Multimodal AI for Personalized Drug Discovery Through Joint Protein Structure Generation and Compound Interaction Prediction

Pangestu Waluyo
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
waluyo1971@ucf.edu

Sean-Paul Concannon
Department of Computer Science, George Mason University, Fairfax, VA, USA.
seanpaul.work@gmu.edu

Amin van Westfalen
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
aminvanwestfalen87@colostate.edu

Ottmar Roht
Department of Computer Science, University of North Texas, Denton, TX, USA.
ottmar353@unt.edu

Abstract

Multimodal artificial intelligence is transforming drug discovery from a sequence of disconnected screening tasks into an integrated socio-technical system that couples molecular generation, structural biology, interaction prediction, and clinical decision support. This paper examines the architecture and governance of personalized drug discovery platforms that jointly generate protein structures and predict compound interactions. Such systems combine protein language models, geometric reasoning, generative molecular design, and multimodal representation learning to propose candidate compounds conditioned on patient-specific biological context. The central claim is that joint generation and prediction can improve therapeutic relevance, but only when the platform is designed as a governed infrastructure rather than a single predictive model. We analyze the structural trade-offs among model scale, interpretability, latency, privacy, and clinical accountability. We then discuss data provenance, federated learning, model cards, datasheets, and regulatory evidence generation as necessary complements to algorithmic performance. Particular attention is given to robustness under distribution shift, fairness across ancestral and socioeconomic groups, environmental sustainability of large-scale training and inference, and policy mechanisms for post-market surveillance. The paper argues that personalized drug discovery requires hybrid evaluation that combines computational benchmarks, wet-lab validation, and real-world clinical feedback. It also identifies failure modes that emerge when generative models optimize for plausible structures or molecules without sufficient interaction-level grounding. The conclusion is that multimodal AI can support more precise and equitable therapeutic discovery if architecture, infrastructure, evaluation, and governance co-evolve. The proposed perspective is system-level and forward-looking, emphasizing deployability, accountability, and long-term maintenance over narrow benchmark gains.

Keywords

multimodal artificial intelligence; personalized drug discovery; protein structure generation; compound interaction prediction; biomedical infrastructure; clinical translation; governance; fairness; sustainability; robust machine learning.

1. Introduction

Personalized drug discovery has become one of the most demanding test cases for contemporary artificial intelligence because it requires the integration of heterogeneous biological signals, uncertain chemical design spaces, and high-stakes clinical decision-making. Traditional pipelines separate target identification, protein structure determination, virtual screening, lead optimization, and clinical validation into distinct organizational and technical stages. Recent advances in protein structure prediction, molecular generation, and representation learning have blurred these boundaries, making it possible to imagine a unified system that jointly proposes candidate proteins or binding interfaces and predicts how small molecules or biologics will interact with them [1], [3], [4], [5]. The promise is not merely faster screening but a more personalized form of discovery in which therapeutic hypotheses are conditioned on patient-specific genotype, phenotype, and microenvironment. At the same time, the integration of generative modeling with interaction prediction creates new system-level risks that cannot be addressed by model accuracy alone. A platform that can generate plausible protein structures and compound candidates may also generate confident but biologically meaningless outputs if its data, governance, and evaluation infrastructure are weak [12], [14].

The title of this paper foregrounds two technical operations that are often treated separately: joint protein structure generation and compound interaction prediction. Protein structure generation concerns the production of three-dimensional conformations or design proposals that satisfy biophysical constraints and functional objectives. Compound interaction prediction concerns the estimation of binding affinity, selectivity, toxicity, and downstream pharmacological behavior for candidate molecules. When these operations are coupled, the system can optimize compounds against generated or dynamically inferred target geometries rather than against a fixed crystal structure. This coupling is attractive for personalized medicine because many clinically relevant targets are polymorphic, conformationally flexible, or incompletely characterized. However, coupling also amplifies error propagation. A small bias in the generative structural prior can shift the entire chemical search toward compounds that bind an unrealistic conformation. Conversely, an interaction predictor trained on narrow chemical libraries may reward molecules that exploit dataset artifacts rather than biological mechanism. Therefore, the system-level question is not whether multimodal models can be combined, but how their interfaces, uncertainty estimates, and governance controls should be designed.

