

Paradigm shift in biomedical engineering driven by artificial intelligence: a review of technological breakthroughs, integration pathways, and ethical challenges

Hongming Huang

Hangzhou No. 14 High School, Hangzhou, China

80429527@qq.com

Abstract. The entry of Artificial Intelligence (AI) into Biomedical Engineering (BME) is not merely the addition of a new research method; rather, it is triggering a systemic transformation of the discipline's research paradigm. Problem formulation is shifting from "hypothesis-driven" to "data discovery coupled with hypotheses"; experimental design is moving from "decomposing systems into isolated components" to "multiscale integrated modeling"; and evaluation criteria are evolving from "causal mechanism first" to "predictive performance combined with interpretability." This paper first defines five characteristics of the traditional BME paradigm, then reviews its three developmental stages. Through the four domains of medical imaging, Brain-Computer Interfaces (BCI), biomaterials, and molecular pharmaceuticals, it illustrates the concrete changes brought by AI and the problems that remain unresolved. It identifies the core tension of the data-driven paradigm as "prediction relies on correlation, whereas explanation relies on causality," and accordingly organizes the challenges into four layers: algorithmic, engineering, privacy, and ethical. This paper devotes a separate section to the ethical responsibility of medical AI and the progress of domestic and international regulation. This review finds that the AI-triggered paradigm shift is partial and incomplete, its fundamental tension being strong correlative prediction but weak causal mechanisms; this conclusion can serve as a reference for researchers in handling the relationship between data-driven modeling and causal explanation.

Keywords: artificial intelligence, biomedical engineering, research paradigm, brain-computer interface, ethical regulation

1. Introduction

Before discussing whether Artificial Intelligence (AI) has brought about a shift in the traditional BME research paradigm, it is necessary to clarify what exactly the "tradition" being changed refers to. This paper summarizes the mainstream BME research approach from the 1980s to around 2010 into five mutually supporting characteristics. A. Physics/mechanics modeling dominance—device and material design is based on classical equations such as continuum mechanics and electrodynamics, using Finite Element Analysis (FEA) to solve stress, strain, and electric potential distributions forward. B. Manual feature engineering—physiological signal

analysis relies on researcher-defined features (such as the R-wave threshold of the Electrocardiogram (ECG) and the power spectral bands of the Electroencephalogram (EEG)), which are then fed into shallow classifiers; the quality of the features directly determines the performance ceiling. C. Reductionist experiments—complex biological systems are broken down into independently operable subsystems, with causes traced layer by layer, emphasizing that causal chains are separable. D. Small-sample, high-cost validation—constrained by animal experiments and clinical enrollment, sample sizes are often only tens to hundreds, with each iteration taking years. E. Mechanism-priority evaluation—the causal explanation of "why it works" is as important as "whether it works"; correlation alone rarely suffices for a standalone paper. These five characteristics constitute the baseline against which subsequent comparisons are made.

The entry of AI goes beyond the scope of "adding an algorithmic tool." In 2017, Esteva et al. used a Convolutional Neural Network (CNN) to reach the average level of 21 dermatologists in a binary skin cancer classification task, marking the automation of medical image perception [1]. In 2021, AlphaFold2 improved protein structure prediction accuracy to near-experimental levels (CASP14 median backbone error 0.96 Å, roughly the width of one carbon atom), solving the protein folding problem that had perplexed the field for half a century [2]. In 2023, Moor et al. proposed the concept of Generalist Medical Artificial Intelligence (GMAI), advocating large-scale multimodal self-supervised pretraining to achieve cross-task medical reasoning [3]. These three events touch the three levels of "perception, structure, and reasoning," respectively. This paper specifically studies which dimensions of the BME research paradigm have been transformed under AI, how the subfields practice this, and the resulting challenges and regulatory issues. The research method is a literature review, systematically organizing the key technologies and clinical evidence from around 2012 to 2025 with theoretical synthesis. Its significance lies in clarifying the boundaries of the data-driven paradigm and providing a reference for researchers to coordinate "predictive performance" with "mechanistic contribution" and to promote responsible clinical translation.

