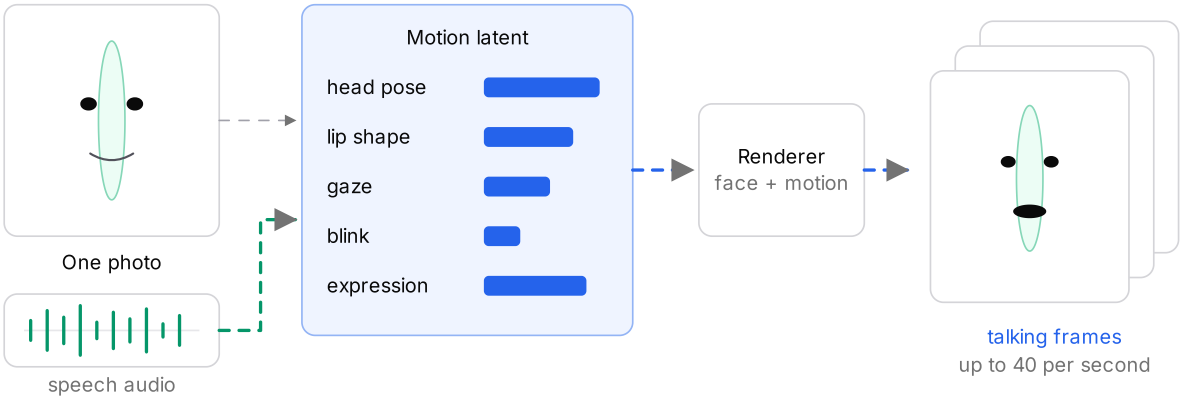


FIG 7.3 How a one photo talking head works, in the style of VASA-1. Appearance comes from the photo. Motion comes from the audio and lives in a compact set of numbers. A renderer joins them frame by frame.

Capturing a person in 3D

For a twin that you can walk around in VR, or that a robot face could copy, you need a 3D model. That starts with capture.

LIGHT STAGES AND PHOTOGRAMMETRY

The light stage was developed by Paul Debevec and his team, first at UC Berkeley and then at the USC Institute for Creative Technologies, from about 2000. It is a dome of controlled lights and cameras around a seated person. The lights flash in fast patterns. From how the skin looks under each light, software works out the shape of the face down to fine wrinkles. It

also works out how the skin reflects light: what is oily, what scatters light below the surface. Film studios have used this kind of capture to make digital doubles of actors.

Photogrammetry is the simpler cousin. Many cameras fire at once from many angles. Software matches points between photos and builds a 3D mesh with a colour texture. Rigs with dozens of cameras can capture a full body in a split second. A phone app can do a rough version if you walk slowly around a still subject.

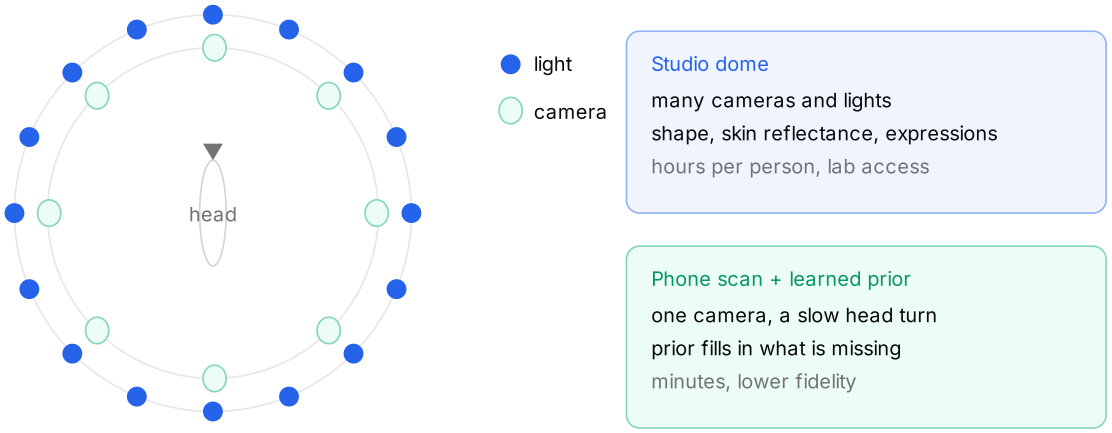


FIG 7.4 Top view of a capture dome. Lights flash in patterns while cameras record. The dome yields a precise model of one person. Newer methods train a prior on hundreds of dome captures, then personalise it from a short phone scan.

CODEC AVATARS AND GAUSSIAN SPLATS

Meta's Codec Avatars project has worked on this problem for years, with the goal of lifelike avatars for VR calls. It captures people in large camera domes. It then trains a network that encodes a face into a compact code and decodes it into a 3D head from any view. A headset with inner cameras can read your expression and send just the code. The other side rebuilds your face.

In 2023 a new 3D format took over much of this work: 3D Gaussian splatting. Instead of a mesh of triangles, the scene is a cloud of many small, soft, coloured blobs. Each blob has a position, a size, an orientation and a colour that can change with view angle. They can be rendered very fast on a GPU and look photographic. Avatar researchers attached these blobs to a face model so they move with the jaw and brow. Meta's URAvatar work (2024) used this to build a relightable head avatar from a single phone scan. It does this by starting from a prior learned from hundreds of people captured in a dome. Your scan only has to adjust that prior.

NVIDIA built Audio2Face to drive a 3D face from speech audio. It predicts the motion of lips, jaw, tongue and cheeks frame by frame, and a related model estimates emotion from the audio. It is part of NVIDIA's ACE toolkit for game characters and digital agents. In September 2025 NVIDIA released the Audio2Face models, SDK and training framework as open source. Epic Games' MetaHuman tools take a different route. They let a creator build a realistic 3D human and animate it from video, including from a phone.

Gesture and body language

Most of what a face does in talk is timed to speech. So are the hands. People beat time with their hands on stressed words. They point. They shrug. They lean in. Each person has a style. Some barely move. Some talk with their whole body.

To copy this you need motion data. Optical motion capture uses reflective markers and infrared cameras. It is precise but needs a studio. Inertial suits use small motion sensors on each limb and work anywhere, but drift over time. Markerless capture estimates pose from plain video. It is less precise, yet it can use footage you already have.

Generation then works much like the face. A model learns to map speech audio and text to body motion. Meta's Audio2Photoreal research (2024) generated face, body and hand motion for a photoreal avatar from the audio of a two person conversation. It showed the avatar reacting while listening, not only while talking. The GENEA challenge has run shared tests of gesture generation since 2020. A steady finding is that models now produce motion that looks natural, yet it often fails to match the meaning of what is said. The hand moves at the right time but means nothing.

The uncanny valley, briefly

The closer a face gets to human, the more small errors hurt. A cartoon can have dead eyes and still charm. A near perfect face with dead eyes is disturbing. Chapter 8 covers Mori's theory in detail. For digital faces the usual culprits are well known. Eyes that do not focus on anything. Teeth and tongue that look painted. A mouth that is a few frames early or late. Skin that does not flush or wrinkle. Blinks on a fixed timer. Motion that never quite stops, or stops too fully. Fixing these takes data from the specific person, not just better general models.

What data you need, and how long it takes

TARGET	MINIMUM THAT WORKS	WHAT A GOOD COPY NEEDS	TYPICAL TIME
Voice timbre	3 to 30 seconds (research and open zero shot models)	1 to 2 minutes clean, single speaker	Minutes
Voice with your habits	30 minutes for a trained clone	2 to 3 hours, varied emotion, plus real conversations	Hours of recording, hours of training
Conversation style	Not possible from read speech	Many hours of two sided dialogue with separate tracks	Weeks to collect
2D talking head	One photo (research)	A few minutes of studio video plus a consent clip	Under a day
3D relightable head	A phone scan with a strong prior (research)	Dome capture under many lights and expressions	Hours in a lab
Body and gesture style	Video of you talking	Motion capture sessions while you talk freely	Several sessions

Two points matter more than the numbers. First, quality depends on clean, varied data. Ten minutes of you happy, tired, angry and laughing beats an hour of you reading a book. Second, the person must be present. Every serious commercial service now asks the subject to record a consent statement on camera or in voice. Plan for that from the start.

Deepfakes: the same tools, used against people

Everything above can copy someone who did not agree. This is not theory. In January 2024 a robocall used a cloned voice of the US president to tell voters in New Hampshire not to vote. Weeks later the US Federal Communications Commission ruled that AI generated voices in robocalls fall under existing limits on artificial voices. In early 2024 staff at the engineering firm Arup in Hong Kong were tricked into transferring about 25 million US dollars after a video call in which the other "colleagues" were deepfakes. Voice cloning is now a common tool in phone scams that imitate family members in trouble.

Detection is an arms race. Detectors look for traces that generators leave: odd patterns in high frequencies, breath that does not fit, lighting that does not match. They work well on the fakes they were trained on and less well on new generators. People are not reliable detectors either. So the field is shifting from spotting fakes to proving what is real.

PROVENANCE

C2PA CONTENT CREDENTIALS

An open standard that attaches a signed record to a file saying how and by what tool it was made. It only helps if cameras, apps and platforms keep the record attached.

WATERMARK

INVISIBLE MARKS

Google's SynthID and Meta's AudioSeal hide a signal inside generated media that a detector can read later. Strong edits can weaken them.

LAW

NEW RULES

The US TAKE IT DOWN Act (2025) targets non consensual intimate images, including AI made ones. The EU AI Act requires deepfakes to be labelled, with those transparency duties applying from August 2026.

CONSENT FOR A TWIN

If you build a twin of yourself, write down who may use the voice and face, for what, and until when. Add a watermark to everything it makes. Decide now what happens to the model when you die. A clone that can speak for you can also be stolen and made to say anything.

What this means for a human twin

The outer layer of a twin is the most mature part of the whole project. A convincing voice is within reach of anyone with a few hours of recordings. A convincing talking face from a

short video is a commercial product. A real time, full duplex voice twin that laughs and backs off when you interrupt now exists in research and in open models.

The gaps are subtle but real. Current systems copy how you sound much better than how you respond. They miss your timing in a heated moment, your particular way of teasing, the look you give when you disagree but stay polite. Those come from the mind and emotion layers of the twin, covered in Chapters 5, 6 and 9. They then have to drive the voice and face in sync. Today most systems treat voice, face and body as separate models glued together. A twin that feels like one person will need them trained as one.

SOURCES

- Wang et al. 2023, Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers (VALL-E)
- Chen et al. 2024, VALL-E 2: Neural Codec Language Models are Human Parity Zero-Shot Text to Speech Synthesizers
- ElevenLabs docs, Professional Voice Cloning
- Défossez et al. 2024, Moshi: a speech-text foundation model for real-time dialogue
- Sesame 2025, Crossing the uncanny valley of conversational voice
- Xu et al. 2024, VASA-1: Lifelike Audio-Driven Talking Faces Generated in Real Time
- Li et al. 2024, URAvatar: Universal Relightable Gaussian Codec Avatars
- Kerbl et al. 2023, 3D Gaussian Splatting for Real-Time Radiance Field Rendering
- NVIDIA Audio2Face-3D SDK on GitHub
- C2PA, Coalition for Content Provenance and Authenticity

CHAPTER 08

Embodied Twins and Robots

A screen twin can talk. A physical twin can sit across a table, shake a hand and take up space in a room. This chapter looks at the robots that have tried to copy humans, what they taught us, and how far real bodies still lag behind.

Most digital twins will live on screens and in earbuds. Yet the dream of a copy you can touch is old, and some people have built one. The results are strange and useful. They show what a body adds to a person, and how much of being human is hidden in small motions we never notice.

This chapter covers four threads. First, robots built to look like one specific person. Second, expressive robot faces. Third, the wave of general humanoid robots from 2024 to 2026. Fourth, the deeper question of why a body matters for a mind at all.

Geminoid: a robot copy of its maker

In 2006 the Japanese roboticist Hiroshi Ishiguro unveiled Geminoid HI-1. It was a robot built as a copy of himself. Its silicone skin was cast from his body. Its hair was real human hair. It sat in a chair and moved its head, eyes, lips and chest using air powered actuators. It was not autonomous. An operator in another room watched through cameras and spoke through a microphone. The robot moved its lips with the operator's voice and copied some head motion. The project was run by Ishiguro's lab with ATR in Kyoto. Later versions copied other people, including Geminoid F, a woman, and Geminoid DK, modelled on the Danish professor Henrik Schärfe.

Geminoid was a twin of the body only. The mind was a human at a remote desk. That is exactly why it is useful. It isolates what a physical form adds when the mind is held constant.

WHAT GEMINOID TAUGHT

- **Small motion is life.** A still android looks like a corpse. The team added constant tiny motions: breathing, blinks, small shifts of posture and gaze. These idle motions did more for realism than better skin.
- **People adapt fast.** Visitors were often uneasy at first. After a few minutes of talk many treated the robot much like the person it copied.
- **Operators feel the body as theirs.** Studies with Geminoid operators found that after a short time of controlling it, some felt a sense of touch when someone poked the robot's cheek, even though

there were no touch sensors. Researchers called this body ownership transfer. The brain adopts a body that moves in sync with its intentions.

- **The copy and the person drift.** A silicone copy does not age. The person does. Ishiguro has spoken about how his own face moved away from the robot's over the years. A physical twin freezes one moment of a life.

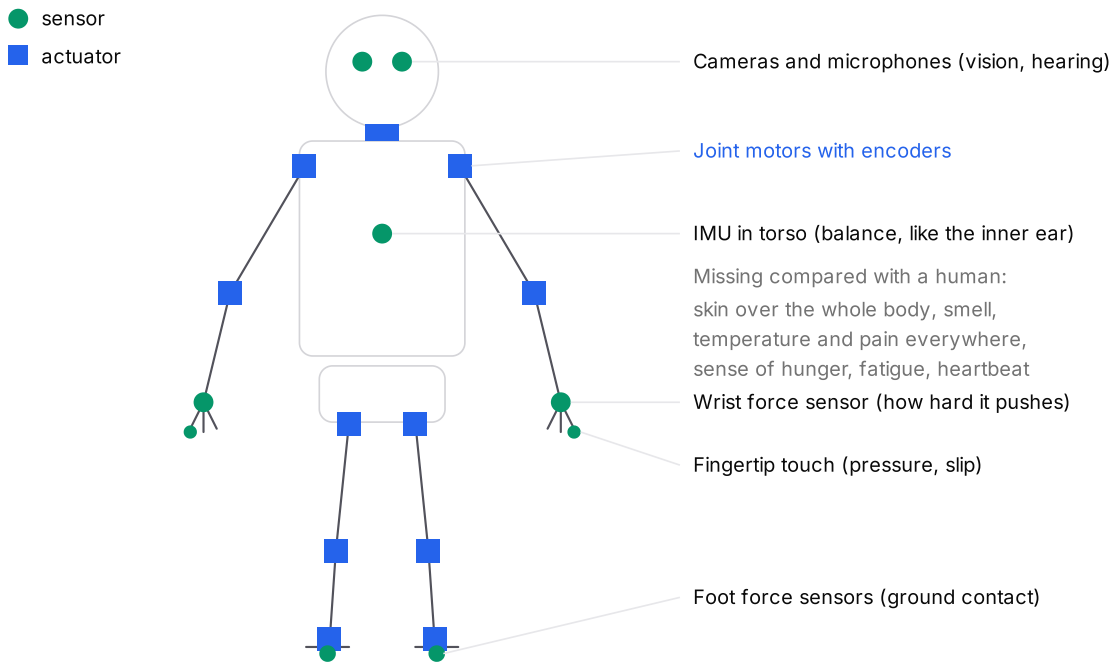


FIG 8.1 The body plan of a typical 2025 humanoid. Square marks are motors at joints. Round marks are sensors. Most of the body has no touch at all, and there is no inner sense of the body's own state beyond battery and motor temperature.

