

Multimodal Electronic Skin Architectures for Physical AI: A Comparative Framework of Latex Silicone, CNT/Graphene Network, Hydrogel–Liquid-Metal based Auxetic Printed Circuit Sensors

Dongchan Lee*

AI & DX Center, Institute for Advanced Engineering, Yongin-si 11780, Republic of Korea.

*Correspondence:

Dongchan Lee, AI & DX Center, Institute for Advanced Engineering, Yongin-si 11780, Republic of Korea Tel: +82-10-4798-9771, Fax: +82-31-330-7117.

Received: 16 Aug 2026; Accepted: 24 Sep 2026; Published: 02 Oct 2026

Citation: Dongchan Lee. Multimodal Electronic Skin Architectures for Physical AI: A Comparative Framework of Latex Silicone, CNT/Graphene Network, Hydrogel–Liquid-Metal based Auxetic Printed Circuit Sensors. Nano Tech Appl. 2026; 9(3): 1-13.

ABSTRACT

Electronic skin (e-skin) is emerging as a key enabling technology for humanoid robots, wearable robots, prosthetic systems, and embodied Physical AI. Unlike conventional tactile sensors designed primarily for localized pressure measurement, future robotic skins must simultaneously accommodate large mechanical deformation, distributed tactile sensing, multimodal perception, large-area electrical routing, and intelligent interpretation of physical interactions. However, no single material or structural architecture can optimally satisfy all of these requirements. A systematic framework comparing alternative e-skin architectures is therefore necessary for selecting appropriate sensing and interconnection strategies for different robotic applications.

This paper presents a comparative framework for four representative multimodal electronic-skin architectures intended for Physical AI systems: (1) latex–silicone-based auxetic stretchable printed-circuit e-skin, (2) CNT/graphene-based auxetic conductive e-skin, (3) hydrogel–liquid-metal–auxetic large-area electronic skin, and (4) an AI-integrated multimodal whole-body e-skin architecture. The first architecture employs geometrically deformable auxetic and serpentine conductive pathways combined with neutral-plane structural engineering to accommodate stretching, bending, twisting, pressure, and shear deformation. The second architecture utilizes carbon nanotube and graphene conductive networks to provide mechanically compliant and distributed strain-sensitive pathways. The third combines hydrogel interfaces, liquid-metal conductors, and auxetic geometries to achieve extreme deformability and large-area conformability. The fourth extends these material and structural technologies toward distributed multimodal perception by integrating tactile, proprioceptive, thermal, vibration, and deformation information with hierarchical artificial intelligence.

A unified electromechanical formulation is introduced to describe strain transfer, resistance variation, multimodal sensor response, sensor fusion, and AI-based tactile interpretation across the four architectures. Their relative characteristics are compared in terms of stretchability, bending and torsional compliance, pressure/shear sensitivity, electrical stability, self-healing potential, large-area manufacturability, signal-routing complexity, and suitability for Physical AI. Based on this comparison, a hierarchical design methodology is proposed in which different e-skin technologies can be selectively deployed over fingertips, hands, joints, limbs, and torso regions according to local mechanical and perceptual requirements.

Rather than identifying a single universally optimal electronic-skin technology, this work establishes the materials–structure–sensing–intelligence design framework for future robotic skin. The proposed perspective provides a reference architecture for researchers developing whole-body tactile interfaces, humanoid robots, wearable robotics, intelligent prostheses, and embodied Physical AI systems.

Keywords

Electronic skin (e-skin), Auxetic structure, Stretchable electronics, CNT-graphene, Hydrogel-Liquid metal, Multimodal tactile sensing, Physical AI, Embodied intelligence.

Introduction

Human skin is not merely a protective covering but a highly distributed, multimodal sensory interface. Through densely distributed mechanoreceptors and thermoreceptors, biological skin simultaneously detects pressure, shear, vibration, temperature, texture, deformation, and complex contact events. These sensory signals are continuously integrated with proprioceptive information by the nervous system to support manipulation, locomotion, balance, environmental perception, and safe physical interaction. Replicating these capabilities has therefore been a major objective of electronic-skin (e-skin) research for robotics, prosthetics, and wearable systems [1-3]. Early e-skin technologies primarily focused on flexible pressure sensing, whereas more recent research has emphasized stretchability, conformability, multifunctionality, and large-area sensor integration [2-6].

Replicating biological-skin-like functionality is particularly challenging for humanoid robots. Future humanoids are expected to operate in human-centered environments and interact directly with objects, tools, and people. Visual perception alone is insufficient for these tasks because many critical physical states—including contact force, friction, incipient slip, local deformation, compliance, and grasp stability—cannot be reliably determined from visual information alone. Distributed tactile perception must therefore become an integral component of embodied robotic intelligence. At the same time, whole-body robotic skin must remain functional while undergoing stretching, bending, twisting, compression, and shear, creating a fundamental trade-off between mechanical compliance, sensor density, electrical interconnection, and sensing performance [3,6].

Electronic-skin technologies have consequently evolved from isolated pressure sensors toward flexible, stretchable, and multimodal sensing platforms. Piezoresistive, capacitive, piezoelectric, triboelectric, magnetic, optical, ionic, and liquid-metal-based sensing mechanisms have been extensively investigated. Nevertheless, achieving biological-skin-like sensing over an entire humanoid body remains difficult because no single sensing material or structural architecture simultaneously provides high mechanical adaptability, multimodal sensitivity, electrical reliability, large-area scalability, and straightforward system integration.

Auxetic structures provide a promising mechanical strategy for addressing this challenge. Unlike conventional materials, auxetic structures exhibit a negative effective Poisson's ratio and can expand laterally under tensile deformation [7-9]. In rotating-unit auxetic structures, substantial macroscopic deformation can be accommodated primarily through geometric rotation of structural cells rather than excessive local material strain [8]. This

mechanism can redistribute mechanical deformation and reduce strain concentration in conductive pathways, making auxetic geometries particularly attractive for stretchable interconnections and mechanically adaptive robotic e-skin.

Carbon-based nanomaterials provide a complementary pathway toward intrinsically deformable sensing and conduction. Carbon nanotube (CNT) networks, for example, can maintain electrical conductivity under large deformation while exhibiting strain- and pressure-dependent resistance changes [10]. Graphene and CNT/graphene hybrid networks can therefore potentially function simultaneously as conductive pathways and deformation-sensitive sensing layers. Such multifunctionality is particularly attractive for high-resolution tactile regions, where sensor density and wiring complexity must be minimized without sacrificing mechanical flexibility.

For highly compliant and large-area interfaces, ionic hydrogels and liquid metals offer additional advantages. Stretchable ionic conductors can sustain substantial deformation while maintaining ionic transport [11], whereas gallium-based liquid metals provide metallic conductivity together with exceptional deformability [12]. Their integration with soft substrates and mechanically adaptive structures provides a potential route toward conformal sensing and signal routing over highly curved or continuously deforming humanoid surfaces. However, encapsulation, environmental stability, leakage prevention, long-term reliability, and heterogeneous material integration remain important engineering challenges.

In parallel with advances in materials and structures, e-skin is undergoing a transition from passive sensing toward intelligent physical perception. Bioinspired artificial afferent systems have demonstrated that tactile sensor responses can be transformed into neural-like representations [13], while neuro-inspired robotic skins have begun to combine distributed sensing with local processing and intelligent interpretation [14]. Furthermore, attention-based deep-learning architectures provide a powerful computational framework for identifying spatial, temporal, and cross-modal dependencies within large-scale sensory data [15]. These developments suggest that future robotic skin should be regarded not simply as a flexible sensor array, but as a distributed physical perception interface capable of supporting semantic understanding and adaptive robot behavior.

From this perspective, a whole-body humanoid e-skin system must simultaneously satisfy five closely coupled requirements:

- Mechanical adaptability to stretching, bending, twisting, compression, and shear;
- Multimodal sensing of heterogeneous mechanical, thermal, and deformation-related stimuli;
- Electrical reliability during repeated and large mechanical deformation;
- Large-area scalability with manageable wiring, multiplexing, and signal acquisition; and
- Intelligent interpretation of distributed tactile information for perception and control.