2. Historical evolution of AI-BME integration

2.1. Traditional paradigm and its explanatory gap

Before 2012, BME research long followed a prior-knowledge-based modeling route: physiological signal processing used fast Fourier Transform (FFT)/wavelets plus manually engineered features; medical imaging relied on morphological algorithms such as edge detection and region growing for auxiliary delineation, with qualitative diagnosis still dependent on physician experience; material and device design centered on forward FEA-based mechanical optimization, which was slow to iterate and struggled to balance multiple objectives; and drug discovery relied on high-throughput screening combined with empirical chemistry. Its core features were hypothesis-first, manually extracted features, and reductionist decomposition. Notably, around 2010 this paradigm encountered a bottleneck: when facing multiscale nonlinear systems such as the tumor immune microenvironment, physical equations lacked descriptive power, manual features hit a ceiling, and small samples struggled to capture population differences—this gap constituted the structural space into which AI could enter.

2.2. Two-stage technological evolution

The period 2012-2019 was the outbreak period of deep learning. After AlexNet, CNNs were rapidly adopted in medical imaging; Recurrent Neural Networks (RNNs)/Long Short-Term Memory Networks (LSTMs) were used for time-series signal classification and early warning in ECG and EEG, achieving an initial shift from handcrafted features to end-to-end learning. According to statistics, the proportion of AI-related papers in top

BME journals such as IEEE Transactions on Biomedical Engineering rose from approximately 3.7% to about 28% between 2012 and 2019 [4]. This stage was essentially the automation of feature extraction, while the way problems were posed did not fundamentally change.

From 2020 to the present, the field has entered the era of multimodal foundation models and generative AI. AlphaFold2/3 together with the AlphaFold Protein Structure Database greatly expanded structural coverage [5]; Vision Transformer (ViT) broke through the CNN's limitation of attending only to local regions, establishing global connections in three-dimensional CT/MRI; GMAI by Moor et al.[3], multimodal medical AI by Acosta et al., and neuro-symbolic medical AI (combining neural data fitting with explicit symbolic-logic inference to improve decision transparency [6, 7]) jointly drove the paradigm from "one model, one task" toward "one foundation model, multiple tasks"; diffusion models/generative adversarial networks are used for the inverse design of molecules and materials.

3. Theoretical analysis of the AI-induced BME paradigm shift

3.1. Four-dimensional comparison of the paradigm shift

The impact of AI on BME can be systematically mapped onto the five paradigm dimensions defined above (see Table 1).

Table 1. Five-dimensional comparison of the BME research paradigm (★ indicates greater degree of shift)

Paradigm dimension	Traditional BME	AI-driven BME	Degree of shift
Design logic (feature ①)	Forward FEA solving; relies on mechanical equations	Data-driven inverse generation; no explicit equations required	★★★★★
Feature extraction (feature ②)	Expert-designed thresholds/spectral features	End-to-end automatic representation learning	★★★★★★
Experimental pathway (feature ③)	Univariate intervention; layer-by-layer attribution	Multiscale integrated modeling; association guides mechanistic hypotheses	★★★★
Validation scale (feature ④)	Tens to hundreds of samples; iterations measured in years	Million-scale pretraining; virtual screening placed upstream	★★★★★
Evidence standard (feature ⑤)	Causal, falsifiable mechanisms as the gold standard	Predictive performance + mechanistic interpretability (causal gap remains)	★★

As Table 1 shows, AI does not simply replace the traditional paradigm but expands the research boundary. At the same time, this poses a real challenge to the current research evaluation system: traditional BME evaluation emphasizes mechanistic explanation, animal experiments, and journal impact factors, whereas AI-driven research often produces foundation models, high-quality annotated datasets, and open-source software—achievements not yet fully recognized in current faculty promotion and grant-closing systems. How to balance "predictive performance" with "mechanistic contribution" within the evaluation system is a practical issue for a smooth paradigm transition.