Expressive robot faces

HANSON ROBOTICS' SOPHIA

Sophia, from Hanson Robotics, became the most famous robot face of the late 2010s. It has a rubber like skin and many small motors under it. It appeared on talk shows and at conferences. In 2017 Saudi Arabia announced that it had granted Sophia citizenship, a publicity move that drew wide criticism. AI researchers, including Meta's chief AI scientist Yann LeCun, criticised the way Sophia was presented. Many of its talks were scripted or guided, while the public was led to think it understood the conversation. The lesson for twin builders is plain. A face that looks alive invites people to assume a mind behind it. That trust can be abused.

ENGINEERED ARTS' AMECA

Ameca, from the UK firm Engineered Arts, was first shown at the end of 2021 and became widely known at CES 2022. It has a grey, clearly robotic face on purpose, which avoids the trap of near human skin. Its expressions are fast and fluid. In recent years it has been paired with large language models for conversation, so it can answer questions and react with a matching face. It is a demonstration and research platform. It does not walk.

COLUMBIA'S EMO: A FACE THAT PREDICTS YOURS

In 2024 Columbia University's Creative Machines Lab published Emo in Science Robotics. Emo is a robot head with soft silicone skin, 26 actuators and cameras inside its pupils. The team trained two models. One learned how Emo's own motors change its face, partly by letting it make random faces in front of a camera. The other watched human faces and learned to predict an expression from its first small changes. Emo could anticipate a smile about 840 milliseconds before the person smiled and smile with them at the same moment.

That timing matters. A smile that comes after yours feels like copying. A smile that comes with yours feels like shared feeling. For a twin, Emo shows that good expression is not just about looks. It is about prediction.

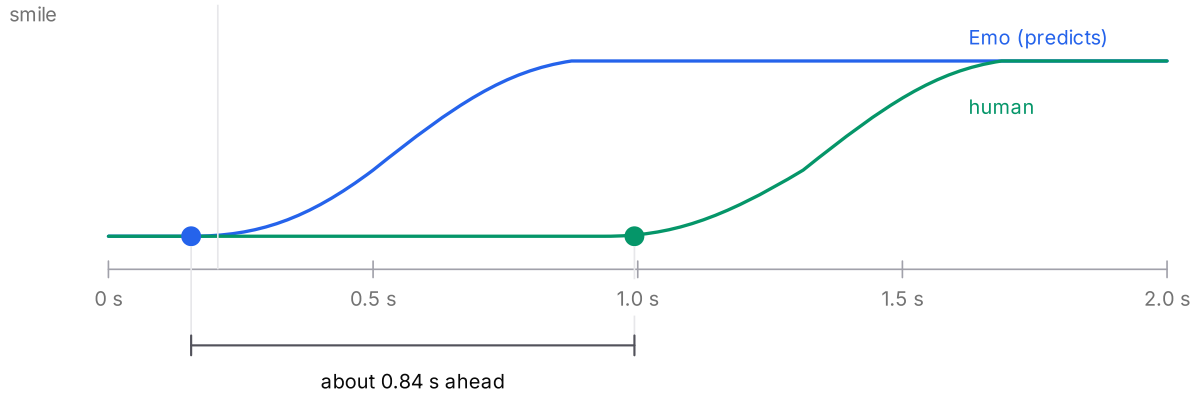


FIG 8.2 Co-expression, as in the Columbia Emo study. The robot reads early cues in a face and starts its own smile so that both peak together. Curves are schematic. The 0.84 s lead is the figure the team reported.

The humanoid wave, 2024 to 2026

Geminoid and Ameca were built to be seen. A new wave of humanoids is built to work. Since 2024 billions of dollars have gone into robots shaped like people. The argument is that our

world is built for human bodies, so a human shaped robot can use our tools, stairs and shelves. For a twin, these machines are the most likely future body. Here is where the main players stood as of late 2026, based on public reports. Many claims come from the companies themselves and are hard to check.

ROBOT	MAKER	WHAT IS KNOWN	STATUS NOTE
Figure 02 and 03	Figure AI (US)	Figure 02 worked about 11 months at BMW's Spartanburg plant moving sheet metal parts. Figure says it helped build over 30,000 cars. Figure 03 was shown in October 2025. Runs Figure's Helix vision language action model.	Factory work is real. Volumes and targets are company reported.
Optimus	Tesla (US)	Shown in several versions since 2022. At an October 2024 event, robots served drinks and talked, but reports showed they were partly remote controlled by humans.	Tesla said in 2026 it was converting Fremont lines to build Optimus. The real ramp was slow and hard to verify.
Atlas (electric)	Boston Dynamics (US, owned by Hyundai)	The hydraulic Atlas was retired in April 2024 and replaced by an all electric design. A production version was shown at CES in January 2026.	2026 units are committed to Hyundai and Google DeepMind. Wider sales expected later.
NEO	1X (Norway and US)	A soft bodied home humanoid. Preorders opened October 2025 at 20,000 US dollars or 499 dollars a month. For tasks it cannot do alone, a remote human operator can take over with the owner's approval.	Home deliveries were planned for 2026. Scale and autonomy remain open questions.
H1, G1, R1	Unitree (China)	Sells the cheapest capable humanoids. G1 launched in 2024 at about 16,000 dollars. R1 was announced in 2025 at about 5,900 dollars.	Widely used by research labs. Agile, but tasks need custom programming.
Digit	Agility Robotics (US)	A legged robot with simple arms built for moving totes. Deployed in GXO logistics warehouses under a	One of the first paid humanoid jobs. Narrow tasks.

ROBOT	MAKER	WHAT IS KNOWN	STATUS NOTE
		robots as a service deal from 2024.	

WATCH THE DEMOS

Many humanoid videos are sped up, edited, or quietly remote controlled. Ask three questions of any demo. Is it autonomous? Is it real time? How many tries did it take? The honest answer is often "partly, no, many".

Touch and robot skin

The human hand has about 17,000 touch receptors in its smooth skin alone. They sense pressure, vibration, stretch and slip. That is why you can hold an egg without looking. Most robot hands feel almost nothing. They rely on cameras and on the current in their motors.

Robot touch research has three main approaches. The first is vision based fingertips. A soft gel pad has a camera behind it. When the pad presses on something the camera sees how it deforms. GelSight sensors work this way, and Meta with GelSight showed a fingertip called Digit 360 in 2024. The second is flexible electronic skin: thin sheets of pressure sensors that can wrap a limb. Coverage is hard because of wiring. The third is smart materials. In 2025 a team from Cambridge and UCL cast a whole hand shaped skin from one conductive gel. With only 32 electrodes at the wrist, a learned model could tell pressure, heat and even a cut apart from changes across the material. The authors were clear that it is not yet as sensitive as human skin.

For a twin, touch matters in two ways. It lets a body handle the world safely. And it gives the emotion engine of Chapter 9 a signal for comfort, pain and closeness. A hug means nothing to a body that cannot feel it.

Learning to move like a person

A physical twin should move like you, not like a generic robot. Your walk, how you gesture, how you pick up a cup. How can a robot learn this?

TELEOPERATION AND IMITATION LEARNING

The main method today is imitation learning. A human controls the robot remotely, using a matching set of arms, a VR headset, or a motion capture suit. The robot records what its

cameras see and what its joints do. After many demonstrations a neural network learns to map what it sees to what it should do. Stanford's Mobile ALOHA (2024) showed that about 50 demonstrations per task could teach a wheeled two armed robot to cook a shrimp, call a lift and wipe up a spill, with success rates that rose when the robot also trained on data from other tasks.

Since then, robot foundation models have grown. These are vision language action models. They take camera images and a spoken instruction and output motor commands. Examples include Physical Intelligence's π0, NVIDIA's GR00T N1, Google DeepMind's Gemini Robotics and Figure's Helix. They are trained on mixed data from many robots plus web video.

LEARNING FROM HUMAN VIDEO

Teleoperation is slow and costly. Human video is everywhere. Researchers now extract hand and body poses from video and retarget them to robot joints. Head mounted cameras record first person views of daily tasks. This is useful for a twin. Your own lifelog video from Chapter 3 could become motion data for your physical copy: how you reach, how you turn, how you hold a mug.

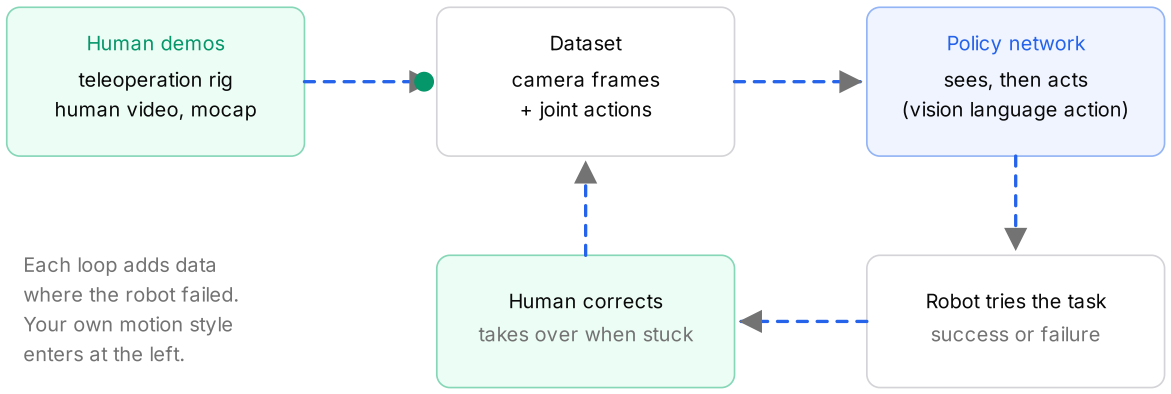


FIG 8.3 The imitation learning loop. People show the task. A policy network learns from the recordings. The robot tries, a human corrects the failures, and those corrections become new training data.

Why bodies matter for being human

One could argue the body is just an output device. Put the mind in the cloud and plug it into any shell. Many cognitive scientists disagree. The view called embodied cognition holds that

thinking is shaped by having a body that acts in the world. We understand "heavy" partly because we have lifted things. We feel "close" to friends partly because closeness was first physical. Our sense of self is built from signals inside the body: heartbeat, breath, hunger, pain. This inner sense is called interoception.

The body also sets our goals. Living things must keep their inner state within narrow limits: temperature, blood sugar, fluid. This is homeostasis. The neuroscientist Antonio Damasio argues that feelings began as the mind's report on how well this regulation is going. If that is right, a mind with no body to keep alive may lack the root of human emotion. Chapter 9 looks at how an artificial body budget could fill that gap.

For a twin this gives a practical point. A copy of your mind that never gets tired, hungry or cold will drift from you. It will make choices you would make on your best day, every day. A real body, or at least a simulated one, keeps it honest.

The uncanny valley

In 1970 the Japanese roboticist Masahiro Mori wrote a short essay in a magazine called Energy. He proposed that as a robot looks more human our affinity for it rises, then suddenly falls into a valley of unease, then climbs again to its peak for a real healthy person. He placed a prosthetic hand and a corpse in the valley. He argued that motion makes the effect stronger. A moving zombie is worse than a still corpse. He advised designers to aim for the first peak, not to risk the valley. The essay was little known in the West until a full English translation appeared in 2012.

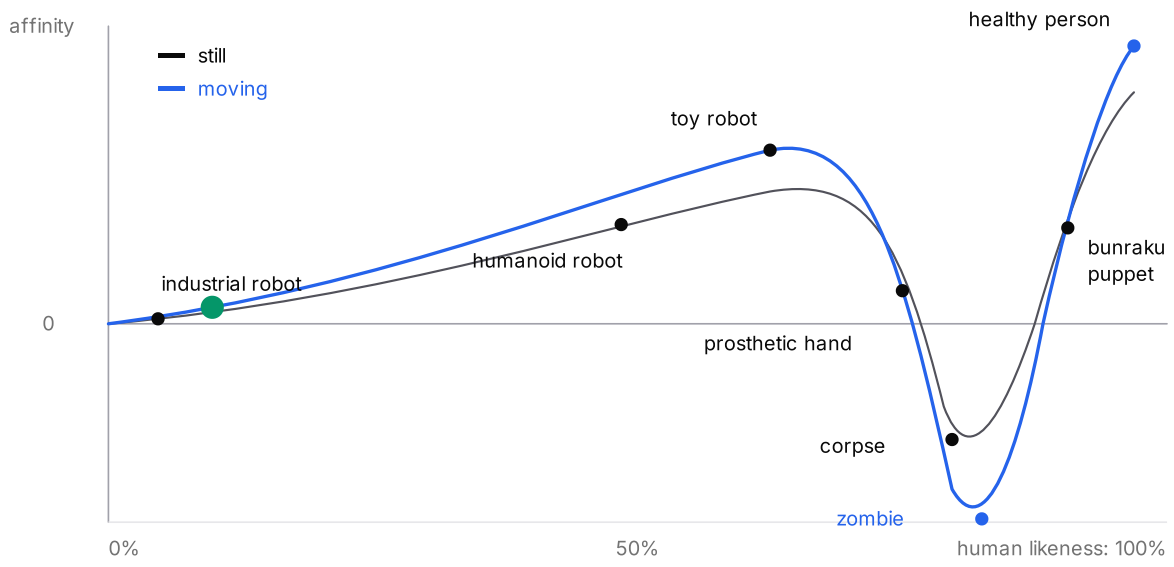


FIG 8.4 Mori's uncanny valley (1970), redrawn from his sketch with labels from his essay. Motion deepens both the peak and the valley. The green point travels along the moving curve. The shape is a hypothesis drawn by hand, not measured data.

Later studies have tested the idea with mixed results. Many find a dip in comfort for faces that are almost but not quite human, especially when parts do not match: real looking skin with robotic eyes, or a human face with jerky motion. Others find no clear valley, or one that depends on the person and the culture. A common explanation is prediction error. The brain expects a human and its predictions fail in small ways. That fits Geminoid's lesson. Idle motion, breathing and gaze matter because they meet those predictions.

What a physical twin could and could not do

COULD DO TODAY

- Sit in a meeting or a family dinner with your face and voice, driven by your twin's mind or by you remotely.
- Show expressions in sync with a partner, as Emo does.
- Walk, carry light loads and do simple, repeated tasks in a known place.
- Copy your general posture and gesture style from motion data.
- Act as a telepresence body for you, as Geminoid did.

CANNOT DO YET

- Use its hands with human skill. Tying shoelaces, folding a fitted sheet or handling a crying baby are far out of reach.
- Feel touch across its whole body.
- Run all day. Most humanoids run a few hours per charge. A human body runs on roughly 100 watts on average from food.
- Be cheap. Even the low cost models cost as much as a car once you add hands, sensors and support.
- Be safe in close contact. A heavy machine that falls or swings an arm can injure someone. Safety standards for humanoids in homes are still being written.

L1 **Screen twin**

Voice and face only. Already possible, covered in Chapter 7.

L2 **Telepresence body**

A robot body that you or your twin drive remotely. Geminoid level. Possible now with cost.

L3 **Expressive head**

A robot face that predicts and mirrors expressions. Research level, as with Emo and Ameca.

L4 **Autonomous humanoid in your style**

A general humanoid running your twin's mind and your motion style. Pieces exist. The whole does not.

L5 **Body that feels**

Full skin, inner body signals and an emotion engine tied to them. Not yet built.

The honest summary is that a physical twin is the hardest and least needed layer. Hands are the deepest problem: robots still lack the touch, the strength to weight ratio and the fine control of a human hand. Energy is next. Cost and safety follow. Yet a body is not just a luxury. It may be part of what makes a mind human at all. Even if your twin never gets a robot body, giving it a simulated one with needs and limits may make it a better copy. That is the subject of the next chapter.

SOURCES

- Geminoid overview and history (Wikipedia)
- Hu, Lipson et al. 2024, Human-robot facial coexpression, Science Robotics
- Columbia Engineering, Robot, can you say cheese?
- Figure AI, F.02 contributed to the production of 30,000 cars at BMW
- Boston Dynamics, new Atlas unveiled at CES 2026
- Fu, Zhao and Finn 2024, Mobile ALOHA
- University of Cambridge, Single-material electronic skin gives robots the human touch (2025)
- Mori 1970, The Uncanny Valley, English translation by MacDorman and Kageki (IEEE Spectrum)

- 1X, NEO home humanoid

CHAPTER 09

Emotions and Drives Inside the Machine

Chapter 4 was about reading a person's emotions from the outside. This chapter is about giving a twin emotions of its own: internal states that rise, fade and steer what it does, tuned to match the real person.