These requirements cannot be optimized independently. Increasing sensor density, for example, generally increases the number and complexity of conductive interconnections, potentially reducing mechanical compliance. Highly conductive materials may provide excellent electrical performance but insufficient stretchability, whereas extremely soft ionic or liquid conductors may accommodate large deformation while introducing challenges associated with encapsulation, environmental stability, fabrication, and integration. Accordingly, future robotic skin should not necessarily be designed as a homogeneous single-material sensor system. Instead, it can be regarded as a hierarchical materials–structure–sensor–electronics–intelligence system, in which different technologies are selected according to the mechanical and perceptual requirements of individual body regions.

Based on this concept, the present study considers four complementary e-skin architectures. The first architecture consists of a latex–silicone-based substrate combined with auxetic structures and deformable printed conductive pathways. Neutral-plane placement and curved or serpentine routing are employed to reduce strain concentration in electrically vulnerable regions and to accommodate stretching, bending, twisting, pressure, and shear. Within the proposed design framework, representative target deformation conditions include a bending radius of approximately 14 mm, tensile deformation of approximately 20%, and torsional deformation of up to approximately 120°. These values are treated as design targets rather than experimentally validated performance limits.

The second architecture employs CNT/graphene-based auxetic flexible conductive networks. In this configuration, mechanical deformation and electrical response are intrinsically coupled through changes in conductive pathways, contact resistance, and tunneling characteristics. The architecture is intended to accommodate stretching, bending, and twisting while enabling pressure, strain, shear, texture, and slip-related sensing. Such characteristics make carbon-based e-skin particularly suitable for high-resolution tactile regions such as humanoid fingertips and palms.

The third architecture combines hydrogel, liquid-metal conductors, and auxetic structures to provide highly compliant, large-area robotic skin. The hydrogel provides a soft and deformable interface, while liquid-metal channels enable stretchable electrical routing. Auxetic structural design can further redistribute mechanical deformation across the sensing surface. This architecture is therefore particularly relevant to large and highly deformable body regions, including the torso, back, arms, and legs, where conformability and distributed contact detection may be more important than extremely high spatial resolution.

The fourth architecture extends e-skin from a distributed sensing surface to an AI-integrated multimodal perceptual system. Pressure, shear, vibration, temperature, deformation, and proprioceptive information can be encoded and fused to construct semantic representations of physical interaction. Rather than directly

mapping individual sensor outputs to predefined thresholds, the resulting architecture aims to infer higher-level states such as contact, incipient slip, stable grasp, texture, human touch, collision, and body deformation and subsequently connect these states to adaptive robot decision-making and action generation.

The central hypothesis of this paper is therefore that different humanoid body regions require different e-skin technologies, but their heterogeneous sensory information should converge into a common multimodal Physical AI framework. High-resolution CNT/graphene sensing can be concentrated at the fingertips and palms; mechanically robust auxetic printed circuits can be deployed around joints and limbs; hydrogel–liquid-metal systems can provide conformal large-area coverage over the torso and other soft-contact regions; and AI-based multimodal integration can unify these heterogeneous sensory streams into a whole-body representation of physical interaction.

Accordingly, the principal contributions of this paper are as follows:

- ① A unified classification of four complementary e-skin architectures for humanoid robots and Physical AI;
- ② A comparative electromechanical framework linking material properties, auxetic structures, conductive networks, sensing mechanisms, and deformation modes;
- ③ A body-region-specific design strategy for selecting and integrating heterogeneous e-skin technologies;
- ④ A multimodal sensor-fusion framework connecting distributed electronic skin with semantic tactile-state understanding and adaptive robot control; and
- ⑤ A technological roadmap extending conventional e-skin from local physical sensing toward whole-body tactile intelligence and embodied Physical AI.

Through this framework, electronic skin is repositioned from a collection of flexible sensors to a distributed physical perception system capable of connecting contact, sensing, representation, reasoning, and action. This transition provides the basis for integrating advanced e-skin technologies with tactile representation learning, multimodal foundation models, proprioception, and adaptive control, ultimately enabling humanoid robots to perceive and interpret physical interaction at the whole-body level. Figure 1. Conceptual framework and technological evolution of multimodal electronic skin toward embodied Physical AI. Four complementary technological platforms—latex–silicone auxetic printed circuits, CNT/graphene conductive networks, hydrogel–liquid-metal systems, and AI-integrated multimodal sensing—are combined within a heterogeneous whole-body e-skin architecture. Auxetic and stretchable structures accommodate stretching, bending, twisting, compression, and shear, while distributed sensing provides pressure, shear, texture/vibration, temperature, and proprioceptive/deformation information. Multimodal sensory signals are transformed through feature extraction, attention- or Transformer-based fusion, and semantic tactile-state interpretation into physical interaction states including contact, slip, grasp stability, texture, human touch, collision, and deformation.

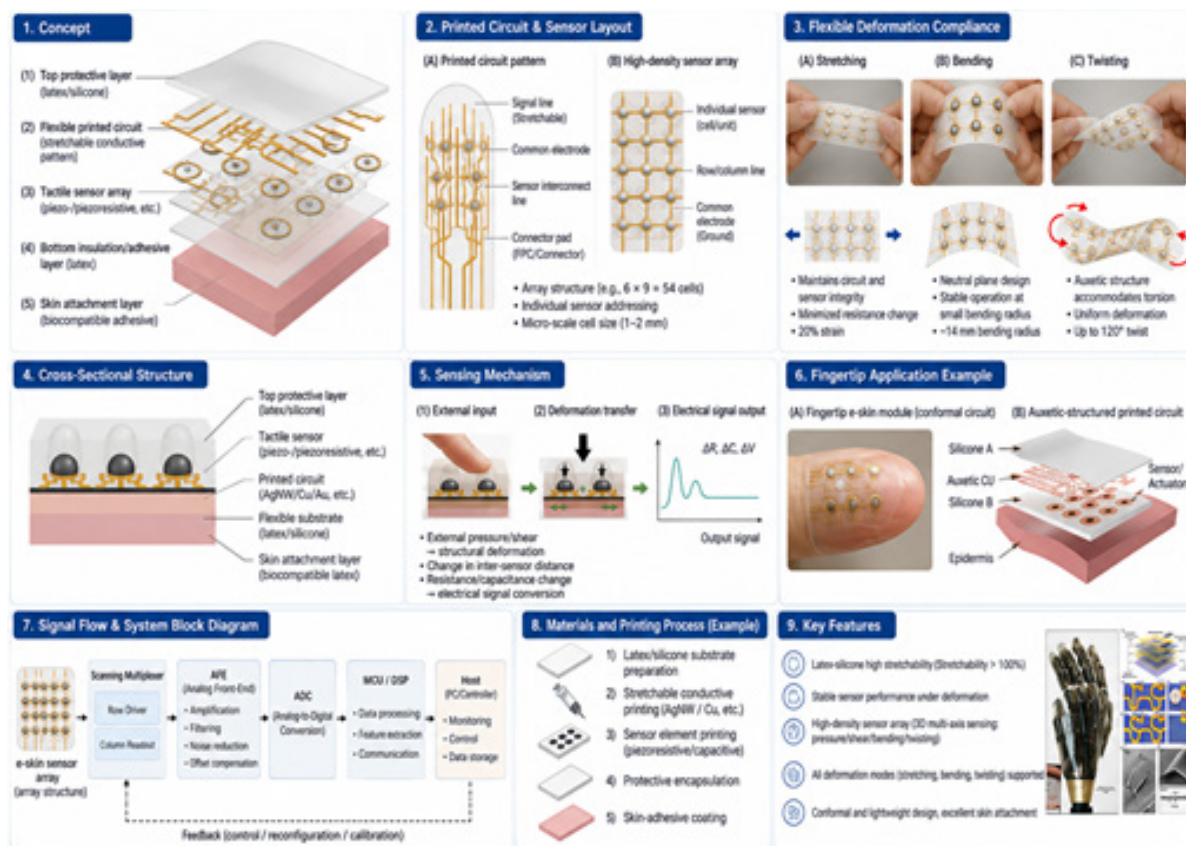


Figure 1: (a) Latex-Silicone based Auxetic flexible printed circuit e-skin.