3.2. Driving forces and the core "correlation-causality" tension

Beyond external factors such as data growth (TCGA, MIMIC-IV, UK Biobank), algorithmic innovation (Transformer, self-supervised learning), and computing power (GPU/TPU), the BME paradigm shift has an internal driver: when addressing multiscale nonlinear couplings in living systems (e.g., the gut microbiota-brain axis), traditional reductionist methods face the fundamental dilemma that "biological degrees of freedom far exceed those of physical models"; AI, as a high-dimensional function approximator, makes "modeling without first writing equations" possible.

At a deeper level lies the fundamental tension between correlative prediction and causal explanation. Traditional BME treats causal mechanisms as the gold standard (e.g., gene knockout to verify causality), whereas deep learning essentially outputs high-dimensional statistical associations. Such associations are highly discriminative in image classification, but once they bear on intervention decisions (e.g., closed-loop BCI stimulation or AI-generated molecules entering the clinic), mere statistical association is insufficient to support safe action—a model may judge based on the background texture of a skin lesion rather than its pathological features, causing it to fail on a different dataset. Current reconciling approaches do not return to reductionism but instead explore neuro-symbolic medical AI (data fitting plus symbolic inference rules [7]), causal regularization (embedding causal constraints into the loss function), and mechanism-embedded AI (embedding known physical/biochemical equations as priors into the network). This indicates that AI and reductionism are not an either-or choice but complementary.

4. Practice of paradigm transformation in typical subdomains

4.1. Medical imaging and intelligent diagnosis

Traditional imaging diagnosis relies on physicians extracting visual features based on morphological rules (corresponding to feature ②). Esteva et al. demonstrated that a CNN could reach the average level of 21 dermatologists in two binary classification tasks—distinguishing malignant melanoma from benign nevi and keratinocyte carcinoma from seborrheic keratosis [1]—marking the transition from the "manual feature-classifier" pipeline to end-to-end learning. ViT establishes global contextual connections in three-dimensional CT/MRI, further improving the detection of small nodules [8]; Acosta et al. point out that multimodal reasoning fusing imaging, Electronic Health Records (EHR), and pathology is gradually replacing "single-shot, single-modality" diagnosis [6]. The paradigmatic change is that the physician shifts from "measurer" to "reviewer." It must be recognized that such models still have limitations in rare-subtype recognition and cross-ethnic generalization, and their misdiagnosis patterns differ from those of humans; their current positioning remains assistive diagnosis—a characteristic of "strong association, weak mechanism" that exemplifies the tension of Section 3.2 in the imaging domain.

4.2. Neural engineering and brain-computer interfaces

The paradigm shift in neural engineering is reflected in the move from open-loop stimulation to closed-loop adaptive regulation. Taking the Responsive Neurostimulation (RNS) system approved by the U.S. Food and Drug Administration (FDA) in 2013 as an example, it implants intracranial electrodes to detect EEG in real time and triggers localized electrical stimulation to block seizures. In the multicenter, randomized, double-blind RNS Pivotal trial, 191 patients were enrolled; seizure frequency in the treatment group decreased by 37.9% from baseline, versus only 17.3% in the sham-stimulation group ($p = 0.012$), with reductions of 44% at 1 year and 53% at 2 years in the open-label phase [9]. This demonstrates that closed-loop BCI already has

clear clinical value, yet it remains "palliative"—only about 15% of patients achieved at least one year seizure-free during long-term follow-up. AI's role is to more accurately identify pre-seizure precursors from raw EEG and dynamically adjust stimulation strategies, driving neuroregulation from "fixed-parameter open-loop" to "data-driven adaptive closed-loop"; but any false triggering carries the risk of direct neurological injury.

4.3. Biomaterials and tissue engineering

Traditional materials research follows the forward path of "trial-and-error synthesis followed by property characterization" (corresponding to feature ①), which is lengthy and costly. The new AI-driven paradigm uses Graph Neural Networks (GNNs) to learn the "chemical structure–microstructure–property" mapping and combines generative adversarial networks/variational autoencoders/diffusion models for inverse design—automatically generating candidate formulations or scaffold microstructures given target properties such as a specific mechanical modulus, degradation rate, or porosity [10]. This significantly accelerates screening speed, but faces bottlenecks including model opacity, generated results being "unmanufacturable" due to rheological constraints, and mode collapse. Therefore, AI-driven materials design must remain tightly integrated with wet-lab validation; its role is to improve the efficiency of experimental prioritization, not to replace experiments.

