

Causal Machine Learning for Personalized Recommendation of Bioactive Peptides in Aging-Related Disease Prevention

Federigo Salgari

Department of Computer Science, Colorado State University, Fort Collins, CO, USA.

federigo832@colostate.edu

Jatin Devi

Department of Computer Science, University of North Texas, Denton, TX, USA.

jdevi@unt.edu

Kathi Wahner

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.

contactkathi@ku.edu

Jabal Utama

School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.

utama956@uc.edu

Abstract

Aging-related diseases represent a convergent burden shaped by biological aging, chronic inflammation, metabolic dysregulation, microbiome shifts, and cumulative environmental exposures. Bioactive peptides derived from food and animal sources have attracted attention because they may modulate oxidative stress, immune signaling, metabolic pathways, and tissue repair, yet their effects are heterogeneous and context dependent. Most existing recommendation systems for nutrition and preventive interventions rely on associational learning, which limits their capacity to distinguish causal effects from confounding, selection bias, and reverse causation. This paper examines causal machine learning as a system-level architecture for personalized recommendation of bioactive peptides in aging-related disease prevention. It discusses how causal representation learning, heterogeneous treatment effect estimation, reinforcement learning, and network medicine can be integrated with multimodal data infrastructures spanning clinical records, omics, microbiome profiling, dietary behavior, wearables, and product composition databases. The analysis emphasizes structural trade-offs involving personalization, interpretability, robustness, fairness, privacy, and sustainability. It also considers governance requirements for deployment, including validation, monitoring, model updating, consent, equitable access, and regulatory oversight. The central argument is that causal machine learning can improve peptide recommendation only if it is embedded within a socio-technical system that treats evidence generation, data stewardship, and translational accountability as first-order design concerns. Such a system must remain modest about uncertain effects while enabling safe, adaptive, and equitable preventive pathways.

Keywords

causal machine learning; bioactive peptides; personalized nutrition; aging-related diseases; recommender systems; fairness; governance; translational medicine.

1. Introduction

Aging-related diseases are not isolated outcomes of single genetic defects or lifestyle choices. They emerge from interacting processes that include cellular senescence, mitochondrial dysfunction, chronic low-grade inflammation, proteostasis loss, metabolic inflexibility, and altered immune surveillance [10]. Geroscience has reframed aging itself as a shared upstream risk condition for cardiovascular disease, neurodegeneration, metabolic disorders, musculoskeletal decline, and frailty [11]. Precision medicine has similarly argued that prevention should move from population averages toward stratified and eventually individualized interventions informed by molecular, behavioral, and environmental data [6]. Systems medicine extends this logic by treating health as a dynamic network property in which diet, gut microbiota, immune regulation, and metabolic signaling continually reshape disease trajectories [7]. In this context, bioactive peptides have become a plausible component of preventive nutrition because they can be produced from food proteins and may act on multiple physiological targets, including antioxidant defense, angiotensin-converting enzyme activity, immune modulation, and satiety signaling [12]. However, the evidence base remains uneven, with many studies focusing on biomarkers or short-term outcomes rather than long-term disease prevention.

The central technical challenge is that personalized recommendation is fundamentally a causal problem. A system that observes that individuals who consume certain peptides have lower disease incidence cannot conclude that the peptides caused the reduction. Such associations may reflect healthier diets, higher socioeconomic status, reverse causation, or unmeasured lifestyle factors. Causal inference provides formal tools for distinguishing intervention effects from observational patterns [1]. Causal representation learning further seeks to recover latent variables and mechanisms that remain stable across environments, which is important when dietary, microbiome, and clinical data are heterogeneous [2]. Heterogeneous treatment effect estimation allows models to ask not only whether a peptide works on average but for whom, under what conditions, and with what expected benefit or harm [3]. Reinforcement learning in healthcare offers sequential decision frameworks for adapting recommendations over time, although it also introduces safety and interpretability risks that require careful constraints [4]. These methods do not eliminate uncertainty; they make uncertainty more explicit and therefore more governable.

This paper develops a system-level account of causal machine learning for personalized recommendation of bioactive peptides in aging-related disease prevention. It does not propose a single algorithm or a fixed pipeline. Instead, it analyzes the architectural, infrastructural, ethical, and policy conditions under which such a system could be credible, deployable, and sustainable. The discussion is organized around several tensions. First, personalization requires rich data, but rich data intensify privacy and equity risks. Second, causal models can improve decision support, but they depend on assumptions that are often untestable in observational settings. Third, recommendation systems can scale preventive advice, but scalability may amplify bias or commercial pressure. Fourth, long-term aging outcomes require longitudinal evidence, yet deployment often proceeds under short-term metrics. These tensions are not peripheral; they define the feasibility of causal peptide recommendation as a socio-technical infrastructure.