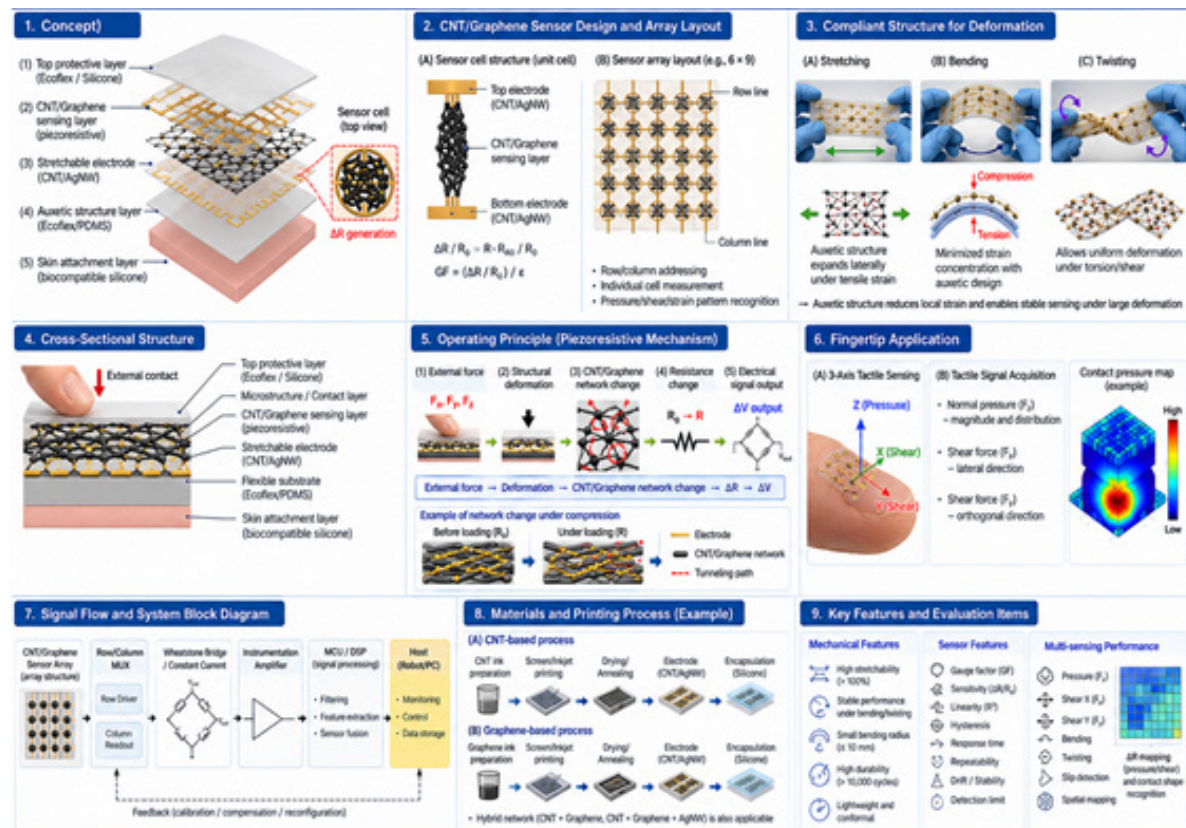


Figure 1: (b) CNT-Graphene based Auxetic printed circuit e-skin.

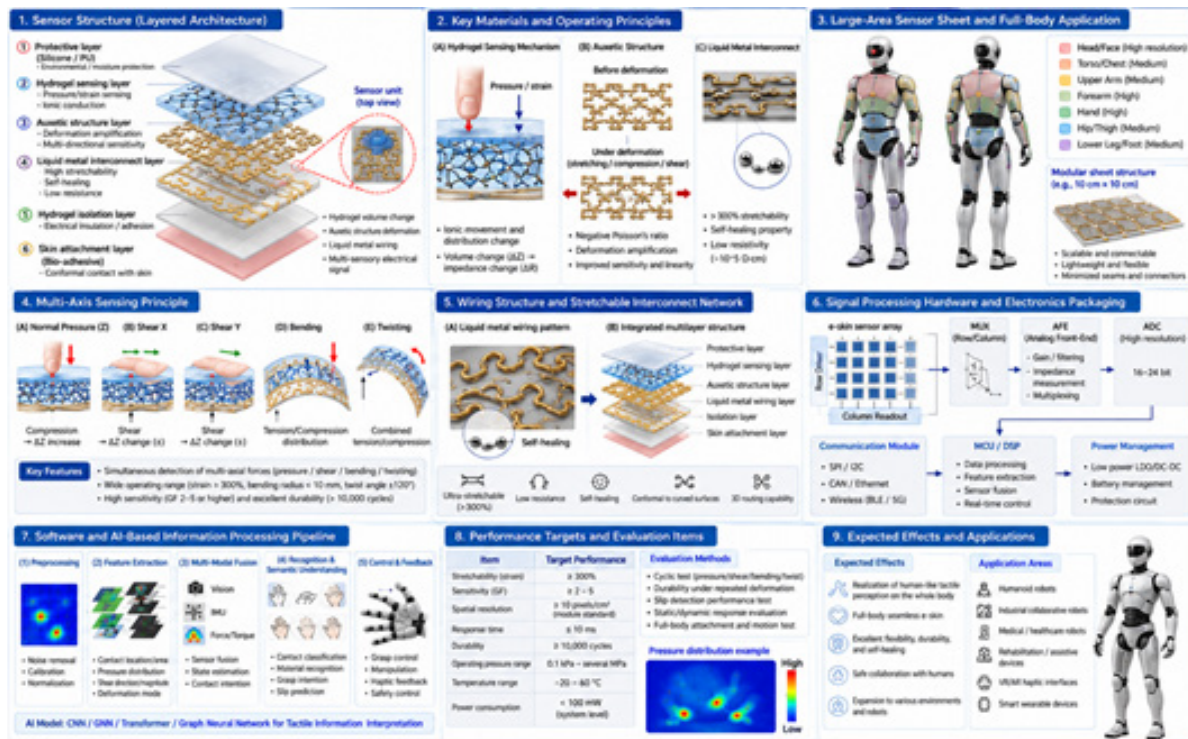


Figure 1: (c) Hydrogel-Liquid Metal based auxetic circuit printed e-skin.

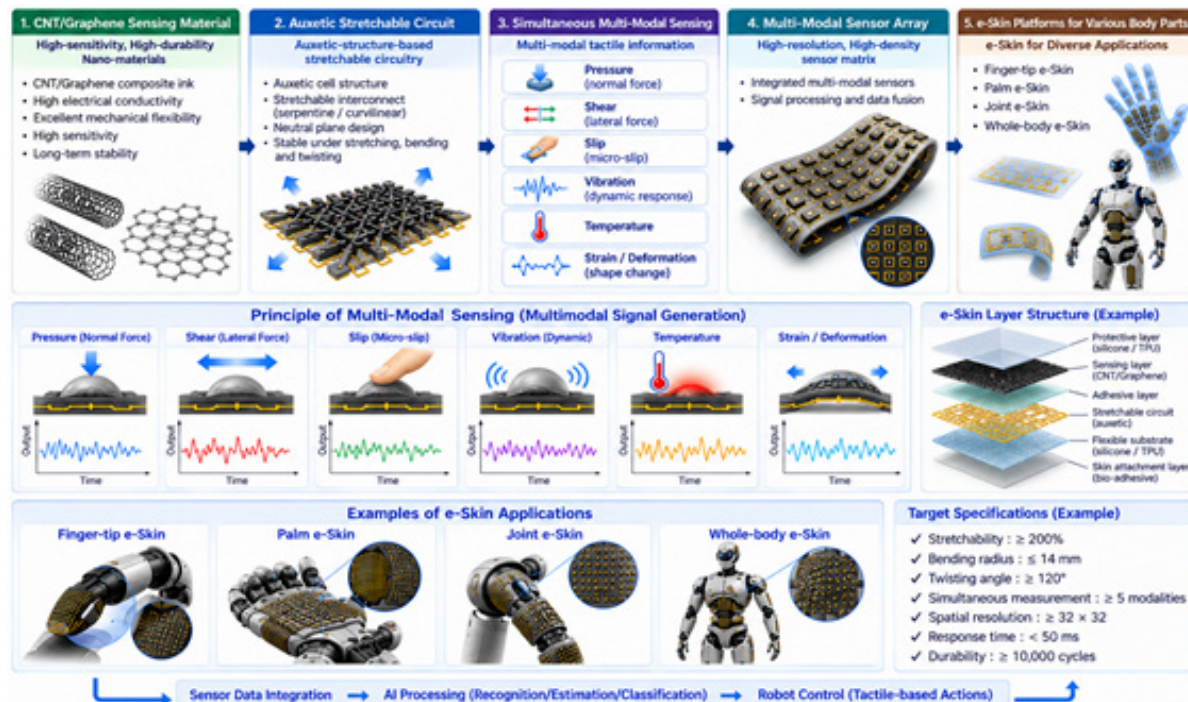


Figure 1: (d) Multimodal e-skin sensor array.