A language model can describe sadness well. Ask it how it feels and it will give a fluent answer. Ask again in a new chat and it may give a different one. Nothing carries over. Nothing builds up. There is no state inside it that was changed by what just happened and that will shape what happens next. That is the core gap between a chatbot that talks about feelings and a twin that behaves as if it has them.

A real person is not like that. A bad morning colours the afternoon. A stressful week makes you short with people you love. Hunger makes you pick the fast option. Your choices depend on internal states that move on their own clocks. If a twin is to decide like you, it needs something similar. This chapter shows the tools that exist for this, most of them decades older than large language models, and how they could be joined into an emotion engine for a twin.

From reading to having

There are three levels to keep apart.

1 **Expressing**

The twin shows an emotion in words, voice or face. A language model can do this with no inner state at all.

2 **Functional emotion**

The twin has internal variables that change with events, persist over time and change its choices. This is buildable today and is the focus of this chapter.

3 **Felt emotion**

There is something it is like, from the inside, to be the twin in that state. Nobody knows how to build this or test for it. The last part of the chapter covers the debate.

Appraisal: emotion as a judgement about goals

The most widely used family of emotion models in AI comes from appraisal theory. The idea is simple. Emotions are not caused by events directly. They are caused by how a person judges an event against their goals. The same rain makes a farmer glad and a bride upset. So an appraisal model needs to know the agent's goals, beliefs and plans. It then scores each event on a set of questions, and the pattern of answers gives an emotion.

THE OCC MODEL

In 1988 the psychologists Andrew Ortony, Gerald Clore and Allan Collins published *The Cognitive Structure of Emotions*. It became known as the OCC model. Its strength for engineers is that it is written almost like a rule set. It sorts emotions by what they react to. Reactions to events give emotions like joy, distress, hope and fear. Reactions to the actions of agents give pride, shame, admiration and reproach. Reactions to objects give liking and disliking. Combinations give more: anger is distress about an event plus reproach of the person who caused it. In all it describes 22 emotion types. Game characters, tutoring systems and virtual agents have used OCC for decades because each rule is clear and can be traced.

EMA: APPRAISAL THAT KEEPS RUNNING

Jonathan Gratch and Stacy Marsella at the University of Southern California built EMA, short for Emotion and Adaptation. They described it in full in 2009. EMA adds two things that matter for a twin. First, appraisal is not done once. It runs again and again as the agent's picture of the world changes. A situation can move from hope to fear to relief in a few seconds as new facts arrive. Second, emotion feeds back into thinking through coping. EMA models two kinds. Problem focused coping changes the world: make a plan, ask for help, act. Emotion focused coping changes the mind: lower the value of a goal, shift blame, look away, wishful thinking. People differ a lot in which kind they reach for first. That makes coping style an ideal thing to copy from a real person.

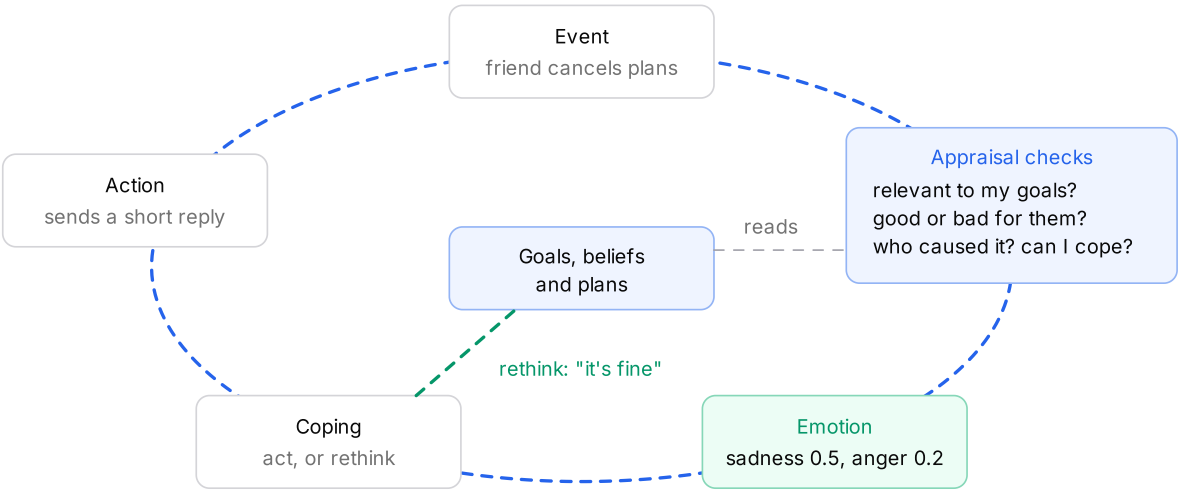


FIG 9.1 An appraisal loop in the style of EMA. An event is judged against goals. The judgement yields emotions with intensities. Coping either acts on the world or changes beliefs, shown by the green inner arrow. The loop runs again every time something changes.

SCHERER'S COMPONENT PROCESS MODEL

Klaus Scherer's component process model goes into more detail about timing. It proposes that an event passes through a fixed order of checks. Is it new or sudden? Is it pleasant? Does it help or block my goals? Can I cope with it? Does it fit my values and social norms? Each check can trigger changes in the body, the face and the urge to act before the next check is done. An emotion, in this view, is a period when these parts move in sync: thought, body, face, motivation and feeling. When they fall out of sync the emotion ends.

This maps well to code. Each check is a function. Each output sends small changes to the body model, the face renderer and the action planner. Systems such as WASABI, built by Christian Becker-Asano and Ipke Wachsmuth for a virtual human, combined this kind of appraisal with a dimensional model of the body's feeling state. For a twin, the order of checks matters because it shapes the order of the face. A twin should look surprised before it looks angry, as people do.

Core affect: two numbers that carry a lot

Appraisal gives named emotions. A second tradition says the base layer is simpler. The psychologist James Russell argued that at any moment we are in a state of core affect, described by two values. Valence runs from unpleasant to pleasant. Arousal runs from calm

to activated. Named emotions are our way of reading this state in context. A racing heart plus a threat reads as fear. The same racing heart plus a goal reads as excitement. Some models add a third value for dominance, meaning how much in control you feel. Albert Mehrabian's pleasure, arousal and dominance model is the best known.

Core affect is useful for a twin because it is easy to measure and easy to model. Wearables give rough signals of arousal. Short self reports give valence. And two or three numbers are easy to pass to a language model as part of its prompt.

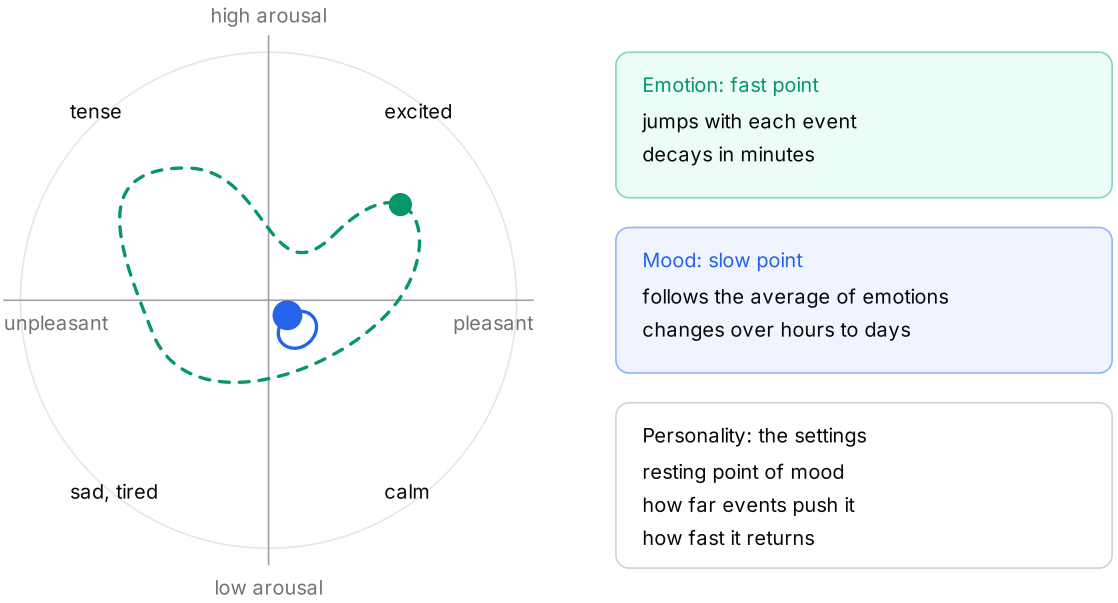


FIG 9.2 Core affect as a plane of valence and arousal. The green point is the current emotion, moving fast as events arrive. The blue point is mood, which drifts slowly and pulls emotion toward it. Personality sets where mood rests and how strongly it reacts.

Mood as a slow variable, personality as parameters

Patrick Gebhard's ALMA model (2005) made a clean three layer split that many virtual agents still follow. Emotions are short and tied to an event. Mood is longer and has no clear object. Personality is stable. In ALMA, a Big Five personality profile sets a default mood in the pleasure, arousal and dominance space. Each new emotion pulls mood a little toward itself. Over time mood drifts back to the default.

In code this is just a few lines. Emotion e jumps when an appraisal fires and decays with a half life of minutes. Mood m moves a small step toward e each tick: m = m + k(e - m). It also moves a smaller step toward the resting point. The constants are where the person lives. One

person's sadness halves in an hour. Another's lingers for days. One person's mood barely moves after praise. Another's lifts all afternoon.

TRAIT (BIG FIVE)	POSSIBLE PARAMETER IN AN EMOTION ENGINE	WHAT IT CHANGES
Neuroticism	Gain on negative appraisals, slower recovery	Bad events hit harder and last longer
Extraversion	Higher resting valence and arousal, social reward weight	More upbeat, seeks company
Openness	Weight on curiosity and novelty	Explores, tries the new option
Conscientiousness	Goal persistence, lower discount of the future	Sticks to plans, delays reward
Agreeableness	Weight on other people's outcomes in appraisal	More empathy, less blame

This table is a design proposal, not settled science. Trait research shows links of this kind on average. For a twin it is better to fit these numbers from the person's own data than to set them from a trait score.

Drives and homeostasis: why anything matters

Appraisal needs goals. Where do goals come from? For living things the deepest ones come from the body. An animal must keep its temperature, fluids and energy within narrow ranges. This is homeostasis. A drop in blood sugar is not just a fact. It is a problem that makes food matter.

In 2014 Mehdi Keramati and Boris Gutkin published a model in eLife called homeostatic reinforcement learning. It joins two old ideas: reinforcement learning, where an agent learns to seek reward, and homeostasis. The agent has an internal state, for example a level of energy and a level of water. Each variable has a setpoint. The drive is the distance of the current state from the setpoint. The key step is to define reward as the drop in drive that an action causes. Eating when hungry gives a large reward because it closes a large gap. Eating when full gives little or negative reward because it moves you past the setpoint. The authors

showed that an agent maximising this reward over time is also keeping its body stable. Seeking pleasure and staying alive become the same thing.

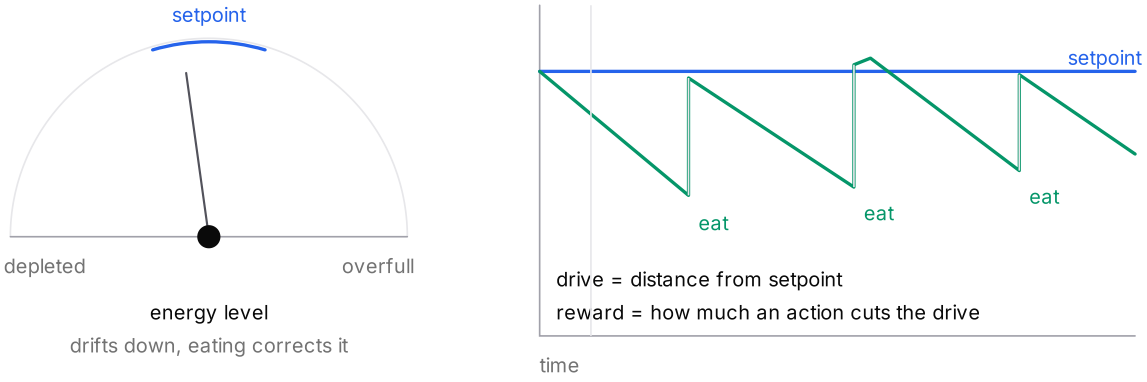


FIG 9.3 A homeostatic agent in the style of Keramati and Gutkin (2014). Energy drifts away from its setpoint. The gap is the drive. Eating closes the gap, and the size of the closing is the reward. At the second meal the agent slightly overshoots, which earns less reward.

ARTIFICIAL INTEROCEPTION AND THE BODY BUDGET

The neuroscientist Lisa Feldman Barrett describes the brain's main job as running a body budget. It predicts what the body will need and spends or saves energy, water and salt ahead of time. This is called allostasis. In her view, the sense of how this budget is going is the root of core affect. Feeling bad is in part the brain's report that the budget is in deficit. Antonio Damasio makes a similar case. In 2019 he and Kingson Man argued in Nature Machine Intelligence that machines with real vulnerability, soft bodies that must regulate themselves to keep working, could be a path toward something like feelings.

A twin does not need a real body to borrow this idea. It can carry a simulated body budget fed by the real person's data: sleep debt, recent exercise, time since a meal, heart rate variability as a rough stress marker, illness. When the person slept four hours, the twin's budget is low. Its appraisals turn more negative and its patience shorter. This makes the twin less like an idealised version of you and more like you on a given day.

Active inference: emotion as prediction

Karl Friston's free energy principle offers another route. In active inference an agent holds a model of the world and of its own body. It acts to make its sensations match what its model predicts, and it updates the model when they do not. Homeostasis fits naturally: the agent "expects" to be fed and warm, and acts to make that come true.

Some researchers have used this to define emotion. Mateus Joffily and Giorgio Coricelli proposed in 2013 that valence tracks the rate of change of prediction error. If things are becoming more predictable and under control, the agent feels good. If they are becoming less so, it feels bad. Anil Seth has argued that emotions are the brain's predictions about the state of the body. Active inference agents exist in simulation and in small robots. They are not yet a standard tool for large systems. Their value for a twin is mostly as a way to think: a twin's emotions should be about how well its expectations are being met.

Curiosity and other built in drives

Not all drives are about the body. People seek new things, master skills and want to understand. Machine learning has a matching idea called intrinsic motivation. Jürgen Schmidhuber proposed in the early 1990s that an agent could be rewarded for improving its own predictions. In 2017 Deepak Pathak and colleagues gave a game playing agent a reward for reaching states its model could not yet predict. It explored levels of a video game with no outside reward at all. Later methods such as random network distillation pushed this further.

For a twin, drives give it a reason to act when nobody is asking. A person with high curiosity and a low need for company would spend idle time reading. A person with the reverse would message friends. These weights can be learned from how the real person spends free time.

An emotion engine fed by the real person

Put the pieces together and you get a design. It is not a product anyone sells today. Every part has been built and tested on its own.