This paper argues that personalized drug discovery platforms should be understood as socio-technical infrastructures. They are composed of data pipelines, model registries, compute clusters, laboratory automation, clinical feedback loops, regulatory documentation, and organizational accountability structures. Their performance depends on long-term maintenance, versioning, monitoring, and domain adaptation as much as on initial training. The analysis is organized around architectural trade-offs, infrastructure requirements, robustness and fairness, sustainability, and policy implications. The goal is to provide a forward-looking account of how multimodal AI can support personalized drug discovery without reducing the problem to benchmark optimization. In doing so, the paper connects

algorithmic design to deployment realities that shape whether a discovery system can be trusted, audited, and sustained.

2. Conceptual Foundations and System Requirements

The conceptual foundation of joint protein generation and compound interaction prediction rests on representation learning across biological and chemical modalities. Protein language models demonstrate that large-scale sequence data contain structural and functional signals that can be extracted without explicit supervision [6]. Transformer architectures provide a general mechanism for modeling long-range dependencies across residues, atoms, and molecular substructures [2]. Geometric deep learning and attention-based protein models further show that three-dimensional context can be incorporated into predictive representations [8], [9], [10]. These developments suggest that a multimodal system can learn shared latent spaces in which protein conformations, ligand graphs, and interaction outcomes are not isolated tasks but mutually informative views of the same biological problem. The system requirement is therefore not simply multimodality as concatenation, but multimodal alignment in which each modality improves the calibration and usefulness of the others.

A second requirement is conditional generation. Personalized drug discovery implies that outputs should depend on patient-specific variables, including genetic variants, expression profiles, immune context, and prior treatment history. A generative model that produces a single optimal molecule for a target population is insufficient if the clinical objective is to match therapy to an individual or subgroup. Conditional generation introduces difficult questions about how patient data are represented, how missing data are handled, and how privacy is preserved. It also creates a tension between specificity and generalizability. A model that overfits to an individual patient may propose compounds that are clinically irrelevant beyond that case, while a model that ignores individual variation may reproduce population-level biases. The architecture must therefore support conditional control at multiple levels, from target conformation to pharmacological constraints and safety margins.

A third requirement is interaction grounding. Generative models of proteins and molecules can produce outputs that appear novel and plausible but are not necessarily synthesizable, stable, or efficacious. Interaction prediction serves as a grounding signal that connects generated structures and compounds to measurable biological outcomes. Recent work on unified multimodal Bayesian flow for all-atom protein design illustrates the direction of travel toward integrated generative and predictive reasoning over atomic detail [11]. The lesson is that generative capability and predictive discrimination must be trained and evaluated together, because separation encourages models to exploit modality-specific shortcuts. The system should therefore expose interaction predictions as part of the generation loop, not as a post hoc filter. This integration requires uncertainty-aware scoring, because a single point estimate of binding affinity is inadequate for deciding whether a generated candidate merits expensive experimental validation.

A fourth requirement is lifecycle governance. Biomedical discovery systems are not static software products; they operate in environments where pathogens evolve, standards change, new assays become available, and regulatory expectations shift. The platform must support data versioning, model retraining, provenance tracking, and rollback capabilities. Without these capabilities, a multimodal discovery system may accumulate hidden technical debt and become difficult to audit or reproduce [18]. Governance is not an external layer added after deployment. It is a design requirement that determines which data can be used, which outputs can be released, and which human experts must review decisions before they affect patients.

The conceptual foundation of personalized drug discovery is therefore both technical and institutional, combining representation learning with accountable infrastructure.