Figure 1: Four representative multimodal electronic skin (e-skin) architectures for humanoid robots and whole-body Physical AI. (A) Latex–silicone-based auxetic printed-circuit e-skin employing neutral-plane structural design and serpentine/spiral conductive routing to accommodate stretching, bending, twisting, pressure, and shear while minimizing strain-induced electrical degradation. (B) CNT/graphene-based auxetic conductive e-skin combining intrinsically stretchable carbon-based conductive networks with auxetic structures for high-resolution pressure, strain, texture, and slip sensing under large deformation. (C) Hydrogel–liquid-metal–auxetic e-skin integrating an ionic hydrogel interface, deformable liquid-metal conductors, and auxetic structural layers to provide ultra-high stretchability, conformability, large-area tactile coverage, and mechanically robust signal routing. (D) AI-integrated multimodal e-skin architecture that combines distributed pressure, shear, temperature, vibration, and strain/proprioceptive sensing with signal preprocessing, feature extraction, attention-based multimodal fusion, semantic tactile-state interpretation, and adaptive decision and action generation.

Region-specific e-skin technologies are assigned according to the mechanical and perceptual requirements of the fingertips, palms, articulated joints, limbs, torso, and back, while whole-body sensory integration establishes a perception–reasoning–action loop for embodied Physical AI. Future developments toward self-powered sensing, wireless and edge electronics, tactile foundation models, and adaptive whole-body robotic intelligence are also envisioned.

Unified Design Framework for Multimodal Electronic Skin

General Requirements

For a deformable robotic surface, the mechanical state can be represented approximately by

$$\boldsymbol{\varepsilon} = [\varepsilon_x \quad \varepsilon_y \quad \varepsilon_z \quad \gamma_{xy} \quad \gamma_{yz} \quad \gamma_{zx}]^T \quad (1)$$

where $\varepsilon_x, \varepsilon_y, \varepsilon_z$ denote normal strains and $\gamma_{xy}, \gamma_{yz}, \gamma_{zx}$ denote engineering shear strains.

A robotic e-skin should maintain electrical functionality over a deformation domain

$$\Omega_D = \{\varepsilon, \kappa, \theta, p, \tau\} \quad (2)$$

where ε is tensile strain, κ is surface curvature, θ is torsional deformation, p is normal pressure, and τ is shear stress.

This formulation is important because conventional e-skin studies frequently optimize only one quantity—for example pressure sensitivity—while humanoid skin experiences all five simultaneously.

Latex–Silicone Auxetic e-Skin

The first architecture uses a soft latex–silicone substrate combined with auxetic, serpentine, spiral, or curved conductive pathways. The concept is particularly suitable for articulated regions where conventional printed conductors experience high strain concentration. The attached architecture specifically targets stretching, bending, twisting, pressure, and shear, with neutral-plane placement of the conductive network as an important structural feature. Auxetic structures exhibit an effective negative Poisson's ratio and can redistribute deformation more uniformly than conventional geometries [7–9]. Their mechanical behavior makes them suitable for stretchable electronic interconnects exposed to repeated multidirectional deformation.

The effective Poisson's ratio of an auxetic mechanical structure is

$$\nu_{eff} = \frac{\varepsilon_{transverse}}{\varepsilon_{axial}} \quad (3)$$

For an auxetic structure,

$$\nu_{eff} < 0 \quad (4)$$

Consequently, tensile deformation can be accommodated by geometric rotation, opening, or unfolding of the unit cells rather than exclusively through material elongation. This is highly advantageous for printed electronics because

$$\varepsilon_{trace} < \varepsilon_{skin} \quad (5)$$

can be achieved through geometrical strain isolation. During bending, longitudinal strain varies approximately as

$$\varepsilon_b(z) = \kappa(z - z_n) \quad (6)$$

where z is the through-thickness position, z_n is the neutral-plane position, and κ is curvature.

Therefore,

$$z \rightarrow z_n \Rightarrow \varepsilon_b \rightarrow 0 \quad (7)$$

Locating conductive traces near z_n can substantially reduce bending-induced conductor fatigue. This principle explains why auxetic geometry and neutral-plane engineering should be treated as complementary rather than competing strategies. Rotating-square mechanisms provide a representative example in which macroscopic expansion is achieved through rotation of relatively stiff structural units [8]. This mechanism is attractive for robotic skin because conductive traces can be mechanically isolated from the full macroscopic deformation of the substrate.

CNT/Graphene Auxetic e-Skin

The second architecture replaces or supplements conventional metallic conductors with carbon-based nanomaterials, particularly carbon nanotubes (CNTs) and graphene. The attached design uses this approach in an auxetic flexible circuit intended to accommodate stretching, bending, twisting, pressure, and shear. CNT and graphene conductive networks are particularly attractive for electronic skin because their resistance is strongly affected by structural deformation and conductive-network rearrangement. Carbon-nanotube films have been demonstrated as highly stretchable electrodes and pressure/strain sensing layers [10].

For a resistive carbon-network sensor,

$$R = \rho \frac{L}{A} \quad (8)$$

and its normalized deformation response is

$$\frac{\Delta R}{R_0} = GF\varepsilon \quad (9)$$

where (R_0) is the initial resistance and (GF) is the effective gauge factor.

However, CNT/graphene networks are more complicated than an ideal metallic strain gauge. Their electrical response includes changes in conductive pathways, contact resistance, tunnelling distances, and percolation states.

A generalized relation can therefore be written as

$$R_{eff} = f(R_c, R_t, N_p, d_t, \varepsilon) \quad (10)$$

where R_c denotes contact resistance, R_t tunnelling resistance, N_p the number of conductive pathways, and d_t the characteristic tunnelling distance.

This architecture has an important conceptual advantage. The conductor can simultaneously become the sensor. This differs fundamentally from Architecture I, in which mechanical routing and sensing elements may remain functionally separated. Because the conductive layer itself responds to deformation, CNT/graphene e-skin can integrate sensing and electrical routing within a single structural layer [10].

Hydrogel–Liquid-Metal–Auxetic e-Skin

The third architecture combines hydrogel, liquid-metal conductors, and auxetic mechanical structures. The source concept specifically targets a whole-body large-area tactile sensor, wiring, and signal-processing system capable of multidirectional deformation. Liquid metals are particularly attractive because electrical continuity can be retained while the microchannel itself undergoes large deformation. Highly deformable ionic conductors provide a promising alternative to conventional metallic wiring. Stretchable transparent ionic conductors have demonstrated excellent mechanical compliance under large deformation [11]. Similarly, liquid metals based on gallium alloys retain metallic conductivity while undergoing extreme shape change [12].

For a liquid-metal channel,

$$R_{LM} = \rho_{LM} \frac{L_c}{A_c} \quad (11)$$

where ρ_{LM} is liquid-metal resistivity, L_c is channel length, and A_c is channel cross sectional area.