2. Aging-Related Disease Prevention and the Peptide Evidence Base

Aging-related disease prevention operates across timescales that are difficult for conventional clinical trials and consumer technologies to capture. The hallmarks of aging include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, disabled macroautophagy, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication, chronic inflammation, dysbiosis, and dysregulated nutrient sensing [10]. These processes interact, so an intervention that alters one pathway may produce secondary effects elsewhere. Geroscience suggests that targeting fundamental aging mechanisms could delay multiple chronic conditions simultaneously, but it also warns against assuming that any single dietary compound will produce uniform benefits [11]. Network medicine reinforces this caution by showing that diseases are connected through shared molecular and environmental modules rather than discrete boundaries [18]. A peptide recommendation system therefore cannot treat aging-related disease as a simple label to be predicted; it must model a dynamic, multi-morbid, and partially observable system.

Bioactive peptides are short amino acid sequences that can be released from parent proteins through enzymatic hydrolysis, fermentation, gastrointestinal digestion, or food processing [15]. Their activities depend on sequence length, amino acid composition, charge, hydrophobicity, resistance to digestion, absorption, and interactions with host receptors or microbiota. Milk protein-derived peptides have been studied for blood pressure regulation, immune modulation, and metabolic effects [14]. Marine-derived peptides have shown antioxidant, antihypertensive, and anti-inflammatory potential [16]. Food-derived peptides with biological activity have been reviewed extensively, with attention to their production, bioavailability, and translation into functional foods [17]. These reviews are valuable, but they also reveal a persistent gap between in vitro activity, animal models, and human preventive outcomes. A recent narrative review of animal-derived peptides and aging-related diseases synthesizes evidence that such peptides may influence oxidative stress, inflammation, and metabolic dysfunction, while also emphasizing the need for stronger human data [13]. This evidence base supports cautious optimism, not deterministic recommendation.

The gut microbiome is a critical mediator and modifier of peptide effects. Gut health is not a single biomarker but a functional state involving mucosal immunity, barrier integrity, microbial metabolism, and resistance to pathogens [8]. The microbiota changes across age, diet, environment, medication use, and disease, making it a moving target for personalization [9]. Peptides may be metabolized by microbial enzymes, alter microbial composition, or interact with immune receptors in the gut. Conversely, microbiome composition may determine whether a peptide reaches systemic circulation or is degraded before producing a physiological effect. This creates a reciprocal causal structure that associational recommender systems struggle to capture. A system that recommends a peptide based on average effects may overlook that the same peptide could be beneficial for one microbiome state and ineffective or even detrimental for another. Causal machine learning offers a vocabulary for modeling such effect modification, but only if the data infrastructure can measure the relevant mediators and moderators over time.

3. Causal Machine Learning Architecture for Personalized Recommendation

A causal machine learning architecture for peptide recommendation must begin with a clear target of inference. The system may aim to estimate the effect of a specific peptide, dose, duration, and delivery format on a defined outcome such as inflammatory burden, glycemic variability, cognitive decline, or frailty progression. Structural causal models provide a way to represent assumptions about which variables confound, mediate, or modify the effect [1].

Causal representation learning can help compress high-dimensional molecular and behavioral data into latent factors that preserve causal structure across settings [2]. This is particularly relevant when integrating proteomic, metabolomic, metagenomic, and clinical variables that are noisy, sparse, and batch dependent. Without causal representation, a model may learn shortcuts tied to a particular cohort, sequencing platform, or dietary questionnaire, leading to recommendations that fail when deployed in another population.

Heterogeneous treatment effect estimation is the core predictive task. Instead of estimating a single average effect, the system estimates conditional effects that vary with individual characteristics, environmental context, and time. Metalearners and related approaches allow flexible machine learning models to estimate such effects while preserving some interpretability at the subgroup level [3]. In practice, the system might combine randomized trial data, observational cohorts, electronic health records, and n-of-one experiments. Randomized evidence provides stronger identification but is often limited in duration and population diversity. Observational data provide scale but require careful adjustment for confounding, selection, and measurement error. The most credible architecture is therefore pluralistic: it does not privilege one data source but explicitly models the limitations of each and updates recommendations as evidence accumulates. This aligns with precision medicine's emphasis on evidence stratification rather than one-time classification [6].