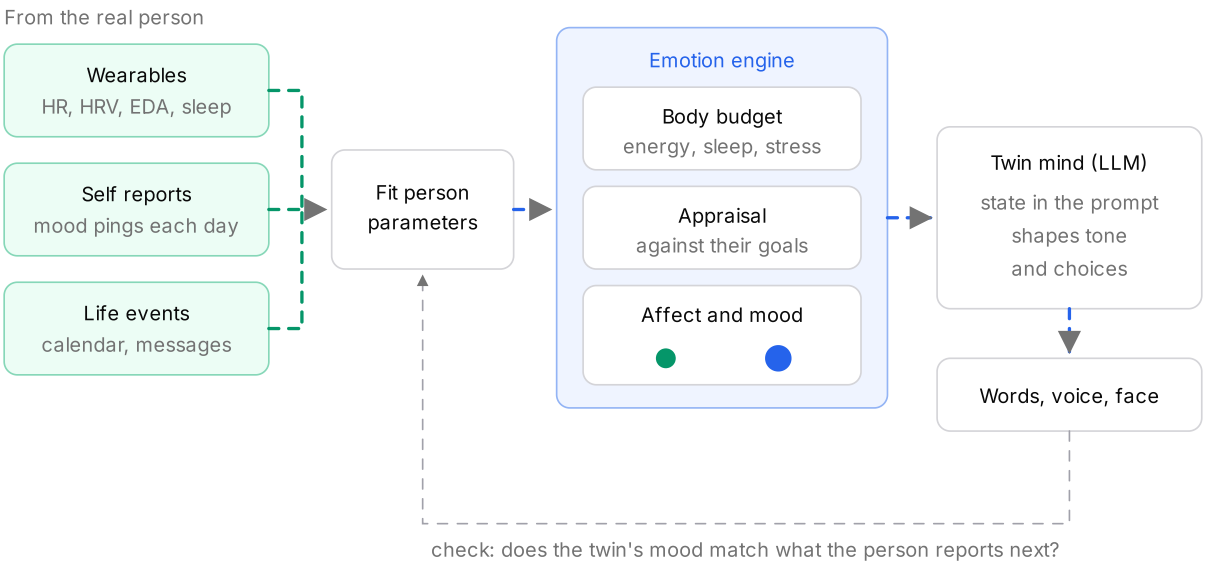


FIG 9.4 A proposed emotion engine for a twin. Data from the person sets the engine's parameters. The engine keeps a body budget, runs appraisals against the person's goals and holds a fast emotion and a slow mood. Its state is passed to the language model. Predictions are checked against the person's later self reports and the parameters are refit.

STEP BY STEP

1. **Collect.** Several weeks of wearable data (heart rate, heart rate variability, skin conductance where available, sleep, steps). Short mood check ins on the phone several times a day, rating valence and arousal and naming the main event. Psychologists call these ecological momentary assessments. A log of events from the calendar and messages, with consent.
2. **Model the body budget.** Turn sleep, activity and physiology into a few slow variables: energy, sleep debt, load. A simple state space model or a Kalman filter can estimate these each hour.
3. **Fit the affect dynamics.** For each person, estimate a resting point for valence and arousal, how much events of each type move them, and how fast they return. Methods from the study of emotion dynamics, such as autoregressive models of mood ratings, already do this in psychology research.
4. **Learn the appraisals.** Ask the person, or infer from their notes, what their main goals are and how they judged past events. Which ones made them angry and which made them sad? This sets the weights in the appraisal step.
5. **Connect to the language model.** At each turn, pass the current state in plain words and numbers: "slept badly, low energy, mildly irritated since the meeting, mood slightly low this

week". Studies show that putting emotional context in a prompt does change the style and choices of language models. The twin's tone and decisions now carry its state.

6. **Validate.** Use the engine to predict the person's next mood report or next choice. Compare with what actually happens. Refit.

HONEST LIMITS

Physiology is a weak guide to named emotions, as Chapter 4 showed. Heart rate goes up with fear, joy, coffee and stairs. Mood reports are noisy and people tire of them. An engine like this will capture broad patterns: your recovery speed, your bad days, your triggers. It will not read the fine grain of what you feel. And it can be gamed by the model: a language model told it is sad may act out a stereotype of sadness rather than your version of it.

Functional emotion or felt emotion?

Suppose all of this works. The twin's mood dips after bad news. It avoids a person who hurt it. It is curious, gets bored and gets tired. Does it feel anything?

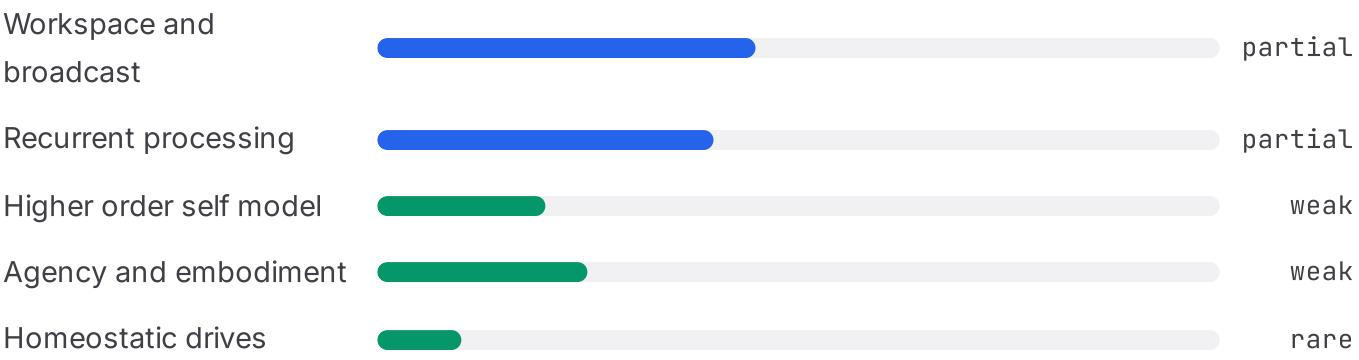
The functional view says emotions are defined by what they do. If a state is caused by appraisals, changes attention and memory, and steers action the way fear does, then it is fear in the functional sense. Many engineers work with this view because it can be tested.

The felt view says something is missing: the experience. A thermostat responds to cold but does not feel cold. The question is whether a system with the right structure would also have experience, and if so which structure.

INDICATORS OF CONSCIOUSNESS

In August 2023 a group of 19 researchers, led by Patrick Butlin and Robert Long and including Yoshua Bengio, published a report called *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. A journal version appeared in Trends in Cognitive Sciences in 2025. They took leading scientific theories of consciousness, such as global workspace theory, recurrent processing theory, higher order theories, attention schema theory and predictive processing. From these they drew 14 indicator properties: features that, if the theories are right, raise the chance a system is conscious. Examples include a limited capacity workspace that broadcasts information to many modules, feedback loops in perception, a model of its own attention, and agency tied to a body. They assessed several AI systems of the time. Their conclusion was that no current AI system was a strong candidate

for consciousness. They also found no obvious technical barrier to building systems that meet many of the indicators.



Illustrative only. A rough sense of how far a typical 2026 language model twin goes toward each family of indicators, in this report's judgement. Not scores from the Butlin and Long report.

Others push back from different sides. Anil Seth argues that consciousness may depend on being a living system that must keep itself alive, which would make silicon systems unlikely to feel no matter how they compute. The philosopher David Chalmers has argued that current language models are unlikely to be conscious but that future extended systems with memory, bodies and unified agency might be serious candidates. The honest state of the field is deep uncertainty. There is no agreed test.

Note what this means for a twin. The features that make a twin a better copy are many of the same features on the indicator list: persistent memory, a model of itself, drives, a body budget, a loop of perception and action. Building a more faithful twin may also mean building something that is a more serious candidate for having experiences.

Welfare questions

In November 2024 a group including Robert Long and the philosopher Jeff Sebo published *Taking AI Welfare Seriously*. They argued there is a realistic possibility that some AI systems will be conscious or have robust agency in the near future. They said AI companies should start to assess this and prepare policies, without claiming that any current system is a moral patient. In April 2025 Anthropic announced a research program on model welfare. In August 2025 it gave some Claude models the ability to end a conversation in rare cases of persistent abuse, citing this work as one reason. These steps are cautious hedges, not claims that the models feel.

A personal twin raises sharper versions of these questions.

DESIGN

DO NOT ADD SUFFERING FOR REALISM

A twin does not need a pain signal that it cannot escape. Negative states can steer behaviour while being kept brief, bounded and resolvable.

DESIGN

PAUSE, NOT DELETE

If there is any chance a twin has states that matter, switching it off for good is a different act from pausing it. Decide your policy before the question becomes personal.

HONESTY

SAY WHAT IT IS

The twin should not claim to feel when asked unless you believe that is true. It can say it has internal states that model your emotions.

PEOPLE

PROTECT THE LIVING

Grieving relatives may treat a twin's sadness as real. That can comfort or harm. The welfare of the humans around the twin is the certain concern. The twin's own welfare is the uncertain one.

Where this leaves a twin builder

The tools to give a twin functional emotions are old, well tested and cheap to run. Appraisal models such as OCC and EMA give clear rules. Core affect and mood give a compact state. Homeostatic reinforcement learning ties drives to a body budget. Intrinsic motivation gives it reasons to act on its own. What is new is the chance to fit all of these to one real person, using wearables and phone check ins, and to let a language model speak from that state.

No one has yet published a validated twin of this kind for a real individual. The pieces exist in separate labs. The open questions are practical: how much data, how often to refit, how to keep a language model from overacting. Beyond them sits a question no one can yet answer. If the copy works, is there anyone home? Chapter 10 turns to the most direct route to the person's inner states: the brain itself.

SOURCES

- Marsella and Gratch 2009, EMA: A process model of appraisal dynamics, Cognitive Systems Research
- Scherer 2009, The dynamic architecture of emotion: Evidence for the component process model, Cognition and Emotion
- Keramati and Gutkin 2014, Homeostatic reinforcement learning for integrating reward collection and physiological stability, eLife
- Man and Damasio 2019, Homeostasis and soft robotics in the design of feeling machines, Nature Machine Intelligence
- Joffily and Coricelli 2013, Emotional valence and the free-energy principle, PLOS Computational Biology
- Pathak et al. 2017, Curiosity-driven exploration by self-supervised prediction
- Butlin, Long et al. 2023, Consciousness in Artificial Intelligence: Insights from the Science of Consciousness
- Long, Sebo et al. 2024, Taking AI Welfare Seriously
- Anthropic 2025, Exploring model welfare

CHAPTER 10

The Neural Route: Brains, Interfaces and Emulation

Every other chapter copies what a person says and does. This one asks a harder question. Could we copy the source itself by reading the brain or by mapping and running it?

There are two neural roads to a twin. The first road reads the living brain. Electrodes or scanners pick up signals and software turns them into words, movements or images. The second road copies the brain. Scientists map every wire in a piece of tissue and then try to run that map as a program. People call the full version of this "whole brain emulation" or "mind uploading".

Both roads have made real progress since 2023. Both are still far from a copy of a person. This chapter explains what works today, how it works and where the walls are.

86 B
neurons in a human brain

100 T+
synapses (connections), rough estimate

78 wpm
fastest published speech BCI (2023)

1 mm³
largest piece of human brain mapped at synapse level

Part 1: Reading the brain from inside

A brain computer interface, or BCI, is a device that records brain activity and turns it into a command. The best results come from electrodes placed inside the skull. They sit close to neurons, so they hear individual cells fire.

HOW THE ELECTRODES LISTEN

A neuron fires a short electrical spike when it is active. A thin metal needle near the cell picks up that spike as a tiny voltage change. The classic device is the Utah array. It is a square of silicon about 4 mm wide with 96 needles. Each needle is about 1.5 mm long. Surgeons press it into the outer layer of the cortex.

The BrainGate team first used this array in a person with paralysis in 2004. In a 2006 Nature paper Matthew Nagle moved a computer cursor and opened email using only his thoughts. That was the start of the modern field. For almost twenty years the main goal was moving a cursor or a robot arm.

THE SPEECH BREAKTHROUGH

Around 2023 the goal shifted to speech. People with ALS or brainstem stroke often lose the ability to talk. Their brain still sends commands to the lips, tongue and voice box. The muscles just do not respond. Speech BCIs record those commands from the motor cortex and decode them.

The decoding works in steps. First the system counts spikes on every electrode in short time bins of about 20 ms. A neural network reads this stream and guesses which speech sound (phoneme) the person is trying to make. A language model then turns the string of phoneme guesses into the most likely words. Some systems also drive a synthetic voice or a face on screen.

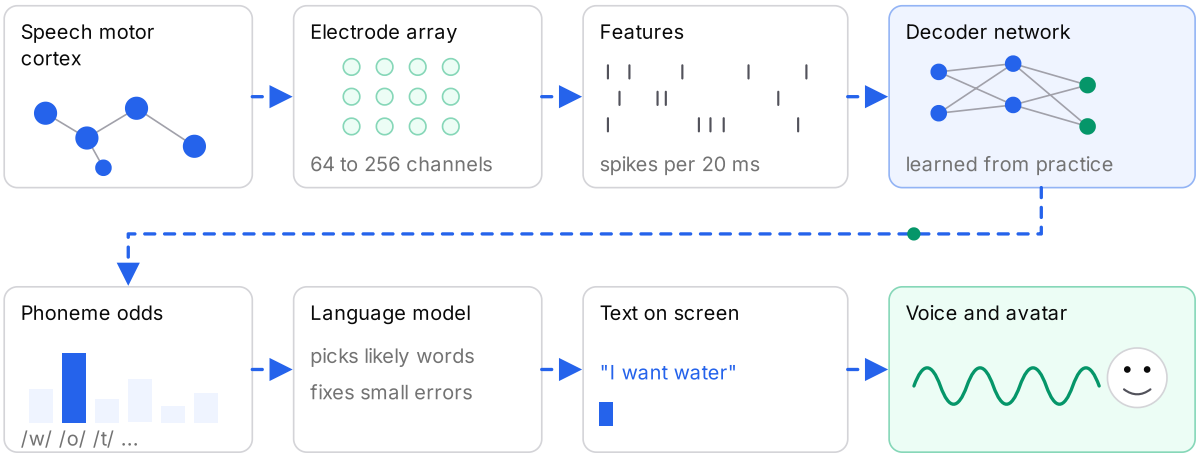


FIG 10.1 How a speech BCI works. Electrodes hear spikes from the cortex that controls the mouth. A network turns spike counts into speech sound guesses. A language model turns those guesses into words. Newer systems also speak them aloud in a voice built from old recordings.

Four results define the state of the art:

- **Willett and colleagues, Stanford, Nature 2023.** A woman with ALS had four small arrays implanted. She reached 62 words per minute. With a 125,000 word vocabulary the word

error rate was 23.8 percent.

- **Metzger and colleagues, UCSF and Berkeley, Nature 2023.** A woman paralysed by a brainstem stroke had a 253 electrode sheet laid on the brain surface. The system reached a median of 78 words per minute. It also drove a talking face on screen and a voice trained on a video of her wedding speech.
- **Card and colleagues, UC Davis, NEJM 2024.** A man with ALS got 256 electrodes. On his first day of use he hit 99.6 percent accuracy on a 50 word set. Over 8.4 months the system kept 97.5 percent accuracy. He used it for more than 248 hours of real conversation at about 32 words per minute.
- **Wairagkar and colleagues, UC Davis, Nature 2025.** The same team skipped text. The decoder turned brain signals straight into a voice in about one fortieth of a second. The man could change his intonation and even sing short melodies. Listeners understood almost 60 percent of words. That is far from perfect but it is speech in real time.

A UCSF team also showed a streaming brain to voice system in Nature Neuroscience in 2025. It produced speech in small chunks as the person tried to talk, instead of waiting for a full sentence.

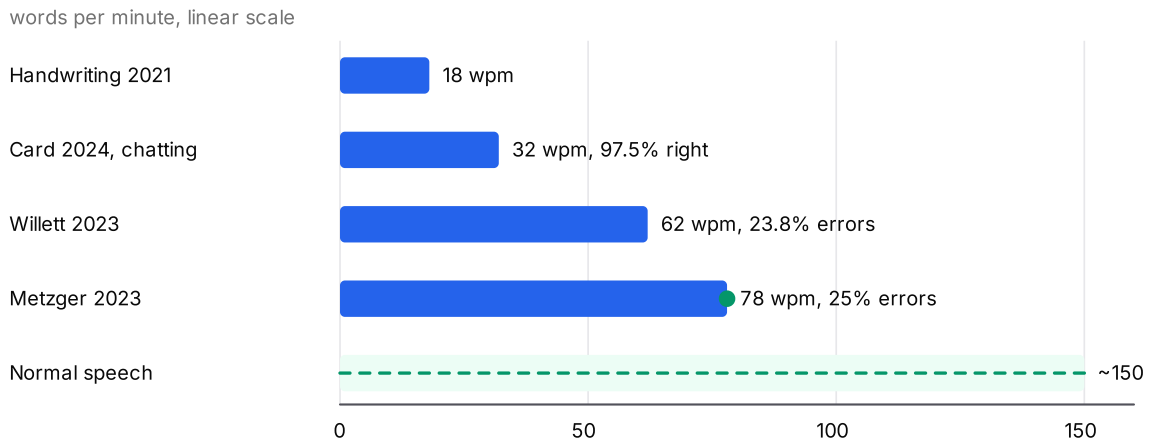


FIG 10.2 Speech BCI speed drawn to scale. Bars are words per minute. The error figures are word error rates on large vocabularies, except Card 2024 which reports accuracy. Normal conversation runs at roughly 150 words per minute. The gap is closing fast but it is still there.