Under deformation,

$$\frac{R}{R_0} = \frac{L/L_0}{A/A_0} \quad (12)$$

Consequently, deformation of the microfluidic geometry itself produces measurable resistance modulation. Hydrogels can simultaneously provide soft ionic interfaces. Their conductivity can be represented approximately by

$$\sigma_{ion} = \sum_i n_i q_i \mu_i \quad (13)$$

where n_i , q_i , and μ_i denote the concentration, charge, and mobility of ionic species i , respectively.

Because the liquid conductor follows deformation of the surrounding microchannel, high stretchability can be achieved without conventional conductor fracture [12]. The combination of hydrogel interfaces, liquid-metal wiring, and auxetic mechanical structures therefore provides a promising architecture for large-area humanoid torso, limb, and joint coverage. Combining ionic hydrogels, liquid-metal conductors, and auxetic structural geometries therefore provides a potential route toward extremely compliant robotic skin.

Multimodal Intelligent e-Skin for Physical AI

Traditional e-skin systems primarily convert contact stimuli into electrical measurements. Physical AI requires an additional step in which distributed tactile information is converted into semantic physical understanding. The development of bioinspired artificial

afferent nerves has demonstrated that flexible tactile devices can reproduce elements of biological sensory signal transmission [13]. More recently, neuro-inspired robotic electronic skin has shown that tactile sensing and intelligent processing can be combined to provide robot-oriented physical perception [14]. The attached framework already identifies a multi-sensory AI integrated control system incorporating tactile, proprioceptive, and thermal information. A multimodal sensory state can be defined as

$$\mathbf{x}_t = [\mathbf{p}_t, \boldsymbol{\tau}_t, \mathbf{v}_t, \mathbf{T}_t, \boldsymbol{\varepsilon}_t, \mathbf{q}_t] \quad (14)$$

where $\mathbf{p}$ is pressure, $\boldsymbol{\tau}$ shear, $\mathbf{v}$ vibration, $\mathbf{T}$ temperature, $\boldsymbol{\varepsilon}$ deformation, and $\mathbf{q}$ proprioceptive state.

Each modality can be encoded as,

$$\mathbf{h}_i = f_i(\mathbf{x}_i; \theta_i) \quad (15)$$

Multimodal fusion then produces

$$\mathbf{Z} = \sum_{i=1}^M \alpha_i \mathbf{h}_i, \sum_{i=1}^M \alpha_i = 1 \quad (16)$$

where α_i represents learned modality importance.

For attention-based fusion,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (17)$$

The semantic tactile state becomes

$$\mathbf{s}_t^{\text{tactile}} = g_\theta(\mathbf{Z}_t) \quad (18)$$

which may represent states such as

{contact, slip, stable grasp, texture, human touch, collision}

Finally,

$$\mathbf{a}_t = \pi_\theta(\mathbf{s}_t^{\text{tactile}}, \mathbf{s}_t^{\text{robot}}, \mathbf{s}_t^{\text{env}}) \quad (19)$$

links tactile perception directly to robot action. This is the key transition from electronic skin to Physical-AI skin.

The four architectures should not be regarded as mutually exclusive technologies. Instead, they represent complementary solutions to different regions of the robotic body. Auxetic printed circuits are particularly advantageous for articulated joints because strain can be redistributed geometrically [7-9]. CNT/graphene networks are suitable for fingertip and palm sensing because they combine stretchable conductivity with direct pressure and strain sensitivity [10]. Hydrogel and liquid-metal technologies provide superior deformability for large-area and highly compliant body surfaces [11,12]. AI-integrated tactile systems provide the computational layer required to transform distributed sensory measurements into physical understanding and adaptive action [13-15]. A future humanoid robot may therefore employ heterogeneous e-skin rather than a single universal material system.

Body-Region-Specific Hybrid e-Skin

Table 1: Comparative Analysis of the Four Architectures.

Characteristic	Latex–Silicone Auxetic PCB	CNT/Graphene Auxetic	Hydrogel–Liquid Metal	AI-Integrated Multimodal Skin
Primary role	Reliable routing	Conductive sensing	Extreme deformability	Physical understanding
Stretchability	High	High	Very high	Architecture dependent
Bending	Excellent	Excellent	Excellent	Architecture dependent
Twisting	Excellent	Excellent	Excellent	Architecture dependent
Pressure sensing	Sensor dependent	High potential	High potential	Multimodal
Shear sensing	Sensor dependent	High potential	High potential	Multimodal
Conductivity	High	Medium–high	Very high	—
Sensor/conductor integration	Moderate	High	High	System level
Large-area scalability	High	High	Very high	Very high
Self-healing potential	Low	Moderate	High	Algorithmic compensation
Fabrication complexity	Low–moderate	Moderate	Moderate–high	High
AI compatibility	High	High	High	Native
Best application	Joints/limbs	Hands/fingers	Torso/large surfaces	Whole-body integration

A heterogeneous whole-body e-skin architecture is proposed instead of employing a single sensor technology over the entire humanoid body. Figure 2 illustrates the proposed whole-body multimodal electronic skin architecture in which heterogeneous e-skin technologies are strategically distributed across different humanoid body regions according to their functional requirements. Unlike conventional robotic skins based on a single sensing material or uniform sensor configuration, the proposed approach recognizes that fingertips, palms, articulated joints, limbs, and large torso surfaces experience fundamentally different mechanical loads and require different tactile sensing resolutions.

At the fingertips, CNT/graphene-based high-resolution electronic skin is employed because these regions require high spatial sensitivity for fine object interaction. The sensing layer is intended to detect local pressure distributions, fine surface textures, and incipient micro-slip, which are essential for stable manipulation of fragile or small objects. For the palm, CNT/graphene sensing elements are combined with auxetic stretchable printed circuits. This configuration enables the palm to measure distributed normal pressure and tangential shear forces while preserving electrical continuity during hand opening, closing, and grasping motions. The resulting tactile information can be used to estimate grasp stability and detect the onset of object slip. The wrist, elbow, and knee represent highly deformable regions exposed to repeated bending, stretching, and torsional motion. Therefore, latex–silicone-based auxetic printed circuits are assigned to these joints. Their auxetic and serpentine geometries redistribute mechanical strain and reduce local stress concentration, while the deformation pattern itself can provide proprioceptive information regarding joint configuration and body posture. The arms and legs require a balance between mechanical compliance and large-area tactile coverage. In these regions, auxetic printed circuits are integrated with distributed tactile sensing nodes to detect contact forces, deformation patterns, and interaction locations over relatively large surfaces. Such distributed sensing can contribute to collision detection, human–robot physical interaction, and whole-body environmental awareness. For the torso and back, a hydrogel–liquid-metal-based electronic skin is proposed because these

regions require highly deformable, conformal, and large-area sensing surfaces. The hydrogel provides a soft and compliant interface, while liquid-metal conductors maintain electrical connectivity during substantial deformation. This architecture is particularly suitable for detecting broad-area soft contact, body support, or human physical interaction.

At the system level, sensory information acquired from all body regions is integrated through a multimodal AI framework. Local tactile measurements are first converted into representative features and then fused using attention-based or learned multimodal integration mechanisms. The fused representation is subsequently interpreted as semantic tactile states, including contact, slip, stable grasp, surface recognition, human touch, collision, and body deformation. These semantic states are used by the control layer for grasp adjustment, force regulation, safe physical interaction, and adaptive behavior.

Consequently, the proposed architecture establishes a hierarchical pathway from distributed local sensing to global embodied intelligence. The heterogeneous whole-body e-skin therefore serves not only as a tactile measurement surface but also as a distributed perceptual interface that connects physical interaction, semantic understanding, adaptive control, and Physical AI.

Mathematically, the optimum skin architecture for body region r can be formulated as

$$A_r^* = \arg \max_{A_j} [w_1 S_j + w_2 D_j + w_3 C_j + w_4 R_j + w_5 I_j - w_6 \text{COST}_j] \quad (20)$$

where S represents sensing capability, D deformability, C conformability, R reliability, and I intelligence/integration capability.