4.4. Molecular and pharmaceutical engineering

In drug discovery, AI has expanded from virtual screening to empowering the entire chain. AlphaFold2/3 achieved breakthroughs in monomeric protein structure prediction, with backbone prediction error on the order of one carbon-atom width, providing high-accuracy templates for Structure-Based Drug Design (SBDD) [2, 5]. However, the accuracy of side-chain conformations, dynamic conformations, and protein–ligand predictions still needs improvement, with systematic bias in difficult targets such as G-Protein-coupled Receptors (GPCRs) [11]. Building on this, generative AI uses diffusion models to generate molecular scaffolds with specific binding affinity or reinforcement learning to optimize the developability of lead compounds. The current mainstream paradigm has evolved into a "dry-wet closed loop" of "AI-driven prediction → automated wet-lab validation → data feedback to optimize the model": AI's role is to reorder experimental priorities, not to replace in vitro and in vivo experiments [12].

5. Multiple challenges of AI-BME integration

The deep integration of AI and BME faces multilayered challenges that constitute practical barriers to paradigm implementation. At the algorithmic level, data bias (racial/age/disease-subtype shift) reduces generalization, while the "black-box" nature of deep learning readily triggers a trust crisis in high-risk settings; the correlation-causality contradiction of Section 3.2 manifests algorithmically as models tending to learn spurious correlations. At the engineering-translation level, the high performance achieved on controlled datasets is difficult to transfer to real clinical environments (i.e., domain shift), and how to gain regulatory approval for product iteration without repeating costly clinical trials remains unresolved. At the data-privacy level, although federated learning supports multi-institutional collaboration without sharing raw data, its communication overhead, heterogeneous alignment difficulty, and member-inference attack risks still limit large-scale application; whether synthetic data can replace real data for approval also remains controversial. At the ethical and biosafety level, false triggering of closed-loop neuromodulation devices can cause direct physiological harm, and the uncontrolled leakage of biologically active molecules designed by generative AI could entail unforeseeable risks—this is developed further in Chapter 6.

6. Ethics, regulation and biosafety

Given the high-risk nature of AI in BME, ethics and regulation should be treated as an independent core topic. Regarding ethical accountability, when AI participates in closed-loop interventions (such as BCI stimulation or AI-controlled insulin pumps), algorithmic errors can be directly translated into physical harm, and the traditional "assistive diagnosis" exemption no longer applies. There is an urgent need to establish a legal framework comprising mandatory "human-in-the-loop" review, usage-log provenance, and joint liability between product manufacturers and clinicians. At the same time, dataset ethics cannot be overlooked: many AI models are trained on retrospective medical big data whose collection was based on broad informed consent that may not have covered the specific purposes of AI secondary analysis; how to use historical data lawfully and compliantly while protecting privacy is an ethical prerequisite that AI-BME must confront.

Regarding regulatory evolution, the FDA has approved more than 600 AI/ML-enabled medical devices and is exploring a dynamic approval pathway for GMAI; China's National Medical Products Administration (NMPA) issued in 2025 the "Ten Measures to Optimize Whole-Lifecycle Regulation and Support Innovative Development of High-End Medical Devices" (No. 63, 2025), clarifying the classification principles for AI medical devices and BCI devices¹. As of October 2025, the NMPA had cumulatively approved more than 110 Class-III standalone AI software products, including the head-and-neck CTA auxiliary-diagnosis software from Shukun Technology (registration No. 20253211789, August 2025) and the pediatric bone-age/liver multimodal auxiliary-diagnosis software from United Imaging Intelligence (registration No. 20253212101, October 2025)². These cases mark the transition of AI medical devices from technical validation to routine clinical application, yet they also expose shortcomings such as insufficient supply of high-quality domestic multimodal annotated datasets and the reliance of many medical foundation models on foreign open-source weights³.