Sequential recommendation adds another layer. Aging-related prevention is not a single decision but a series of decisions about diet, supplementation, monitoring, and adherence. Reinforcement learning can model these sequential choices and adapt recommendations based on observed responses [4]. However, healthcare reinforcement learning faces well-known challenges, including sparse rewards, delayed outcomes, safety constraints, and off-policy evaluation difficulties. A peptide recommender should not autonomously optimize for engagement or short-term biomarker changes. It should operate as a constrained decision-support system in which clinicians, nutritionists, and users retain authority. The system can propose candidate interventions, estimate expected benefits and harms, and explain uncertainty, but it should not collapse prevention into automated prescription. This distinction matters because aging-related prevention involves long horizons and non-reversible risks, making unconstrained exploration ethically problematic.

Network medicine offers a complementary architecture. Rather than treating peptides as isolated agents, the system can map them onto molecular pathways, protein interactions, and disease modules [18]. A peptide may affect multiple pathways, and a disease may be driven by multiple modules. Network-based reasoning can help identify pleiotropic effects, potential interactions with medications, and biomarkers that mediate response. It can also reveal redundancy: if two peptides target overlapping pathways, recommending both may not produce additive benefit. At the system level, this suggests a graph-structured recommendation layer that combines causal effect estimates with mechanistic plausibility. Mechanistic knowledge does not replace causal inference, but it can inform prior assumptions, guide feature selection, and improve interpretability. A purely data-driven model may find correlations that are statistically strong but biologically implausible; a purely mechanistic model may be too rigid for heterogeneous individuals. The architecture must balance both.

4. Data Infrastructure, Deployment, and Lifecycle Governance

The data infrastructure for causal peptide recommendation must support collection, curation, integration, and updating across clinical, molecular, behavioral, and commercial domains. Clinical data include diagnoses, medications, laboratory values, and outcomes. Molecular data

include genomics, proteomics, metabolomics, and inflammatory markers. Microbiome data include taxonomic and functional profiles that change over time [9]. Behavioral data include dietary intake, physical activity, sleep, and adherence. Product data include peptide sequence, source protein, production method, dose, bioavailability, and batch variability [15]. These data are generated by different institutions, governed by different consent regimes, and measured with different error structures. Without shared standards, the system will produce recommendations that are precise only within narrow data silos.

Deployment should be staged rather than monolithic. A closed-loop clinical trial environment can test the system under careful oversight, with predefined endpoints and stopping rules. A clinician-supervised pilot can evaluate workflow integration, user comprehension, and equity of access. A consumer-facing preventive setting may be considered only after safety, interpretability, and regulatory pathways are mature. Each stage requires different evidence thresholds. In early stages, the system may recommend only well-characterized peptides with favorable safety profiles and robust biomarker evidence. In later stages, it may support more personalized combinations, but only if monitoring can detect adverse effects and if users can withdraw without penalty. The governance challenge is to prevent premature scaling from outrunning evidence. Precision medicine has shown that technological capability alone does not guarantee clinical utility [6]. Systems medicine has shown that interventions must be evaluated within the whole person and environment, not only at the molecular target [7].

Lifecycle governance must address model drift, evidence drift, and product drift. Model drift occurs when population characteristics, data collection practices, or disease patterns change. Evidence drift occurs when new trials or meta-analyses revise the estimated effects of a peptide. Product drift occurs when manufacturing, formulation, or sourcing changes the actual exposure. A causal recommender must therefore maintain versioned models, transparent change logs, and continuous monitoring. It should also distinguish between stable causal mechanisms and context-dependent effect sizes. A peptide that reduces inflammation in one population may have little effect in another because of baseline risk, co-medication, or microbiome differences. Governance should require periodic re-evaluation rather than assuming that initial validation remains valid indefinitely. This is especially important in aging-related prevention, where the target population may change as earlier interventions succeed.

Interoperability is a structural prerequisite. If a hospital cannot exchange data with a nutrition platform, or if a wearable cannot export contextual data into a research pipeline, the causal model remains incomplete. Interoperability also has ethical dimensions. Data minimization, purpose limitation, and consent granularity must be designed into the infrastructure. Users should be able to understand what data are used, why a recommendation is made, and what alternatives exist. Explainability in causal systems is not merely a feature; it is a condition for trust and accountability. A recommendation that cannot be explained cannot be contested. A recommendation that cannot be contested is difficult to govern in clinical, regulatory, or consumer settings.