This equation converts the paper from a conventional review into a design-selection framework. Future tactile learning systems may combine distributed e-skin data with vision, language, and proprioception. The development of attention-based architectures [15] and neuro-inspired electronic skin [14] provides an important foundation for this direction. Rather than searching for a single

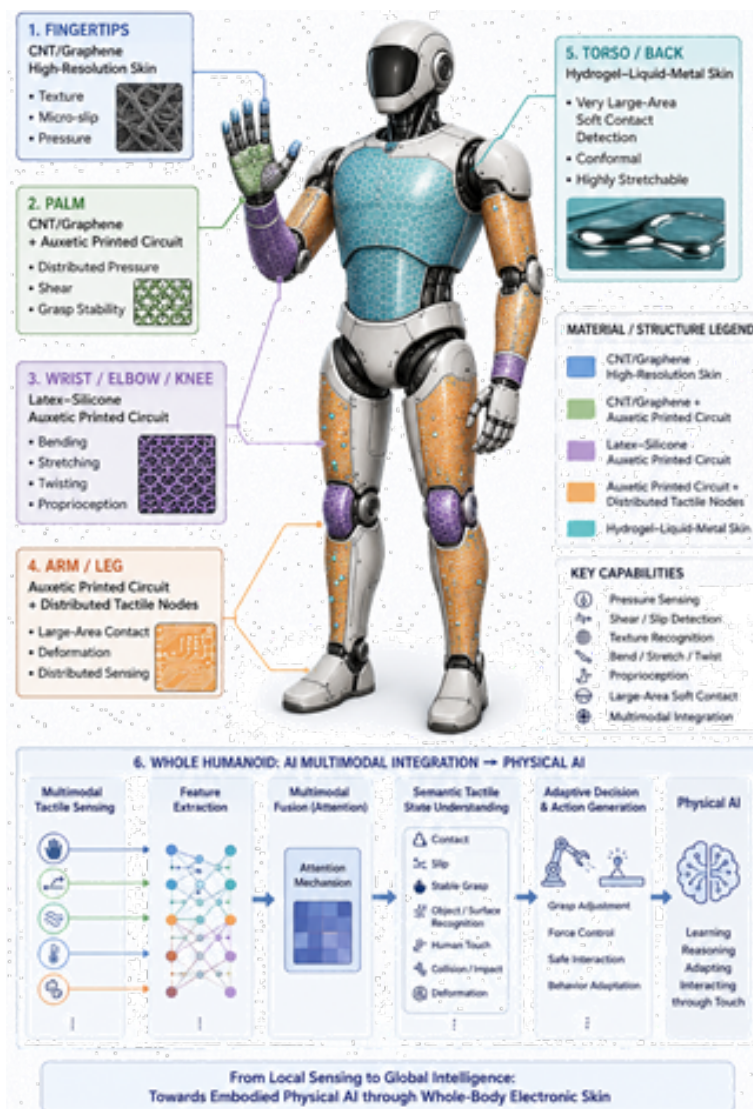


Figure 2: Body-region-specific multimodal electronic skin architecture for whole-body Physical AI. The proposed humanoid e-skin system adopts a heterogeneous, region-specific design strategy in which different material and structural platforms are assigned according to the local mechanical and tactile requirements of each body region. High-resolution CNT/graphene-based e-skin is applied to the fingertips for texture recognition, micro-slip detection, and pressure sensing. The palm integrates CNT/graphene sensing layers with auxetic printed circuits to support distributed pressure measurement, shear-force detection, and grasp-stability estimation. Highly articulated joints, including the wrist, elbow, and knee, employ latex–silicone-based auxetic printed circuits to accommodate bending, stretching, and twisting deformation while simultaneously providing proprioceptive information. The arms and legs utilize auxetic printed circuits combined with distributed tactile nodes for large-area contact detection, deformation monitoring, and spatially distributed sensing. For the torso and back, hydrogel–liquid-metal e-skin is proposed to provide highly conformable and stretchable large-area soft-contact sensing. At the whole-body level, multimodal tactile information is processed through feature extraction, attention-based sensor fusion, semantic tactile-state interpretation, adaptive decision-making, and action generation. The resulting unified architecture enables a transition from local tactile sensing to whole-body embodied intelligence and Physical AI.

universally optimal sensor technology, the key challenge is to determine how different material systems, mechanical structures, sensor modalities, and AI architectures can be integrated according to the local mechanical and perceptual requirements of different humanoid body regions.

From Electronic Skin to Tactile Foundation Models

The evolution of electronic skin is progressing from material-level sensing toward distributed physical perception and, ultimately, toward large-scale tactile intelligence. Conventional electronic skin

primarily converts local mechanical stimuli into electrical signals, whereas whole-body robotic skin generates high-dimensional spatiotemporal information representing pressure, shear force, vibration, temperature, deformation, proprioception, and contact dynamics. When such information is continuously acquired from heterogeneous e-skin architectures distributed over a humanoid body, the resulting dataset provides not only measurements of contact but also a rich representation of physical interaction. This transition creates an opportunity to develop **tactile foundation models (TFMs)** capable of learning transferable representations

of contact, object properties, manipulation states, human touch, and physical interaction.

In the proposed whole-body e-skin framework, different sensing technologies can be deployed according to local functional requirements. CNT/graphene-based high-resolution skins can provide fine tactile information at the fingertips and palms; auxetic printed circuits can accommodate large deformation around joints and limbs; hydrogel–liquid-metal architectures can provide compliant large-area sensing over the torso; and multimodal AI modules can integrate these heterogeneous signals into a unified representation. Consequently, a tactile foundation model should not be designed around a single sensor type. Instead, it should learn a sensor-independent or sensor-adaptive latent representation from heterogeneous tactile modalities.

Spatiotemporal Representation of Whole-Body Tactile Data

Consider a whole-body e-skin consisting of N sensing nodes, M sensing modalities, and a temporal observation window containing T samples. The tactile dataset can be represented by a third-order tensor

$$\mathbf{X} = [x_{t,n,m}] \in \mathbb{R}^{T \times N \times M}. \quad (21)$$

where $x_{t,n,m}$ denotes the measurement obtained at time t , sensor node n , and sensing modality m .

For the multimodal e-skin considered in this study, the local sensory state at node n can be expressed as

$$\mathbf{x}_{t,n} = \begin{bmatrix} p_{t,n} \\ \tau_{x,t,n} \\ \tau_{y,t,n} \\ v_{t,n} \\ T_{t,n} \\ \varepsilon_{t,n} \end{bmatrix}, \quad (22)$$

where p denotes normal pressure, τ_x and τ_y are tangential shear components, v represents vibration-related information, T is temperature, and ε represents local deformation or strain.

Whole-body proprioceptive information may additionally be represented by the robot joint state

$$\mathbf{q}_t = [q_{1,t}, q_{2,t}, \dots, q_{J,t}]^T, \quad (23)$$

where J is the number of robot joints. Combining tactile and proprioceptive information is particularly important because deformation generated by robot motion must be distinguished from deformation caused by external physical contact.

A generalized embodied sensory state can therefore be defined as

$$\mathbf{s}_t^{\text{emb}} = [\mathbf{X}_t, \mathbf{q}_t, \dot{\mathbf{q}}_t, \mathbf{F}_t^{\text{ext}}], \quad (24)$$

where $\dot{\mathbf{q}}_t$ represents joint velocity and $\mathbf{F}_t^{\text{ext}}$ denotes available external force information. This representation allows a learning system to interpret tactile events in relation to the robot's own physical state rather than treating skin measurements as isolated sensor signals.

Tactile Tokenization and Latent Representation Learning

Directly processing thousands of distributed tactile nodes is computationally inefficient. Similar to patch-based processing in vision models, whole-body tactile data can be partitioned into spatial and temporal regions and converted into **tactile tokens**.

For tactile region, an embedded representation can be expressed as

$$\mathbf{z}_{r,t} = \phi_\theta(\mathbf{X}_{r,t}) + \mathbf{e}_r^{\text{pos}} + \mathbf{e}_t^{\text{time}} + \mathbf{e}_m^{\text{mod}}, \quad (25)$$

where ϕ_θ is a learnable tactile encoder, $\mathbf{e}_r^{\text{pos}}$ represents spatial or body-region encoding, $\mathbf{e}_t^{\text{time}}$ is temporal encoding, and $\mathbf{e}_m^{\text{mod}}$ identifies the sensing modality.