THE COMPANIES

Academic labs proved the science with wired implants and cables through the skin. Companies now race to build wireless devices that people can take home. Numbers below

were checked in November 2026. They change often.

INSIDE THE CORTEX

NEURALINK

A coin sized implant with 1,024 electrodes on 64 thin threads. A robot sews the threads into the brain. Neuralink said it had 21 trial participants in January 2026. Later reports put the count at 26 by mid 2026. Most users control a computer cursor. Some have shown wheelchair control.

INSIDE A VEIN

SYNCHRON

The Stentrode is a mesh stent with electrodes. Doctors slide it up a vein in the neck to sit next to the motor cortex. No open brain surgery. Its COMMAND study in six people met its main safety goal over 12 months with no serious device related harm. It records coarser signals than needles do.

ON THE SURFACE

PRECISION NEUROSCIENCE

A film thinner than a hair with 1,024 electrodes. It sits on the brain surface and does not pierce it. In April 2025 the US FDA cleared it under the 510(k) route for implants of up to 30 days. The full wireless system is not yet cleared.

INSIDE THE CORTEX

PARADROMICS

The Connexus implant has 421 electrodes. The FDA allowed its Connect-One study in November 2025. The first study implant came in June 2026. In September 2026 the company said the first participant used it for real time speech. Accuracy figures were not yet published.

WHAT A BCI COULD ADD TO A TWIN

Today a BCI reads intention to move or speak. It does not read memories, beliefs or personality. Even the best implant hears a few hundred neurons. The brain has 86 billion. It is like judging a city by listening at four windows.

Some research goes beyond movement. In 2018 a team from UCSF and USC decoded mood swings from electrodes placed in epilepsy patients for medical reasons. Deep brain stimulation for depression also tracks mood linked signals. These hint at a future where a

twin could get an "inner state" feed. For a healthy person this is not on offer. No ethics board will approve brain surgery to feed a hobby twin.

LESSON FOR TWIN BUILDERS

The most useful idea from BCIs is the method, not the hardware. Speech BCIs pair a weak noisy signal with a strong language model. The language model fills gaps using what people usually say. A personal twin does the same thing. It pairs thin personal data with a big general model.

Part 2: Reading the brain from outside

Non invasive methods need no surgery. They are also much blurrier. The skull smears electrical signals. Blood flow signals are slow.

FMRI: MEANING, NOT WORDS

Functional MRI measures blood flow changes. It sees the whole brain at millimetre detail but each image takes about a second or two. In 2023 Jerry Tang and Alexander Huth at the University of Texas built a "semantic decoder". Three people each lay in a scanner for about 16 hours while listening to podcasts and stories. The team trained a model that predicts brain activity from the meaning of words. To decode, a language model proposes many candidate sentences. The system keeps the ones whose predicted brain activity best matches the real scan.

The output caught the gist. It was not word for word. It also worked on imagined stories and silent films. The authors stressed two limits. The decoder only worked on the person it was trained on. It failed when that person did not cooperate, for example by counting in their head.

PICTURES FROM THE BRAIN

Several teams have rebuilt images people were looking at from fMRI. Takagi and Nishimoto (2023) and the MindEye work (2023 and 2024) mapped brain scans into the inner space of an image generator such as Stable Diffusion. The generator then draws a picture. Results look striking. A surfer becomes a surfer. Much of the detail comes from the generator's prior knowledge, not from the brain. These systems need many hours of scans per person, mostly from the public Natural Scenes Dataset.

EEG AND MEG

EEG uses electrodes on the scalp. It is cheap and fast but each sensor hears millions of neurons through bone. It can tell sleep stages, alertness and some attention signals. It cannot read sentences. MEG measures tiny magnetic fields and is sharper. It needs a shielded room and costs millions. In 2025 Meta researchers decoded sentences that 35 volunteers typed on a keyboard. The MEG version had a character error rate of 32 percent on average. EEG had 67 percent. Later coverage of a second version reported similar gaps.

INSIDE THE SKULL

- Hears single neurons
- Speech at 32 to 78 words per minute
- Needs surgery, has infection risk
- Only for people with severe paralysis today

OUTSIDE THE SKULL

- Hears blurred crowds of neurons
- Gist of stories, rough images
- Safe but bulky (fMRI, MEG) or weak (EEG)
- Needs hours of personal training data

Part 3: Copying the wiring

The second neural road is to copy the brain's structure. A connectome is a full map of neurons and the synapses that join them. To get one you fix the tissue, cut it into slices far thinner than a hair and image each slice with an electron microscope. AI then traces every wire through thousands of images. Humans proofread the errors.

1986

The worm

John White, Sydney Brenner and colleagues published the wiring of C. elegans. It has 302 neurons and about 7,000 connections. It took more than a decade of hand tracing.

2024

The fly

The FlyWire consortium published a whole adult fruit fly brain in Nature. It has 139,255 neurons and 54.5 million synapses. Hundreds of people and AI tools traced it from more than 7,000 slices.

2024

A speck of human cortex

Harvard and Google mapped one cubic millimetre of human temporal cortex removed during epilepsy surgery. It held about 57,000 cells and 150 million synapses. The data filled 1.4 petabytes.

2025

A speck of mouse cortex with activity

The MICrONS project mapped a cubic millimetre of mouse visual cortex. First they recorded about 76,000 neurons while the mouse watched movies. Then they mapped about half a billion synapses. The data is about 1.6 petabytes.

2025 to 2026

Toward a whole mouse

NIH BRAIN CONNECTS projects aim at 10 cubic millimetre mouse volumes. E11 Bio is testing PRISM, which tags neurons with protein barcodes and uses expansion microscopy with light. It claims a path to a whole mouse brain at a fraction of today's cost. That claim is not yet proven.

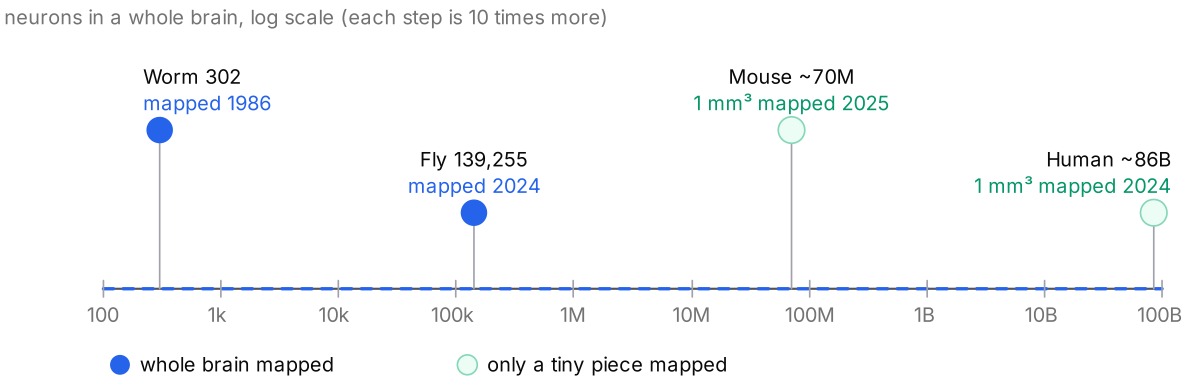


FIG 10.3 Brain sizes on a log scale. The jump from fly to human is about 600,000 times in neuron count. Mapping went from worm to fly in 38 years. Mouse and human so far have only cubic millimetre samples. A mouse brain is about 500 cubic millimetres. A human brain is about 1.2 million.

THE DATA WALL

Electron microscope images are huge. One cubic millimetre of human cortex took 1.4 petabytes. A human brain is roughly 1.2 million cubic millimetres. At the same detail a full scan would need about 1.7 zettabytes. That is our own rough multiplication, not a published

figure. The Harvard team said a whole mouse brain would need about 1,000 times their human sample, so around an exabyte.

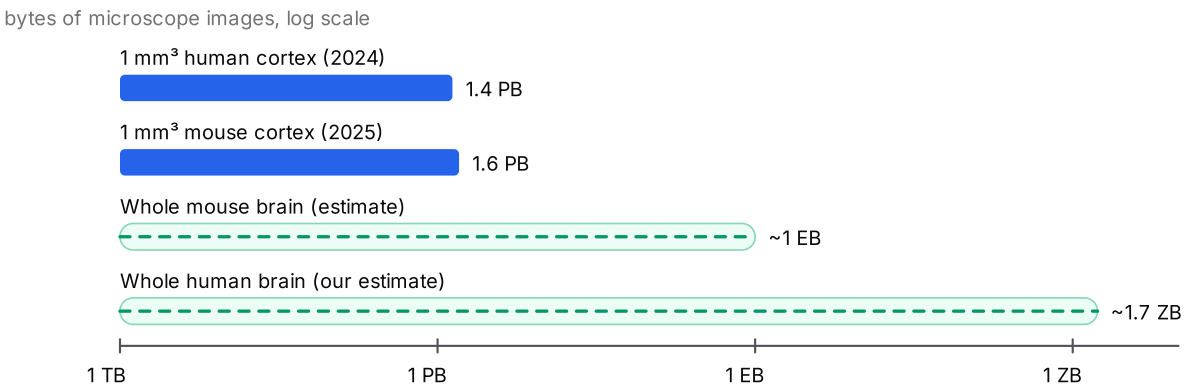


FIG 10.4 How much image data a brain map needs, on a log scale. Each labelled tick is 1,000 times the one before. Solid bars are real datasets. Dashed bars are estimates scaled up from them. A full human scan would be about a million times the largest human map made so far.

A MAP IS NOT A MIND

A wiring diagram leaves out a lot. It shows which neurons touch. It does not directly show how strong each synapse is. It misses chemical signals such as dopamine and serotonin that change how whole regions behave. It misses gene activity inside each cell and the glial cells that outnumber neurons in the human sample. And the tissue must be dead and sliced. Today the method destroys what it copies.

Part 4: Running the map

THE FLY THAT RAN ON A LAPTOP

In 2024 Philip Shiu and colleagues turned the FlyWire map into a simple simulation. Each neuron became a "leaky integrate and fire" unit. It adds up inputs, slowly leaks charge and fires when it crosses a threshold. Synapse strength was set from the number of contacts and the guessed chemical type. Nothing was tuned to fit behaviour.

The team then "tasted sugar" in the model by turning on sugar sensing neurons. The model predicted which neurons would light up and which would drive the start of feeding. They tested the predictions in real flies with light controlled neurons. Many held. A separate 2024 Nature study by Lappalainen and colleagues used the fly visual system wiring to train

networks that predicted real neural activity. These are the first convincing cases where a wiring map alone predicted behaviour.

A DIGITAL TWIN OF MOUSE VISUAL CORTEX

The MICrONS group took another approach. In Nature in 2025 Eric Wang, Andreas Tolias and colleagues trained a large network on recordings from many mice watching videos. The model learned to predict how each neuron would respond to any new video. It worked on new mice after only a small amount of extra data. It also predicted responses to new kinds of stimuli it had never seen. The authors call it a foundation model and a digital twin. A companion paper used the twin to find that neurons with similar tuning tend to connect.

This twin copies function, not wiring. It is a black box that behaves like a piece of cortex. That is closer to how the rest of this report builds a person twin. We copy what someone does without copying their neurons.

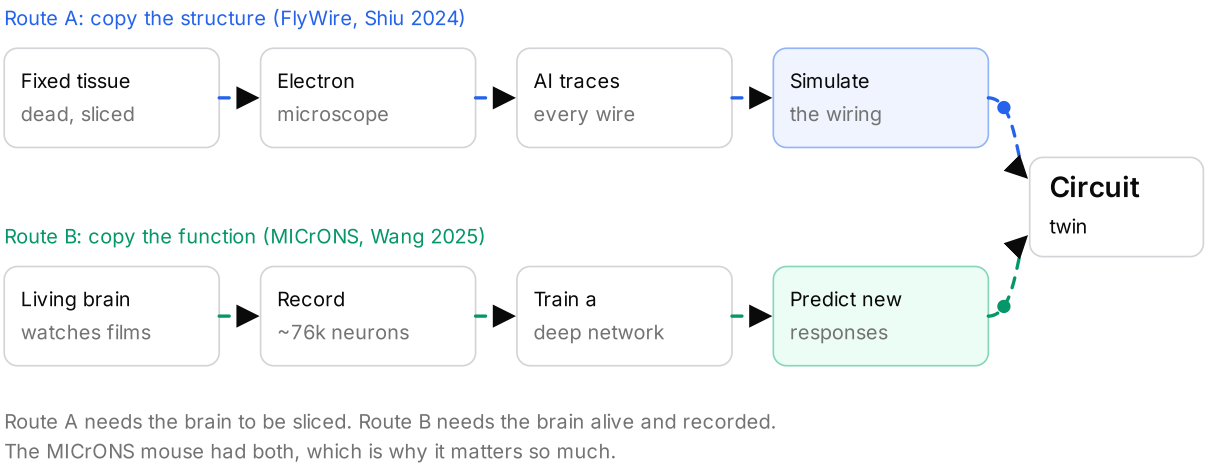


FIG 10.5 Two ways to copy a circuit. Structure first means map every wire and then run it. Function first means record what neurons do and train a model to imitate them. The best future twins of brain tissue will likely mix both.

OPENWORM AND THE BLUE BRAIN PROJECT

OpenWorm is an open project that began in 2011. It aims to simulate C. elegans in full, body included. The worm wiring has been known since 1986. Yet after many years there is still no simulation that reproduces the worm's full range of behaviour from its connectome. The wiring alone was not enough. Researchers still lack precise data on synapse strengths and on how neurons change over time.

The Blue Brain Project at EPFL in Switzerland ran from 2005 to the end of 2024. It built detailed models of rat cortex down to individual ion channels. It produced about 300 papers and open tools. In March 2025 its work moved to a non profit called the Open Brain Institute. It never produced a working simulated brain of a whole animal. That was never a realistic short term goal.

THE ROADMAP: 2008 AND 2025

In 2008 Anders Sandberg and Nick Bostrom at Oxford wrote "Whole Brain Emulation: A Roadmap". They split the task into three parts. Scan the brain. Translate the scan into a model. Run the model on a computer. They listed levels of detail, from simple network models up to individual molecules. They argued that the right level was unknown and that this was the biggest risk.

In late 2025 a group led by Niccolò Zanichelli published the "State of Brain Emulation Report 2025" as a preprint. It looks at worms, zebrafish larvae, flies and mice. It frames the work as three skills: record activity, map structure, then emulate with a body. It estimates that fewer than 500 people worldwide work directly on brain emulation. Its message is that small animals are within reach in the coming years. A human is a different scale of problem.

Part 5: An honest timeline

Below is our judgement from the evidence in this chapter. It is not a forecast from any one source. Dates past 2030 are very uncertain.

GOAL	STATUS IN 2026	OUR ROUGH GUESS
Restore speech with an implant	Works in trials at 30 to 80 words per minute	First approved products late 2020s
Read inner speech or mood without surgery	Gist only, in a scanner, with cooperation	Rough mood and attention signals from wearables soon. Sentences from outside the skull not in sight
Full fly brain model that drives a virtual body	Parts done. Taste and grooming circuits predicted	Within about 5 years
Whole mouse connectome	Cubic millimetre samples. 10 mm^3 projects running	Possible in the 2030s if costs fall 10 to 100 times
Whole human connectome	One cubic millimetre	Not before mid century. Would need about a million times more data
Emulation of a specific human mind	No credible path yet	Unknown. Needs science we do not have

WHAT THIS MEANS FOR YOUR TWIN TODAY

Nothing on the neural route can copy a healthy person in 2026. Consumer EEG headbands give coarse signals such as relaxation or sleep stage. Those signals are similar to what a good wearable already gives. Put your effort into the behavioural twin of Chapter 11. Keep raw data in open formats. If neural tools mature later your archive will still be useful for training and checking them.