3. Joint Multimodal Architecture for Protein Generation and Compound Interaction Prediction

A joint multimodal architecture for personalized drug discovery can be understood as a layered system. The first layer consists of modality-specific encoders that transform protein sequences, structural coordinates, molecular graphs, assay results, and patient context into numerical representations. Protein encoders may combine sequence language models with geometric modules that respect rotational and translational invariance. Molecular encoders may process atom types, bonds, stereochemistry, and pharmacophore features. Clinical encoders may summarize genotypes, expression signatures, and prior outcomes. The second layer aligns these representations through cross-modal attention or shared latent spaces, allowing the system to condition compound generation on protein structure and patient context. The third layer performs generative and predictive functions, proposing candidate structures or molecules while estimating interaction outcomes such as binding affinity, selectivity, and toxicity risk. The fourth layer manages uncertainty, constraints, and human review before candidates enter experimental pipelines.

The main architectural trade-off is between unification and modularity. A fully unified model can learn rich cross-modal interactions and reduce interface errors, but it becomes difficult to interpret, debug, and update. A modular system is easier to validate component by component, yet it risks misalignment between the generative prior and the interaction predictor. In personalized discovery, modularity is often preferable at the boundaries because different modalities have different data governance requirements. For example, patient genotype data may be subject to strict privacy controls, while public protein structures and chemical libraries may be openly shared. A modular architecture allows privacy-preserving components to be isolated without preventing the system from learning cross-modal relationships. However, the interfaces between modules must be specified with uncertainty estimates, version identifiers, and semantic constraints so that a generated protein conformation cannot be silently treated as a confirmed experimental structure.

Generative design in this architecture can take several forms. Molecular generative models may propose compounds with desired properties, while protein generative models may propose binding pockets, scaffolds, or engineered variants. The required reference on rationalized all-atom protein design with unified multimodal Bayesian flow is relevant because it points toward generative processes that operate at atomic resolution while incorporating multiple information sources [11]. In a personalized system, such generative processes could be conditioned on patient-specific target variants or on resistance mutations that alter binding surfaces. The interaction predictor then evaluates whether proposed compounds are likely to engage the generated structure in a therapeutically useful way. This loop resembles inverse design, where the system searches for molecules that satisfy a desired interaction profile rather than screening a fixed library [5]. The difference is that the target itself may be uncertain or generated, which raises the need for conservative uncertainty handling.

Foundation models add both capacity and governance complexity to this architecture. Large pretrained models can transfer knowledge across tasks and reduce the need for labeled data in rare diseases or underrepresented populations [19]. They can also encode biases from pretraining corpora and produce outputs whose provenance is difficult to trace. A

personalized discovery platform should therefore maintain model registries that record pretraining data, fine-tuning procedures, evaluation results, and known limitations. The architecture should support constrained decoding or filtering so that generated molecules satisfy synthesizability, toxicity, and intellectual property constraints. It should also support human-in-the-loop review, because clinical translation requires expert judgment about mechanism, safety, and patient context. The joint architecture is not a replacement for laboratory and clinical expertise; it is a coordination mechanism that makes expert review more targeted and reproducible.

4. Data Infrastructure, Governance, and Personalized Deployment

Data infrastructure is the substrate on which multimodal drug discovery depends. Personalized systems require linked datasets spanning genomics, proteomics, structural biology, chemistry, pharmacology, and electronic health records. These datasets are often fragmented across institutions, governed by different consent agreements, and stored in incompatible formats. The FAIR principles for scientific data management provide a baseline for making data findable, accessible, interoperable, and reusable, but they do not resolve the ethical and legal complexities of patient-derived data [22]. A deployment-ready platform must therefore implement data contracts, access controls, provenance logs, and auditable consent management. It must also support federated or privacy-preserving learning so that models can improve without centralizing sensitive records. Privacy in the age of medical big data is not only a technical encryption problem; it is a governance problem involving reidentification risk, secondary use, and equity in data contribution [17].