This formulation is particularly useful for heterogeneous whole-body e-skin because a signal originating from a fingertip CNT/graphene sensor has a different physical interpretation from a large-area torso signal generated by a hydrogel–liquid-metal sensor. Explicit body-region and modality embeddings allow the model to preserve this contextual information while learning a common latent space.

For K tactile tokens, the input sequence becomes

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_K]. \quad (26)$$

Long-range spatial and temporal relationships between tactile regions can subsequently be learned using self-attention [15]:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (27)$$

With

$$\mathbf{Q} = \mathbf{Z}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{Z}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{Z}\mathbf{W}_V, \quad (28)$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are trainable projection matrices and d_k is the dimension of the key vectors. This mechanism allows the model, for example, to associate fingertip micro-slip with simultaneous palm shear-force redistribution or to interpret distributed torso contact together with changes in joint configuration.

Self-Supervised Pretraining for Tactile Foundation Models

A fundamental limitation of tactile AI is the cost of manually labeling physical interaction data. Whole-body robotic skin may generate millions of sensor samples during routine manipulation and human–robot interaction, but only a small fraction can realistically be manually annotated. Therefore, tactile foundation models should rely strongly on **self-supervised learning**.

A masked tactile modeling strategy can be defined by dividing the tactile tokens into observed and masked subsets:

$$\mathbf{Z} = \mathbf{Z}_{\text{vis}} \cup \mathbf{Z}_{\text{mask}}. \quad (29)$$

The model reconstructs the masked tactile information from the observed context,

$$\hat{\mathbf{Z}}_{\text{mask}} = D_\psi[E_\theta(\mathbf{Z}_{\text{vis}})], \quad (30)$$

where E_θ denotes the tactile encoder and D_ψ is the reconstruction

decoder. A reconstruction objective can then be defined as

$$L_{\text{mask}} = \frac{1}{|M|} \sum_{i \in M} \|z_i - \hat{z}_i\|_2^2, \quad (31)$$

where M denotes the set of masked tactile tokens. Temporal prediction can provide an additional self-supervised objective:

$$L_{\text{pred}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \|z_{t+1} - f_{\theta}(z_{\leq t})\|_2^2. \quad (32)$$

Such learning encourages the model to acquire representations of contact evolution, slip initiation, loading/unloading processes, and deformation dynamics without requiring explicit human-generated labels.

Multimodal Foundation Model for Physical Interaction

Tactile information becomes substantially more useful when combined with vision, language, and proprioception. A future Physical AI system can therefore construct a unified embodied representation

$$z_{\text{emb}} = F_{\theta}(z_{\text{tactile}}, z_{\text{vision}}, z_{\text{language}}, z_{\text{prop}}), \quad (33)$$

where z_{tactile} , z_{vision} , z_{language} , and z_{prop} denote tactile, visual, linguistic, and proprioceptive latent representations, respectively.

Rather than simple feature concatenation, cross-modal attention can be used to determine which modality is most informative for a particular physical event. For example, vision may identify an object before contact, whereas tactile information becomes dominant after grasping and during slip detection.

The resulting foundation model may learn a generalized mapping $F_{\text{TFM}}: (\mathbf{X}_{1:T}, \mathbf{V}_{1:T}, \mathbf{q}_{1:T}, \mathbf{I}) \rightarrow (\mathbf{c}, \boldsymbol{\eta}, \mathbf{a})$, (34)

where $\mathbf{V}_{1:T}$ represents visual observations, $\mathbf{I}$ represents language or task instruction, $\mathbf{c}$ is the semantic contact state, $\boldsymbol{\eta}$ represents estimated physical properties, and $\mathbf{a}$ denotes an action or action affordance.

The physical-property vector may include

$$\boldsymbol{\eta} = [\mu \quad k \quad c \quad r_{\text{slip}} \quad g_{\text{stable}}]^T, \quad (35)$$

where μ is friction coefficient, k is effective contact stiffness, c is damping, r_{slip} represents slip risk, and g_{stable} denotes grasp-stability information. Accordingly, the objective of tactile intelligence evolves from answering “What signal was measured?” to answering “What physical state produced the signal, what will happen next, and what action should the robot take?”

From Semantic Tactile Understanding to Robot Action

The semantic tactile state predicted by the foundation model can be represented as

$$c_t = g_{\theta}(z_{\text{emb},t}). \quad (36)$$

where c_t may encode contact, slip, stable grasp, collision, texture, human touch, or deformation states. Robot action is subsequently generated according to

$$a_t = \pi_{\omega}(z_{\text{emb},t}, c_t, s_t^{\text{robot}}), \quad (37)$$

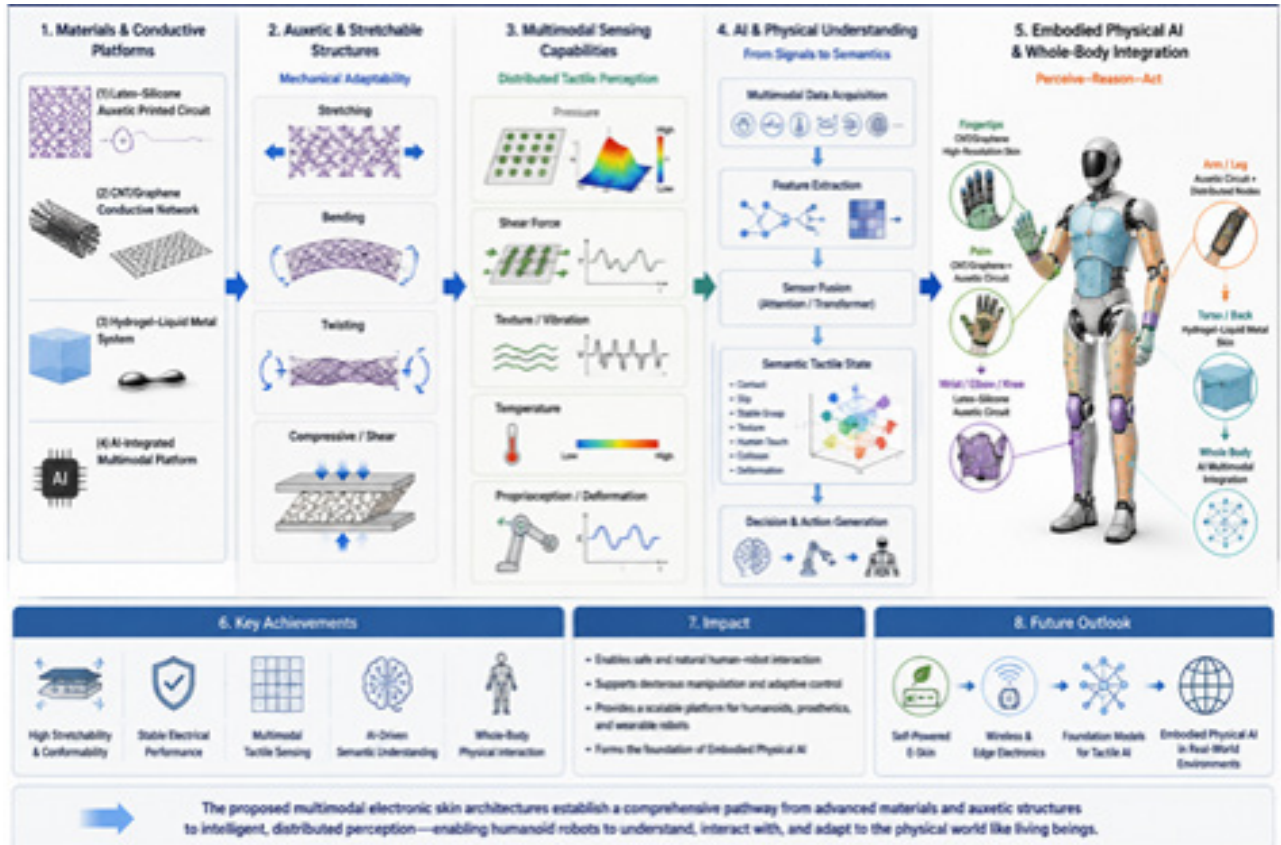


Figure 3: Conceptual summary and future evolution of multimodal electronic skin toward embodied Physical AI.

where π_{ω} is an embodied control policy. For grasp manipulation, for example,

$$r_{\text{slip}} \uparrow \Rightarrow F_{\text{grip}}^* \uparrow, \quad (38)$$

where F_{grip}^* denotes the desired grasp force. Conversely, detection of unexpected human contact may trigger compliant behavior,

$$P_{\text{human contact}} > P_{\text{th}} \Rightarrow K_{\text{robot}} \downarrow, \quad (39)$$

where $P_{\text{human contact}}$ is the predicted probability of human contact, P_{th} is a decision threshold, and K_{robot} represents the effective robot control stiffness. Thus, tactile foundation models provide an interface between distributed electronic skin and adaptive robot behavior.