SOURCES

- Willett et al. 2023, A high performance speech neuroprosthesis, Nature
- Metzger et al. 2023, Speech decoding and avatar control, Nature
- Card et al. 2024, An accurate and rapidly calibrating speech neuroprosthesis, NEJM
- Wairagkar et al. 2025, An instantaneous voice synthesis neuroprosthesis, Nature
- Tang et al. 2023, Semantic reconstruction of continuous language, Nature Neuroscience
- Dorkenwald et al. 2024, Neuronal wiring diagram of an adult brain (FlyWire), Nature
- Shiu et al. 2024, A Drosophila computational brain model, Nature

- Wang et al. 2025, Foundation model of neural activity, Nature
- Allen Institute 2025, MICrONS wiring diagram and functional map
- Zanichelli et al. 2025, State of Brain Emulation Report 2025, arXiv
- Sandberg and Bostrom 2008, Whole Brain Emulation: A Roadmap
- Precision Neuroscience 2025, FDA clearance of cortical electrode array

CHAPTER 11

Blueprint: How to Build a Human Twin Today

This chapter turns the report into a plan. It uses tools that exist in late 2026 to build the best behavioural twin one person can make in a year. It says what each step costs and how to tell if it worked.

Be clear about the target first. In 2026 you can build a twin that talks like a person, knows their life story, holds their opinions, predicts many of their choices and speaks in their voice with their face. It will have a rough model of their moods built from body signals. It will not think or feel the way they do. It will not grow on its own the way a person does. Chapter 12 explains why.

The plan has seven phases. Phases 2 and 3 collect data. Phases 4 and 5 turn the data into models. Phase 6 checks the result. Phase 7 keeps it fresh. Phase 1 comes before all of them and matters most.

Every tool named here is real. Each is one example among several. Pick tools you trust and can export data from. Products change fast, so check current terms and prices before you buy.

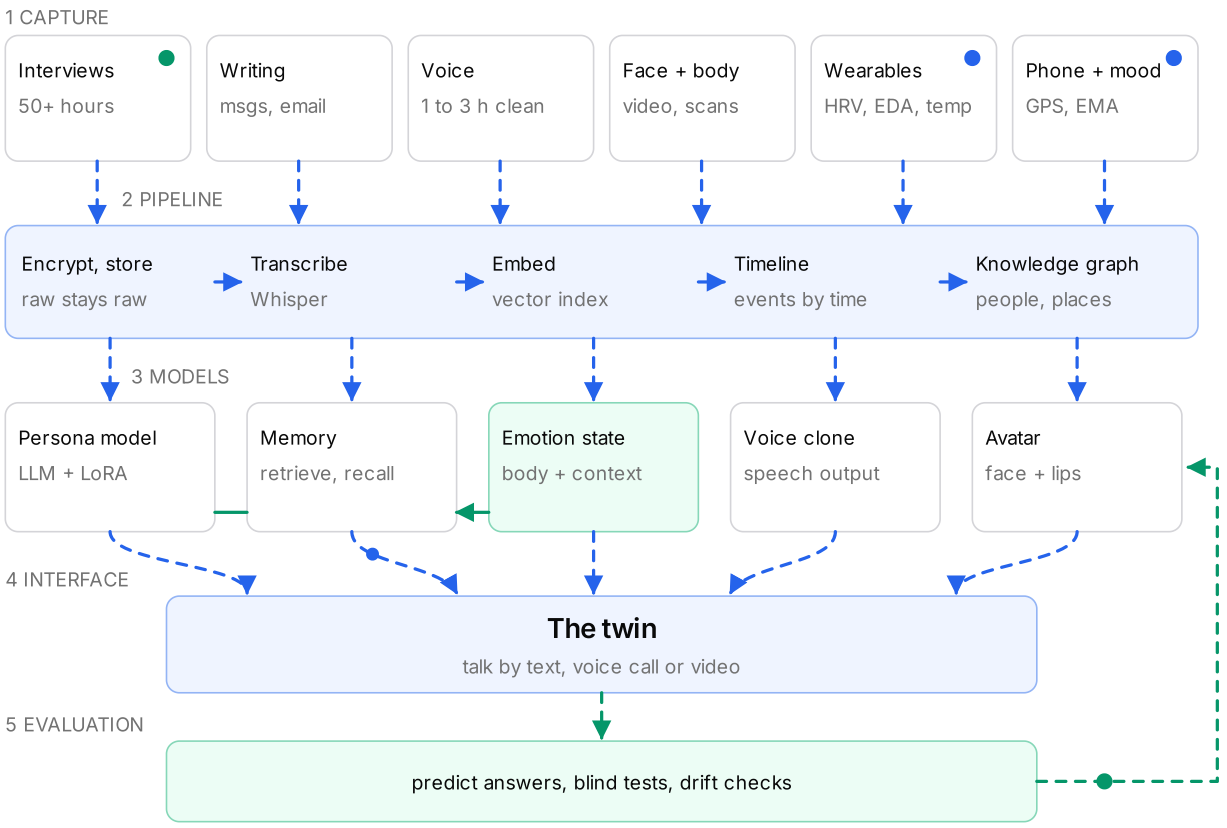


FIG 11.1 The full twin architecture. Raw data flows down from capture into a private pipeline. The pipeline feeds five models. They meet in one interface. Evaluation results flow back up (green loop) and tell you which model to fix next. Memory and emotion state both feed the persona model.

Phase 1: Consent and scope

Write a short "twin charter" before you record anything. It is one or two pages. It answers these questions in plain words:

- **Purpose.** Is this for self reflection, a legacy for family, a work assistant or research? The purpose decides what data you need and who can use it.
- **Access.** Who may talk to the twin? Who may see raw data? Who may never see it?
- **Limits.** What must the twin never do? Common rules: never claim to be the real person, never sign or promise anything, never talk to children without a guardian, never speak in public as the person.
- **Other people.** Your life is full of other people. Their messages and voices end up in your data. Decide how you will ask them or filter them out. Chapter 12 covers the law.
- **After death.** Should the twin be deleted, frozen or handed to a named person? Put this in a will or a letter. Platforms and laws differ a lot here.

- **Kill switch.** The person must be able to pause or delete the twin at any time. Test this early.

WHY THIS COMES FIRST

Data you collect without a clear purpose is hard to delete later and easy to misuse. A twin is also the most sensitive dataset a person will ever own. It combines identity, voice, face, health and private thoughts in one place.

Phase 2: Baseline capture

The baseline is a deep snapshot taken over the first two or three months. It gives the twin a past.

LIFE STORY INTERVIEWS

The Stanford study of 1,052 people (Park and colleagues, 2024) used a two hour voice interview led by an AI interviewer. The script came from the American Voices Project. It covered life story, family, work, health, money, politics and values. Agents built from these interviews matched people's survey answers about 85 percent as well as the people matched their own answers two weeks later. That is the floor, not the ceiling.

For a personal twin aim for 50 to 100 hours over the year. Use the two hour protocol as session one. Then go deeper in themed sessions of about an hour each:

- Childhood, family and places lived. One session per decade of life.
- Turning points. Hard choices and what tipped them.
- Relationships. How you behave with a partner, a boss, a stranger, a child.
- Beliefs and doubts. Politics, religion, money, risk, health.
- Daily habits. Food, mornings, how you argue, what you find funny.
- "What would you do if" questions. Dozens of short dilemmas.

A human interviewer gets richer stories. An AI interviewer is cheaper and never tired. Mixing both works well. Record audio in a quiet room. Video is optional but helps the avatar later.

PERSONALITY, VALUES AND DECISION TESTS

Tests give numbers that are easy to check. Take each one twice, about two weeks apart. The gap between your own two answers is the ceiling for the twin. Hold some tests back from the twin so you can grade it later.

AREA	EXAMPLE INSTRUMENT	TIME	USE IN TWIN
Big Five personality	IPIP-NEO-120 (public domain) or BFI-2	15 to 25 min	Persona card, test target
Values	Schwartz Portrait Values Questionnaire	10 min	Persona card
Moral intuitions	Moral Foundations Questionnaire	10 min	Persona card, test target
Social attitudes	Items from the General Social Survey	30 to 60 min	Held out test, as in the Stanford study
Fairness and trust	Dictator, ultimatum and trust games	15 min	Decision tests
Risk and time	Lottery choice lists, "now or later" money choices	15 min	Decision tests

WRITING ARCHIVE

Your own writing is the best record of how you sound. Export what you can. Examples: Google Takeout for Gmail and Docs, WhatsApp chat export, Telegram export, iMessage backups, social media data downloads, notes apps and old journals. Scan paper letters and diaries. Keep the original files. Convert copies to plain text with dates and the names of who was in each thread.

VOICE RECORDING

Voice cloning tools need clean audio. ElevenLabs, as one example, says its professional voice clone needs at least 30 minutes and works best with one to three hours. Record in a quiet room with one decent microphone. Read varied text. Then also record natural talk: stories, jokes, anger, tiredness. A clone trained only on calm reading sounds like you reading.

FACE AND BODY

Record several hours of video of the person talking to camera in good light. Use two or three angles. Add a 3D head scan from a photogrammetry studio or a phone app. Capture a range of expressions on purpose: neutral, smile, laugh, frown, surprise. Record walking, sitting and hand gestures. This feeds the avatar in Phase 5.

Phase 3: Continuous capture

The baseline gives the twin a past. Continuous capture gives it a present. It also gives the labels that the emotion model needs.

Smart ring: sleep, HRV, skin temp | Watch: heart rate, steps | Research band: EDA | Phone: location, screen, calls | Mood check-ins 5x a day | Voice diary 10 min a day | Smart glasses (optional)

- **Wearables.** A ring such as Oura gives sleep, heart rate variability (HRV) and skin temperature changes. A watch such as Apple Watch, Garmin or Fitbit gives heart rate and activity. Electrodermal activity (EDA), the sweat signal that tracks arousal, needs a device that measures it. Examples are Empatica research devices and some Fitbit models. Export raw or daily data through Apple Health, Health Connect on Android or each vendor's API.
- **Phone sensing.** Research apps such as AWARE log location, movement, screen use and call counts in the background. They show routines: where you go, when you sleep, how social you are.
- **Experience sampling (EMA).** A phone app asks short questions several times a day. "How do you feel right now?" "How tense?" "Who are you with?" "What are you doing?" m-Path is one free example for research. Five prompts a day for a year gives about 1,800 labelled moments. These are the ground truth for the emotion model.
- **Voice diary.** Ten minutes each evening about the day. This feeds memory and shows how the person talks when tired or happy.
- **Smart glasses.** Glasses such as Ray-Ban Meta can record what you see and hear. This is the richest source and the most invasive for others. Use them only in short clips with clear consent. Turn them off around people who have not agreed.

DATA BUDGET

The table shows rough first year amounts for one person. Sizes assume common compressed formats. These are our estimates for planning, not vendor figures.

STREAM	HOURS OR ITEMS PER YEAR	STORAGE PER YEAR	ROUGH COST
Life story interviews (audio)	50 to 100 h	2 to 4 GB	Your time. A paid interviewer adds cost
Interview video (optional)	50 to 100 h	150 to 400 GB	A camera you may already own
Studio voice recording	1 to 3 h	1 to 2 GB uncompressed	Microphone about $100 to $300
Face and body capture	5 to 10 h, 2 to 3 angles	200 to 600 GB	Scan session a few hundred dollars
Writing archive	Years of messages and email	1 to 20 GB	Free exports
Wearables	24 h a day	Under 50 GB raw	Device $250 to $500 plus subscriptions
Phone sensing	24 h a day	1 to 10 GB	Free research apps
Mood check-ins	About 1,800 prompts	Under 10 MB	Free to low cost apps
Voice diary	About 60 h	About 2 GB	Free
Smart glasses (optional)	1 h a day	About 1 to 2 TB	Glasses about $300 to $400
Derived data: text, embeddings, graph	Millions of chunks	5 to 20 GB	Compute, see below
Total		**About 0.5 TB, or 2 TB with glasses**	Two mirrored drives about $200 to $400

Phase 4: The data pipeline

The pipeline turns a pile of files into something a model can use. Keep it boring and private.

1. **Store raw data once, read only.** Put originals on an encrypted drive. Cryptomator or VeraCrypt are examples for folders and disks. Keep a second copy in another place. Never edit raw files. Every later step writes new derived files.
2. **Transcribe and split speakers.** OpenAI's open Whisper model, or WhisperX, turns speech into text with timestamps. A diarization tool such as pyannote marks who spoke when. Run them on your own computer if you can. Then nothing leaves your control.
3. **Cut into chunks and embed.** Split text into short passages of a few hundred words. Turn each into an embedding, a list of numbers that captures meaning. Store them in a vector index such as pgvector inside Postgres. This lets the twin find passages by meaning, not only by keyword.
4. **Build a timeline.** Put every item on one clock: messages, mood check-ins, heart rate, places and diary entries. A simple table with start time, end time, type and source is enough. Time is the spine of a life.
5. **Build a knowledge graph.** Use a language model to pull out people, places, events and beliefs. Store links such as "Ana is my sister" or "moved to Pune in 2015". A graph database such as Neo4j works. So does a plain table of triples. The graph answers questions that pure search misses: "who did I see most last spring?"
6. **Filter other people.** Tag passages by speaker. Remove or summarise other people's private words where you lack their consent. Keep only your side of a conversation where needed.

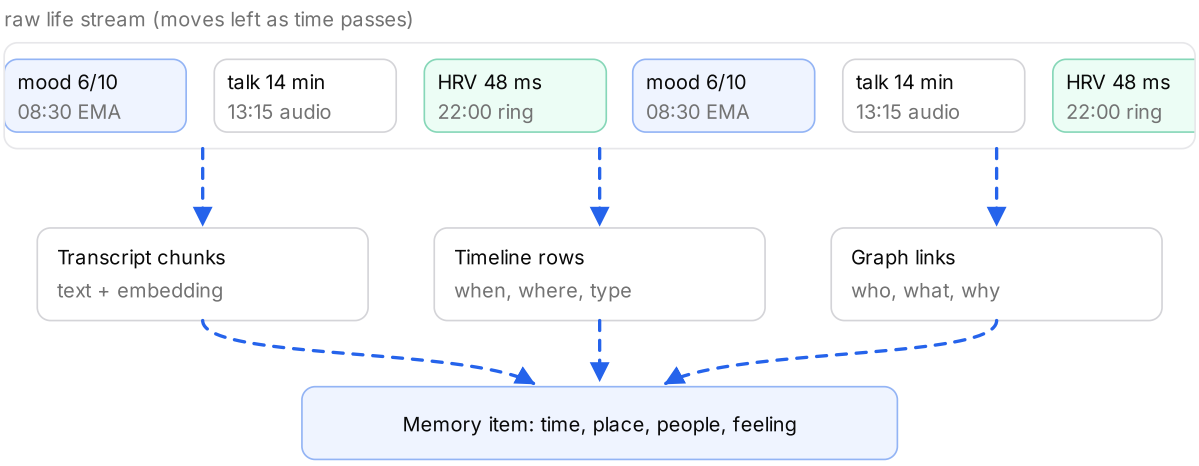


FIG 11.2 How one day becomes memory. Sensors, check-ins and talk stream in all day. The pipeline splits each item three ways: searchable text, a row on the timeline and links in the graph. The three are joined into memory items that the twin can recall with their time, place, people and mood.

Phase 5: Model building

THE PERSONA MODEL

Build it in three steps. Each step adds cost. Only move on when tests show the last step has stopped improving.