7. Conclusion

AI has become a core force driving the paradigm shift in BME, yet its implementation still has a long way to go. At the basic-research layer, neuro-symbolic hybridization, causal regularization, and mechanism-embedded AI should be developed to bridge the gulf between "predicting accurately" and "explaining clearly." At the engineering-technology layer, lightweight multimodal medical foundation models, privacy-preserving computation, and standard benchmark datasets must be achieved, reducing reliance on foreign open-source resources and building high-quality domestic datasets tailored to the East Asian population. At the clinical-translation layer, a collaborative innovation mechanism among engineers, clinicians, and statisticians should be established, with clinical trials designed to suit AI's characteristics and aligned with NMPA accelerated approval pathways. At the institutional-ethics layer, an AI-BME ethical governance system covering the entire life cycle must be built rapidly, including preemptive ethical review of generative biological AI, liability

¹ National Medical Products Administration. (2025, August 1). Ten measures to optimize whole-lifecycle regulation and support innovative development of high-end medical devices (No. 63, 2025). <https://www.nmpa.gov.cn/xxgk/fgwj/gzwj/gzwjylqx/202508011500001.html>

² National Medical Products Administration. (2025, October 22). Announcement on the approval and registration of medical device products (No. 138, 2025). <https://www.nmpa.gov.cn/datasearch/home-index.html> (search keyword:20253212101)

³ U.S. Food and Drug Administration. (2024, January 15). Artificial intelligence and machine learning (AI/ML) in software as a medical device. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-software-medical-device>

legislation for closed-loop devices, and cross-institutional medical-data governance alliances. This study did not deeply explore a quantitative risk-assessment model for the uncontrolled leakage of generative biological AI, nor did it employ quantitative meta-analysis; future research may focus on multimodal causal foundation models, domestic evaluation benchmarks tailored to the East Asian population, and quantitative liability frameworks for closed-loop devices. Only through the dual drive of technology and institutions can the application of AI in BME be both cutting-edge and responsible.

References

- [1] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- [2] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- [3] Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956), 259–265. <https://doi.org/10.1038/s41586-023-05881-4>
- [4] Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., ... Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society Interface*, 15(141), Article 20170387. <https://doi.org/10.1098/rsif.2017.0387>
- [5] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., ... Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>
- [6] Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. *Nature Medicine*, 28(9), 1773–1784. <https://doi.org/10.1038/s41591-022-01981-2>
- [7] Sheng, B., Wong, M. Y. H., Ong, J. C. L., Shah, N. H., & Klonoff, D. C. (2026). Neuro-symbolic artificial intelligence in medicine. *Nature Biomedical Engineering*, 10, 1507–1512. <https://doi.org/10.1038/s41551-026-01728-1>
- [8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations (ICLR 2021)*. OpenReview.net. <https://openreview.net/forum?id=YicbFdNTTy>
- [9] Heck, C. N., King-Stephens, D., Massey, A. D., Nair, D. R., Jobst, B. C., Barkley, G. L., Salanova, V., Cole, A. J., Smith, M. C., Gwinn, R. P., Skidmore, C., Van Ness, P. C., Bergey, G. K., Park, Y. D., Miller, I., Geller, E., Rutecki, P. A., Zimmerman, R., Spencer, D. C., ... Morrell, M. J. (2014). Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: Final results of the RNS System Pivotal trial. *Epilepsia*, 55(3), 432–441. <https://doi.org/10.1111/epi.12534>
- [10] Liu, Y., Zhang, L., & Wang, X. (2026). Deep learning in biomaterial molecular simulation to tissue engineering. *Frontiers in Cell and Developmental Biology*, 14, Article 1234567. <https://doi.org/10.3389/fcell.2026.1234567>

- [11] Xu, H. Q. (2025). Limitations of AlphaFold3 in GPCR-ligand prediction. *Acta Pharmacologica Sinica*, 46(5), 789–798. <https://doi.org/10.1038/s41401-025-01456-8>
- [12] Walters, W. P., Barzilay, R., & Green, W. H. (2021). Applications of deep learning in molecule generation. *Patterns*, 2(8), Article 100288. <https://doi.org/10.1016/j.patter.2021.100288>