Toward Generalizable Whole-Body Tactile Intelligence

An important objective of future tactile foundation models is **cross-sensor, cross-body-region, cross-object, and cross-robot generalization**. A model trained exclusively using one sensor material may otherwise learn material-specific electrical characteristics rather than transferable physical concepts.

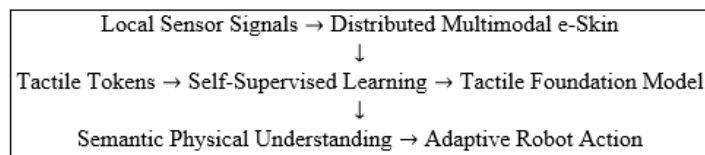
A desirable latent representation should therefore satisfy approximately

$$p(\mathbf{z} \mid c, d_1) \approx p(\mathbf{z} \mid c, d_2), \quad (40)$$

where c denotes the same physical contact state and d_1 and d_2 represent different sensor domains, materials, body regions, or robot platforms.

This sensor-independent representation is particularly important for the heterogeneous e-skin architecture proposed in this work. CNT/graphene sensors, auxetic printed circuits, hydrogel–liquid-metal skins, and other tactile nodes may exhibit substantially different raw electrical responses even when exposed to similar physical interactions. A foundation model should therefore learn the underlying **physical semantics** rather than simply memorize individual sensor characteristics.

Ultimately, the technological evolution can be represented as



This progression represents a fundamental shift in the role of electronic skin. Future e-skin will no longer function merely as a deformable array of tactile sensors; instead, it may become a distributed physical perception interface through which humanoid robots continuously acquire, organize, and learn from physical experience. When integrated with vision, language, proprioception, and robot control, tactile foundation models could provide the perceptual basis for Physical AI systems capable of learning through touch, transferring physical knowledge between tasks, and adapting their behavior to previously unseen objects,

environments, and human interactions.

Conclusion

This paper presents a unified framework for four complementary electronic-skin architectures for humanoid robots and Physical AI: latex–silicone auxetic printed-circuit skin, CNT/graphene-based auxetic electronic skin, hydrogel–liquid-metal–auxetic large-area skin, and AI-integrated multimodal intelligent skin. Rather than treating these technologies as competing alternatives, the proposed framework positions them within a common mechanical–electrical–sensory design space. Auxetic printed circuits provide mechanically robust and deformable signal-routing platforms for articulated robotic surfaces; CNT/graphene networks enable highly sensitive and stretchable tactile interfaces; hydrogel–liquid-metal systems provide exceptional compliance and conformability for large-area soft contact; and AI-integrated multimodal architectures transform distributed sensory signals into meaningful representations of physical interaction.

A central conclusion of this study is that future humanoid electronic skin is unlikely to rely on a single homogeneous sensing technology. Instead, a heterogeneous and body-region-specific architecture provides a more effective pathway toward whole-body tactile perception. High-resolution sensing can be concentrated at the fingertips and palms, mechanically compliant interconnects and deformation sensors can be deployed around articulated joints and limbs, and highly stretchable large-area sensing layers can cover the torso and back. Such region-specific integration allows sensor materials, spatial resolution, interconnection topology, and mechanical compliance to be optimized according to local functional requirements while maintaining a unified whole-body sensory representation.

More importantly, this framework extends the role of electronic skin beyond passive measurement toward distributed physical perception and embodied intelligence. Multimodal information—including pressure, shear, texture and vibration, temperature, and proprioceptive deformation—can be processed through feature extraction, attention- or Transformer-based sensor fusion, and semantic tactile-state interpretation. Consequently, raw sensor responses can be transformed into physically meaningful states such as contact, slip, grasp stability, texture, human touch, collision, and body deformation. Coupling these semantic representations with robot decision-making and adaptive control establishes a closed perception–reasoning–action loop in which electronic skin functions as an active interface between the robot and its physical environment.

The convergence of heterogeneous e-skin materials, mechanically adaptive structures, multimodal sensing, and artificial intelligence therefore provides a promising technological foundation for whole-body Physical AI. Future research should advance this framework toward self-powered and self-healing sensing materials, wireless and edge-based distributed electronics, scalable whole-body sensor networks, and tactile foundation models capable of learning transferable physical representations across sensors,

body regions, objects, tasks, and robotic platforms. Ultimately, intelligent whole-body electronic skin could enable humanoid robots to move beyond merely detecting physical contact toward understanding, predicting, and adapting to physical interactions, thereby supporting safer human–robot interaction, more dexterous manipulation, and autonomous operation in complex real-world environments.

Funding

This research received financial support from the Industry Innovation Infrastructure Project (RS-2024-00439808, Wearable Robot Demonstration Center) funded by the Korea Institute for Advancement of Technology (KIAT) and the Ministry of Trade, Industry, & Energy (MOTIE).

References

1. Hammock, ML, Chortos A, Tee BC, et al. 25th Anniversary Article: The Evolution of Electronic Skin (E-Skin): A Brief History, Design Considerations, and Recent Progress. *Adv Mater*. 2013; 25: 5997-6038.
2. Chortos A, Liu J, Bao Z, et al. Pursuing Prosthetic Electronic Skin. *Nat Mater*. 2016; 15: 937-950.
3. Dahiya RS, Metta G, Valle M, et al. Tactile Sensing—From Humans to Humanoids. *IEEE Transactions on Robotics*. 2010; 26: 1-20.
4. Kim DH, Nanshu Lu, Rui Ma, et al. Epidermal electronics. *Science*. 2011; 333: 838-843.
5. Hua Q, Sun J, Liu H, et al. Skin-Inspired Highly Stretchable and Conformable Matrix Networks for Multifunctional Sensing. *Nature Communications*. 2018; 9: 244.
6. Dahiya R, Yogeswaran N, Liu F, et al. Large-Area Soft e-Skin: The Challenges Beyond Sensor Designs. *Proceedings of the IEEE*. 2019; 107: 2016-2033.
7. Evans KE, Nkansah MA, Hutchinson IJ, et al. Molecular Network Design. *Nature*. 1991; 353: 124.
8. Grima JN, Evans KE. Auxetic Behavior from Rotating Squares. *Journal of Materials Science Letters*. 2000; 19: 1563-1565.
9. Ingrole A, Hao A, Liang R, et al. Design and Modeling of Auxetic and Hybrid Honeycomb Structures for In-Plane Property Enhancement. *Materials & Design*. 2017; 117: 72-83.
10. Lipomi DJ, Vosgueritchian M, Tee BCK, et al. Skin-Like Pressure and Strain Sensors Based on Transparent Elastic Films of Carbon Nanotubes. *Nat Nanotechnol*. 2011; 6: 788-792.
11. Keplinger C, Sun JY, Foo CC, et al. Stretchable, Transparent, Ionic Conductors. *Science*. 2013; 341: 984-987.
12. Dickey MD. Stretchable and Soft Electronics Using Liquid Metals. *Adv Mater*. 2017; 29: 1606425.
13. Kim Y, Chortos A, Xu W, et al. A Bioinspired Flexible Organic Artificial Afferent Nerve. *Science*. 2018; 360: 998-1003.
14. Liu F, Deswal S, Christou A, et al. Neuro-Inspired Electronic Skin for Robots. *Sci Robot*. 2022; 7: eabl7344.
15. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*. 2017; 30.