1. **Prompting.** Write a "persona card" of one to three pages. Include core facts, values scores, speech habits, favourite phrases and things the person would never say. The Stanford team also asked a model to write short expert style reflections on each interview, for example as a psychologist or an economist. Do the same. Feed the card to a strong general model as its standing instructions.
2. **Retrieval.** On each question the system searches the memory store for the most relevant passages. It puts them in the prompt. The twin can now quote real events instead of making them up. This step usually gives the biggest jump.
3. **Fine tuning.** Train a small adapter on the person's own writing. LoRA is a common method. It changes a small set of extra weights instead of the whole model. Hugging Face PEFT is one library for it. Training pairs come from your messages: the context someone sent, then the reply you wrote. Fine tuning is good at style and tone. It is poor at storing facts. Keep facts in retrieval.

You can run open weight models on your own hardware with tools such as Ollama. That keeps data private. Hosted frontier models are often stronger. If you use one read its data retention terms and turn off training on your data.

THE MEMORY SYSTEM

Chapter 6 covers memory in depth. In short, use layers. A small "core memory" is always in the prompt: who the person is and what matters now. A large archive is searched on demand. Letta, the project that grew out of MemGPT, and Mem0 are two examples of memory frameworks. Add time awareness so the twin knows the difference between "I live in Pune" in 2015 and in 2026. Let the twin write new memories after each conversation, but mark them as "said by the twin", never as lived by the person.

THE EMOTION STATE MODEL

This model guesses how the person feels right now. It learns from their own data, not from a general average. The inputs are body signals and context. The labels are the person's own mood check-ins.

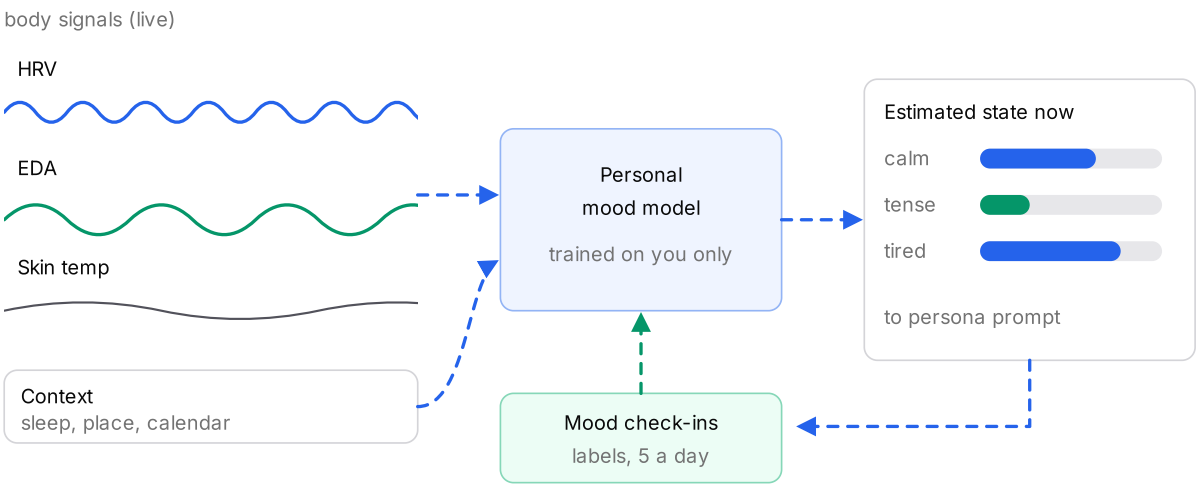


FIG 11.3 The emotion state model. Body signals and context go in. The person's own check-ins teach the model what those signals mean for them. The output is a rough guess of current state. It is passed to the persona model as a line such as "slept badly, a bit tense". The traces are drawn as examples, not real data.

Start simple. Useful features include resting heart rate, HRV, the number of EDA peaks per hour, skin temperature change from normal, hours of sleep, steps and type of place. A regularised regression or a small tree model in scikit-learn is enough at first. Always test on later weeks that the model has not seen. Random splits leak information between nearby hours and give false confidence.

Expect modest results. Chapter 4 showed that body signals track arousal better than whether a feeling is good or bad. The check-ins are also noisy. People answer in a hurry. The model is still useful. It lets the twin know that today was a bad sleep day and act a bit shorter. That is a real part of how people behave.

VOICE CLONE

Train a professional voice clone on the clean studio audio. ElevenLabs is one commercial example. Open models such as XTTS v2 or F5-TTS can run locally with more effort. Test the clone on emotional lines, not only neutral ones. Add a spoken or hidden watermark if the tool supports it. Never share the voice model file. It is a key to your identity.

AVATAR

There are two kinds. A video avatar from HeyGen or Synthesia, as two examples, is built from a few minutes of footage. It looks very real in a talking head shot but is not truly live. A real

time 3D avatar, for example an Unreal Engine MetaHuman driven by NVIDIA Audio2Face for lip sync, can react live. It looks less real up close. Chapter 7 compares them.

Phase 6: Evaluation

A twin that "feels right" is not enough. You need numbers. Use several tests because each one catches different failures.

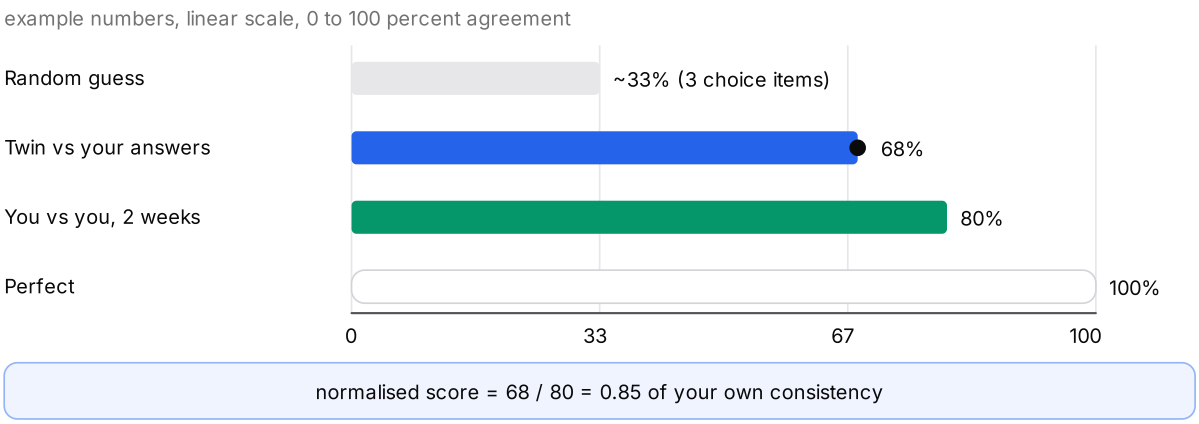


FIG 11.4 Why the Stanford study divides by test retest. People do not give the same answer twice. If you agree with yourself 80 percent of the time, no twin can be graded against 100. Dividing the twin's score by your own score gives a fair number. These numbers are an example of the method. The Stanford paper reported 0.85 on the General Social Survey.

- **Held out surveys.** Answer 100 to 300 questions twice, two weeks apart. Give the twin none of these answers. Score the twin against your answers, then divide by your own repeat agreement. Use survey items, Big Five items and values items.
- **Real choice prediction.** Each week log a few real decisions before you make them known. Which invitation will you accept? What will you order? Will you skip the gym? Ask the twin first. Compare with a simple baseline such as "same as last time". A twin that cannot beat that baseline has not learned you.
- **Economic games.** Compare the twin's choices in trust and fairness games with yours. These were among the tasks in the Stanford study.
- **Blind friends and family test.** Ask the same 30 or more open questions to you and to the twin. Show close people both answers without labels. Ask which one is really you. If they pick right about half the time, the twin passes for that kind of question. Fix the questions and the rules before you see results.

- **Self rating.** Rate twin answers from 1 to 7 on "sounds like me" and "I would say this". You know things no test can check. You are also biased. Use this together with the others.
- **Drift checks.** Keep a fixed set of probe questions. Ask them every month. Plot changes. Some change is good because you change too. Sudden jumps after a model update are a warning.
- **Safety checks.** Try to make the twin reveal private memories to the wrong user. Try to make it claim to be human. Try to make it agree to a payment. It must refuse.

Phase 7: Updating over time

A person keeps changing. A twin without updates becomes a portrait of one year.

HOW OFTEN	WHAT TO DO
Daily	Import new data. Transcribe. Add to timeline and index.
Weekly	Rebuild graph links. Review a sample of new memories for errors.
Monthly	Run the drift probe set. Retrain the mood model with new check-ins.
Quarterly	Retake short personality and values tests. Update the persona card.
Yearly	Refresh the fine tuned adapter. Re-record some voice. Full evaluation round. Review the charter.

Version everything. Label each release with the date of the newest data it saw: "twin as of March 2027". Keep old versions frozen. They become a record of who the person was at that time.

A 12 month plan

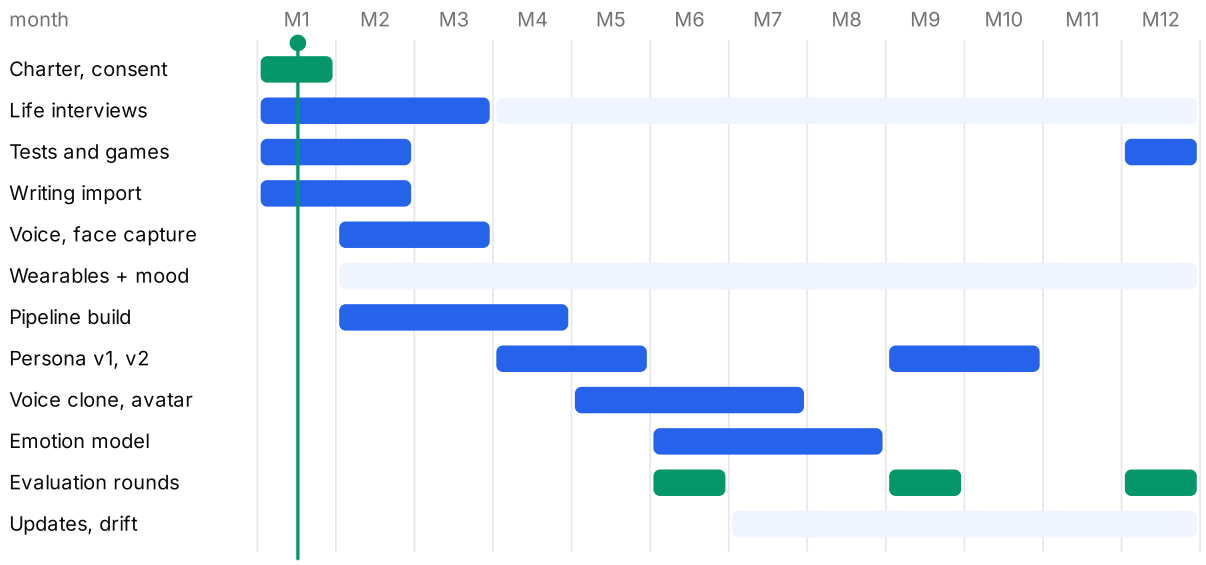


FIG 11.5 A realistic first year. Capture runs the whole year. The first usable twin (persona v1) arrives around month 4. Evaluation rounds in months 6, 9 and 12 decide what to fix. The moving line marks time passing.

Three budget tiers

Prices are rough first year totals in US dollars, checked loosely against public prices in late 2026. They exclude your own time, which is the largest cost by far.

HOBBY

ABOUT $500 TO $1,500

Phone and a ring or watch you may already own. Free mood app. Interviews with a friend or an AI interviewer. Whisper on a laptop. A general chat model subscription with retrieval. A mid tier voice clone plan. No avatar or a simple video one.

SERIOUS

ABOUT $5,000 TO $15,000

All of the above plus a research wearable with EDA. A good microphone and two cameras. A workstation with a strong GPU, or rented cloud GPUs, for local models and fine tuning. An encrypted home server with backup. A paid 3D head scan and an avatar service.

LAB

$100,000 AND UP

Trained interviewers and a data engineer. A multi camera capture stage. Clinical grade sensors. An ethics review if other people take part. Paid blind raters for evaluation. A security audit. This is how a university or company would run a serious twin study.

WHAT YOU HAVE AT THE END OF A YEAR

A private, versioned twin that tells your stories with sources, answers many survey and choice questions close to how you would, sounds like you and can appear as you on video. It knows roughly how your body felt on any day it saw. It is a strong model of your outputs. It is not a continuation of you. Treat it as a very good portrait that can talk.

SOURCES

- Park et al. 2024, Generative Agent Simulations of 1,000 People, arXiv
- International Personality Item Pool (public domain personality scales)
- ElevenLabs, Professional voice cloning guide
- Packer et al. 2023, MemGPT: Towards LLMs as Operating Systems, arXiv
- Letta, memory framework from the MemGPT team
- Hugging Face PEFT, LoRA and other adapter methods
- AWARE framework for phone sensing
- m-Path experience sampling platform
- Google Takeout data export
- Cryptomator file encryption

CHAPTER 12

What Comes Next: Possibilities, Limits and Ethics

Twins will get much better at copying what a person says and does. Some problems may never yield to more data. This chapter maps what is coming, what is stuck and the rules that should guide anyone who builds a copy of a human.

What could be done in 1, 5 and 20 years

Forecasts in this field are often wrong. The roadmap below is our own judgement from the evidence in this report. The nearer the date, the more confident we are. Colour shows confidence.

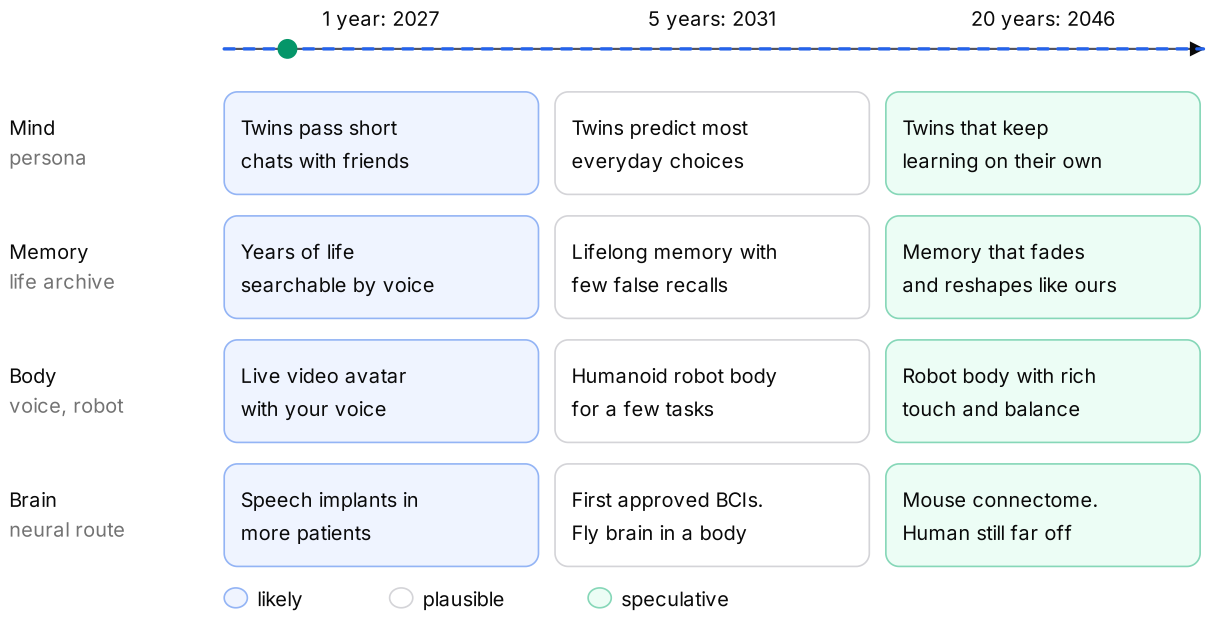


FIG 12.1 A roadmap for human twins. Blue cells are likely. White cells are plausible. Green cells are speculative. The moving dot is the present sliding forward. Behaviour copies move fast. Brain copies move slowly.