5. Robustness, Fairness, and Policy Implications

Robustness in causal peptide recommendation requires attention to confounding, measurement error, missing data, and distribution shift. Observational data are particularly vulnerable to unmeasured confounding. For example, individuals who consume bioactive peptides may also have higher health literacy, better access to care, or different dietary patterns. Causal methods can adjust for measured confounders, but they cannot eliminate

unmeasured ones without strong assumptions [1]. Sensitivity analysis, negative controls, and triangulation across data sources can help, but they do not produce certainty. A responsible system should communicate uncertainty in a way that non-specialists can understand. It should avoid presenting probabilistic estimates as definitive prescriptions. It should also allow users and clinicians to override recommendations without penalty.

Fairness is a central concern because personalized nutrition intersects with socioeconomic inequality. Algorithmic bias can arise from historical inequities, proxy variables, differential access, and measurement bias [19]. If the training data overrepresent affluent, well-monitored populations, the model may perform poorly for underrepresented groups. If the recommendation system is tied to expensive supplements or wearables, it may worsen health disparities. If the system uses race or ethnicity as a proxy for biological difference, it may perpetuate harmful essentialism. Fairness therefore cannot be reduced to a single metric. It requires participatory design, subgroup evaluation, and attention to allocative and representational harms. The ethical implementation of machine learning in healthcare demands similar commitments to transparency, accountability, and patient welfare [20]. These are not optional additions; they are constitutive of legitimate deployment.

Policy implications extend beyond model performance. Regulatory agencies must determine whether a peptide recommendation system is a medical device, a wellness tool, a dietary supplement platform, or a hybrid. The distinction affects evidentiary standards, advertising rules, liability, and reimbursement. Clinicians need guidance on how to interpret causal estimates and how to document shared decision-making. Consumers need protection from exaggerated claims and dark patterns that exploit health anxiety. Researchers need access to interoperable data without compromising privacy. Industry actors need incentives to share product composition and safety data. Policymakers need to consider public funding for independent validation, especially for peptides that cannot be patented and therefore lack strong commercial incentives for large trials. Without public infrastructure, the evidence base may remain fragmented and commercially skewed.

International harmonization is difficult but necessary. Aging-related diseases are global, and peptide sources, diets, and regulatory frameworks vary across regions. A recommendation that is appropriate in one country may be inappropriate in another because of dietary patterns, medication use, or microbiome differences. Cross-border data flows raise privacy and sovereignty issues. At the same time, isolated national systems risk duplicating costs and missing opportunities for validation. A federated approach, in which models are trained across institutions without moving raw data, could help balance privacy and scale. However, federated learning introduces its own governance challenges, including secure aggregation, auditability, and equitable benefit sharing. The policy goal should be a learning health system that is distributed but accountable.

6. Sustainability and Translational Pathways

Sustainability in this context has at least three meanings. The first is environmental: peptide production, packaging, and distribution have ecological footprints. Animal-derived peptides may raise concerns about animal welfare, land use, and emissions, while plant and marine sources raise different sustainability questions [16]. The second is economic: personalized recommendation systems require ongoing data collection, model maintenance, clinical oversight, and user support. If costs are borne by individuals, access will be unequal. If costs are borne by public health systems, budget constraints may limit scale. If costs are borne by commercial platforms, incentives may favor engagement and product sales over health

outcomes. The third is epistemic: the system must remain sustainable as a knowledge enterprise. It must not depend on a single dataset, model, or vendor. It must preserve the ability to revise recommendations when evidence changes.

Translational pathways should be incremental and evidence-gated. A reasonable sequence begins with retrospective causal analyses of existing cohorts to identify promising peptide-outcome relationships and effect modifiers. It then moves to prospective observational studies with prespecified causal questions and rigorous measurement. Randomized trials, including pragmatic and adaptive designs, can test whether recommendation improves outcomes compared with usual care or generic advice. N-of-one trials may be useful for individual-level inference, especially for subjective outcomes such as fatigue or well-being, but they require careful statistical design and cannot replace population-level evidence. Implementation science is essential because even a valid recommendation may fail if users cannot understand it, clinicians cannot integrate it, or health systems cannot pay for it. Sustainability therefore depends on alignment among evidence, workflow, reimbursement, and trust.

A key translational tension is between generalization and personalization. A model that is highly personalized may be difficult to validate because each individual has a unique treatment effect. A model that is too general may ignore important heterogeneity. Causal machine learning can help by estimating effect distributions and identifying subgroups with credible differences. However, subgroup discovery must be preregistered or at least validated independently to avoid false discoveries. The system should distinguish between confirmatory subgroup effects and exploratory hypotheses. It should also avoid using personalization as a marketing term for what is essentially average advice. True personalization requires evidence that the recommendation changes based on individual characteristics in a way that improves outcomes. Without that evidence, personalization is only segmentation.

7. Conclusion

Causal machine learning offers a rigorous framework for personalizing bioactive peptide recommendations in aging-related disease prevention, but it is not a standalone solution. The scientific evidence for peptide effects remains incomplete, heterogeneous, and context dependent. The data required for personalization are sensitive, fragmented, and costly to maintain. The deployment environment is shaped by commercial incentives, regulatory uncertainty, and health inequities. A credible system must therefore be designed as a socio-technical infrastructure rather than a predictive model. It must combine causal inference with mechanistic knowledge, multimodal data with privacy safeguards, adaptive recommendation with human oversight, and personalization with fairness. It must be governed across the lifecycle, with transparent assumptions, continuous monitoring, and the ability to revise or withdraw recommendations as evidence changes.

The most defensible near-term role for causal peptide recommendation is decision support, not automated prescription. Such a system can help researchers identify promising hypotheses, help clinicians discuss uncertain options, and help users understand why a recommendation is made and what remains unknown. It can also help prioritize trials by estimating where effects are most likely to be clinically meaningful. Over time, if evidence and governance mature, the system could support more adaptive preventive strategies. That trajectory depends less on algorithmic novelty than on institutional commitments to validation, equity, and accountability. The promise of causal machine learning in this domain is not that it will eliminate uncertainty. It is that it can make uncertainty visible, testable, and actionable within a system designed to protect human welfare.

References

1. Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
2. Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, 109(5), 612-634.
3. Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10), 4156-4165.
4. Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature Medicine*, 25(1), 16-18.
5. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56.
6. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372(9), 793-795.
7. Hood, L., & Flores, M. (2012). A personal view on systems medicine and the emergence of proactive P4 medicine: Predictive, preventive, personalized and participatory. *New Biotechnology*, 29(6), 613-624.
8. Bischoff, S. C. (2011). Gut health: A new objective in medicine? *BMC Medicine*, 9, 24.
9. Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., & Mele, M. C. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms*, 7(1), 14.
10. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153(6), 1194-1217.
11. Kennedy, B. K., Berger, S. L., Brunet, A., Campisi, J., Cuervo, A. M., Epel, E. S., Franceschi, C., Lithgow, G. J., Morimoto, R. I., Pessin, J. E., Rando, T. A., Richardson, A., Schadt, E. E., Wyss-Coray, T., & Sierra, F. (2014). Geroscience: Linking aging to chronic disease. *Cell*, 159(4), 709-713.
12. Chakrabarti, S., Guha, S., & Majumder, K. (2018). Food-derived bioactive peptides in human health: Challenges and opportunities. *Nutrients*, 10(11), 1738.
13. Liu, T., Song, X., Dai, X., Ma, H., Muhammad, A., He, Z., & Zhao, K. (2025). Effects of animal-derived peptides on aging-related diseases: a narrative review. *Aging Advances*, 2(4), 160-165.
14. Nongonierma, A. B., & FitzGerald, R. J. (2015). The scientific evidence for the role of milk protein-derived bioactive peptides in humans: A review. *Journal of Functional Foods*, 17, 640-656.
15. Udenigwe, C. C., & Aluko, R. E. (2012). Food protein-derived bioactive peptides: Production, processing, and potential health benefits. *Journal of Food Science*, 77(1), R11-R24.

16. Kim, S. K., & Wijesekara, I. (2010). Development and biological activities of marine-derived bioactive peptides: A review. *Journal of Functional Foods*, 2(1), 1-9.
17. Hartmann, R., & Meisel, H. (2007). Food-derived peptides with biological activity: From research to food applications. *Current Opinion in Biotechnology*, 18(2), 163-169.
18. Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews Genetics*, 12(1), 56-68.
19. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
20. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care: Addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981-983.