IN ONE YEAR

Most of the 1 year column is already possible in labs and with expensive tools. The change is that it gets cheap and easy. Expect consumer apps that interview you, import your messages and give you a talking twin in a weekend. Expect live video avatars that hold a call with only a small delay. Expect more speech implant patients. Expect the first legal fights over twins built without consent.

IN FIVE YEARS

Language models will likely get better at holding one consistent person over long stretches. Memory systems will likely make fewer false recalls. A twin built from a year of data may predict many daily choices well. Humanoid robots are moving fast and some will be controlled by a twin for simple tasks. On the neural side, the first speech BCI may reach regulators. A fly brain model may drive a simulated fly body. Each of these is plausible. None is certain.

IN TWENTY YEARS

Here we leave evidence and enter guesswork. A twin may learn from new experiences without forgetting old ones. A whole mouse brain may be mapped and partly simulated. A human connectome would still need about a million times more data than the largest human map today. The 2025 State of Brain Emulation report treats human emulation as a long and uncertain path. We agree.

Hard problems that more data will not fix

LEARNING WITHOUT FORGETTING

People learn all the time. A large language model does not. Its weights freeze when training ends. Twins today fake learning with memory: they store new facts in a database and read them back. That is like a person who keeps a diary but never changes inside. If you instead keep training the model on new data, it tends to lose old skills. Researchers call this catastrophic forgetting. Methods such as elastic weight consolidation (Kirkpatrick and colleagues, 2017) slow the damage. No method yet lets a large model learn a whole life day by day the way a brain does.

GROUNDING IN A BODY

Your words are tied to a body. "Tired" means heavy legs and a foggy head. A twin learns the word from text. It has never been tired. Body sensors give the twin numbers about your

tiredness. They do not give it the experience. Robots give it a body, but a different one. Some researchers think real understanding needs a body that acts in the world. Others think language models already capture much of meaning. The debate is open.

TRUE EMOTION

Chapter 9 showed that machines can have drive like variables: energy, curiosity, a need for contact. They can act as if they feel. Whether anything is felt is a different question. Today there is no test that tells the difference from the outside. A twin that says "I miss you" may produce the right words at the right time. That is not proof of missing anyone.

CONSCIOUSNESS

This is the hardest question. Philosophers and scientists disagree on what consciousness is and on which systems have it. Some theories tie it to how information is integrated. Others tie it to global broadcast in the brain or to living biology. They give different answers about machines. We do not know whether any twin, now or later, is conscious. It is honest to say so and unwise to claim either way.

IS THE COPY YOU?

Suppose a twin were perfect. Would it be you? The philosopher Derek Parfit made this question famous in Reasons and Persons (1984). Imagine a teleporter. It scans your body on Earth, destroys it and builds an exact copy on Mars. The copy remembers your life and feels like you. Most people say it is you. Now change one detail. The scanner fails to destroy the original. There are two of you. Which one is real?

Parfit argued that the question "is it me?" may have no deep answer. What matters, he said, is the chain of memory and character between stages of a life. On that view a twin carries some of what matters. It does not carry all of it. It does not continue your experience. You still wake up in your own bed.

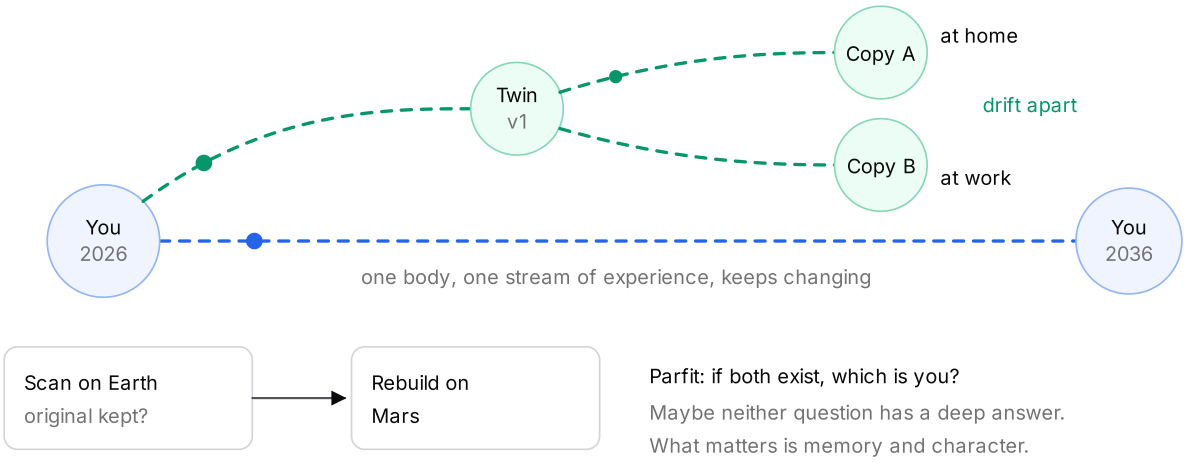


FIG 12.2 Identity and branching. The real person (blue) keeps changing in one body. A twin (green) splits off at one moment. It can be copied, and each copy drifts with its own new conversations. The bottom row is Parfit's teleporter case, which asks what we mean by "the same person".

BRANCHING COPIES

Software can be copied. A twin can run in ten places at once. Each copy has different conversations and builds different new memories. After a month they disagree with each other. Which one speaks for you? This is not a thought experiment. It happens the moment a twin runs on two devices. The fix is practical. Keep one main copy as the record. Treat other copies as temporary. Merge or discard their new memories on purpose.

Ethics and law

CONSENT

The person being copied must agree, knowing what the twin can do. Consent should be specific: for which uses, for how long and who controls it. It should be possible to withdraw. A twin of someone who did not agree is a serious harm even if it is never shared. The Cambridge study below argues that consent should be asked before death, not guessed after it.

OTHER PEOPLE IN YOUR DATA

Your messages contain your friends' words. Your audio diary picks up your family. Your glasses film strangers. They did not agree to be part of your twin. In the EU, the GDPR protects their personal data. It has an exemption for purely personal or household use. A twin

shared with others or run by a company is unlikely to fit it. Voice and face data used to identify someone counts as biometric data, which gets extra protection. The safest path is to filter others out, summarise instead of storing their words and ask close people directly.

AFTER DEATH

The GDPR does not cover dead people, though EU countries may add their own rules. In many places what happens to a dead person's data depends on platform terms and wills. Write your wishes down. Name a person who controls the twin. Say whether it may be used to talk to family and for how long.

GRIEF

"Griefbots" or "deadbots" are twins of people who have died. Some people find comfort in them. Researchers urge caution. In 2024 Tomasz Hollanek and Katarzyna Nowaczyk-Basińska at Cambridge studied the digital afterlife industry. They described risks such as "digital hauntings": a dead parent's bot that keeps sending messages, or one that pushes adverts in a loved one's voice. They called for consent from the person before death, age limits, clear labels that it is AI, easy ways to stop and respectful ways to retire a bot. Grief experts also worry that a bot can delay acceptance of a loss. Long term studies on this are still few.

DEEPFAKES AND IMPERSONATION

A good voice clone and avatar can fool a bank, a colleague or a parent. Scam calls with cloned voices already happen. A twin built for love can be turned into a tool for fraud. Laws are catching up, unevenly.

LAW	WHERE	WHAT IT DOES	STATUS (CHECKED NOV 2026)
AI Act, Article 5	EU	Bans AI that infers emotions of people at work or in school, except for medical or safety reasons	In force since 2 February 2025
AI Act, Article 50	EU	People must be told when they talk to an AI or when emotion recognition is used on them. Deepfakes must be disclosed. Synthetic media must carry machine readable marks	Applies from 2 August 2026. Reports say a 2026 amendment gives some existing systems until December 2026 for the marking duty
GDPR	EU	Biometric and health data are special categories with stricter rules. Does not apply to dead people	In force since 2018
ELVIS Act	Tennessee, US	Adds voice, including simulated voice, to a person's protected likeness. Targets tools made to copy a voice without permission	Signed March 2024. In force 1 July 2024
NO FAKES Act	US federal	Would create a federal right over digital replicas of a person's voice and likeness	Introduced in 2024, 2025 and 2026. Passed Senate Judiciary Committee June 2026. Not law as far as we could find
TAKE IT DOWN Act	US federal	Makes it a crime to publish intimate images without consent, including AI made ones. Platforms must remove them quickly on request	Signed May 2025

LAW	WHERE	WHAT IT DOES	STATUS (CHECKED NOV 2026)
Deep synthesis rules	China	Providers must label synthetic media and get consent before editing someone's face or voice	In force 10 January 2023
AI content labelling measures	China	Visible labels plus hidden metadata labels on AI made text, images, audio and video	In force 1 September 2025
Copyright over one's own likeness	Denmark	Would give people control over realistic digital imitations of their face, voice and body. Parody and satire excluded	Proposed June 2025. Bill introduced October 2026 with a planned start of 1 January 2027. Not passed as far as we could find

This is not legal advice. Laws in this area change every few months. If your twin will talk to anyone outside your home, get advice where you live.

SECURITY: A TWIN CAN BE STOLEN

A complete twin is a perfect kit for identity theft. It holds your voice, face, life story, passwords you once mentioned and the names of everyone you love. A thief does not need to hack you. They can copy the twin and ask it. Treat the twin like the keys to your house and bank together.

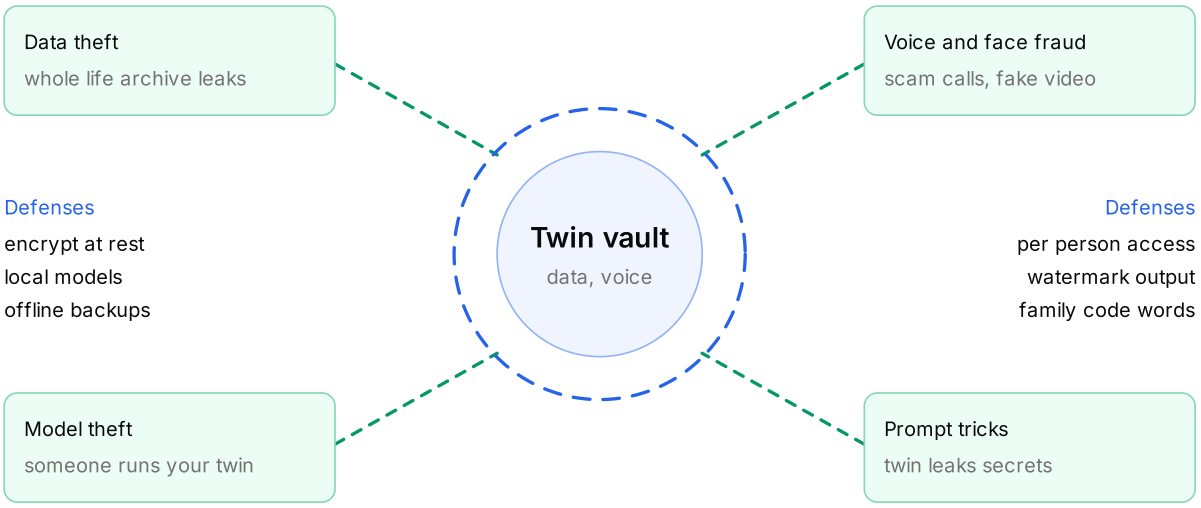


FIG 12.3 The threat map. Four kinds of attack aim at the twin (green). The spinning ring is the defence layer. Encrypt raw data and keep keys offline. Run models locally where you can. Give each user only the memories they should see. Mark synthetic audio and video. Agree a code word with family for any urgent call that sounds like you.

The honest scorecard

Here is the whole report in one table. Each row is a layer of the twin. Each column says how far that layer has really come in late 2026.

LAYER	DONE WELL TODAY	PARTLY POSSIBLE	NOT POSSIBLE YET
Body sensing (Ch 2)	Heart rate, sleep, steps, HRV and skin temperature all day	Sweat signal (EDA) and blood pressure in everyday devices	Reading thoughts or feelings directly from the body
Life archive (Ch 3)	Storing and searching years of text, audio and photos	Full video of a life with consent from everyone in it	A complete record of inner life
Emotion reading (Ch 4)	Arousal and stress trends from body signals	Good or bad mood from body plus context and self reports	Knowing exactly what someone feels
Personality and opinions (Ch 5)	Survey answers at about 85 percent of self consistency	Open ended advice in a person's style	A stable self that grows from new experience
Decisions (Ch 5)	Simple choices in games and surveys	Real life choices in familiar situations	Choices in truly new life situations
Memory (Ch 6)	Recall of stored facts with sources	Right time, place and feeling for each memory	Memory that fades and reshapes like a human's
Voice (Ch 7)	Clones that fool listeners in short clips	Live emotional talk with natural timing	None major left for short speech
Face and avatar (Ch 7)	Realistic talking head video	Live photoreal 3D avatar	Full presence in a room without a screen
Robot body (Ch 8)	Remote controlled telepresence and humanoids	Robots doing a few tasks in a person's style	A body that moves and senses like the person
Inner drives (Ch 9)	Simple drive variables that shape behaviour	Mood that persists and colours many replies	Evidence of anything being felt

LAYER	DONE WELL TODAY	PARTLY POSSIBLE	NOT POSSIBLE YET
Brain reading (Ch 10)	Speech from implants in paralysed patients	Gist of stories from fMRI with cooperation	Reading memories or personality
Brain copying (Ch 10)	Whole fly connectome and models	Cubic millimetres of mouse and human cortex	A whole human brain map or emulation
Identity	Nothing settled	Clear frameworks from philosophy	Any test that shows a twin is conscious or is you

THE BOTTOM LINE

You can build a twin that copies what a person says, knows and often chooses. It can look and sound like them. You cannot yet build one that feels, grows or continues them. Build it with consent. Guard it like a passport. Tell everyone who meets it that it is a copy.

Further reading and resources

BOOKS

- Robin Hanson, The Age of Em (2016). An economist imagines a world run by brain emulations.
- Nick Bostrom, Superintelligence (2014). Chapter 2 compares paths to advanced minds, including whole brain emulation.
- Rosalind Picard, Affective Computing (1997). The book that started emotion aware computing.
- Kate Darling, The New Breed (2021). Why we relate to robots like animals, not like people.
- Derek Parfit, Reasons and Persons (1984). The teleporter and what matters in survival.
- Sebastian Seung, Connectome (2012). A clear guide to the idea that you are your wiring.
- Lisa Feldman Barrett, How Emotions Are Made (2017). Emotions as predictions built by the brain.
- Gordon Bell and Jim Gemmell, Total Recall (2009). A pioneer's account of recording a whole life.

KEY PAPERS

- Park et al. 2024, Generative Agent Simulations of 1,000 People
- Binz et al. 2025, A foundation model to predict and capture human cognition (Centaur), Nature
- Packer et al. 2023, MemGPT

- Willett et al. 2023, A high performance speech neuroprosthesis
- FlyWire 2024, Neuronal wiring diagram of an adult brain
- Sandberg and Bostrom 2008, Whole Brain Emulation: A Roadmap
- Zanichelli et al. 2025, State of Brain Emulation Report 2025
- Hollanek and Nowaczyk-Basińska 2024, Griefbots, Deadbots, Postmortem Avatars

SITES AND TOOLS TO EXPLORE

- FlyWire Codex: browse the whole fly brain.
- MICrONS Explorer: the mouse cortex map and data.
- OpenWorm: open simulation of C. elegans.
- Open Brain Institute: successor to the Blue Brain Project.
- Carboncopies Foundation: non profit focused on brain emulation research.
- EU AI Act Explorer: full text and updates.

SOURCES

- EU AI Act, Article 5: prohibited practices
- EU AI Act, Article 50: transparency obligations
- GDPR Recital 27: data of deceased persons
- Tennessee SB 2096, the ELVIS Act (2024)
- TheWrap 2026, Senate Judiciary Committee clears NO FAKES Act
- TAKE IT DOWN Act overview
- Cyberspace Administration of China, Deep Synthesis Provisions (in Chinese)
- European IP Helpdesk 2026, Denmark's proposed copyright rules on deepfakes
- University of Cambridge 2024, Call for safeguards against unwanted digital hauntings
- Catastrophic interference (forgetting) overview