Governance mechanisms such as datasheets for datasets and model cards for model reporting are especially important in this domain because they make assumptions visible to downstream users [20], [21]. A dataset used to train a compound interaction predictor should document the chemical diversity, assay conditions, target classes, and known gaps. A generative protein model should document the structural families, resolution ranges, and conformational states represented in its training data. Without such documentation, clinical and regulatory reviewers cannot assess whether a model is appropriate for a new patient population or disease context. Governance also extends to deployment. A personalized discovery platform may be used by academic laboratories, pharmaceutical companies, hospitals, and contract research organizations. Each user has different obligations regarding patient consent, biosafety, export control, and clinical validation. The system should therefore support role-based access, purpose limitation, and audit trails that record who generated which candidate and why.

Personalized deployment introduces latency and reliability constraints. A cloud-based system can aggregate large compute resources, but it may be unsuitable for hospitals with strict data residency requirements or limited connectivity. An on-premises system can protect privacy, but it may lack the scale to retrain foundation models. Hybrid architectures can distribute inference to local sites while centralizing model updates under federated governance. Edge deployment may be useful for laboratory automation, where generated candidates must be screened quickly against local assays. However, every additional deployment mode increases the attack surface for model theft, data poisoning, and adversarial manipulation. The infrastructure must therefore include anomaly detection, secure model distribution, and human escalation paths. Clinical integration adds further requirements, because a discovery platform may feed into decision support tools that influence patient care. Machine learning in

medicine has shown promise, but it also requires rigorous validation, monitoring, and workflow integration to avoid harm [14], [16].

The governance of personalized deployment should also address intellectual property and benefit sharing. If patient data contribute to a commercial therapeutic, questions arise about consent, compensation, and access. If a model generates a compound that is later patented, the provenance of training data may become legally relevant. These issues are not peripheral to system design. They determine whether patients and clinicians will trust the platform enough to contribute data and adopt its outputs. A sustainable infrastructure must therefore combine technical interoperability with transparent governance. It must allow institutions to participate without surrendering control over sensitive data, while still enabling the cumulative learning that makes multimodal AI valuable. The alternative is a fragmented landscape in which each institution builds isolated models that cannot generalize or be audited.

5. Robustness, Fairness, Sustainability, and Policy Implications

Robustness is a first-order requirement for personalized drug discovery because distribution shift is the norm rather than the exception. A model trained on one set of cell lines, assay protocols, or patient populations may fail when applied to another. Pathogens mutate, tumors evolve, and chemical libraries change. A multimodal system can improve robustness by integrating multiple views, but it can also create new failure modes when one modality dominates or when missing modalities are imputed incorrectly. Robustness testing should therefore include stress tests for corrupted inputs, novel protein families, rare chemical scaffolds, and shifts in assay conditions. The system should report calibration and uncertainty, not only point predictions, so that downstream users know when to trust a generated candidate. Hidden technical debt in machine learning systems makes this difficult because dependencies, feedback loops, and data drift can accumulate invisibly over time [18].

Fairness is equally important because personalized medicine can reproduce or amplify health disparities. If training data underrepresent certain ancestral groups, rare disease subtypes, or low-resource clinical settings, then generated candidates and interaction predictions may be less reliable for those populations. Research on racial bias in clinical algorithms shows that apparently neutral optimization targets can encode structural inequities [15]. A drug discovery platform could similarly optimize for molecular properties that correlate with population-specific data availability rather than therapeutic benefit. Fairness in this context is not a single metric. It includes representation in training data, equitable access to generated therapies, transparency about model limitations, and inclusion of diverse clinical endpoints. Explainable artificial intelligence can help by revealing which features drive predictions, but explanation alone does not guarantee fairness [13]. Governance must connect explanations to accountability, so that biased or unsafe outputs can be corrected.

Sustainability is an emerging concern for large-scale multimodal AI. Training foundation models for protein and molecular generation requires substantial compute, energy, and water for cooling. Inference at scale, especially during generative search and interaction screening, can also be resource-intensive. In personalized drug discovery, the environmental cost must be weighed against the potential to reduce wasteful wet-lab experimentation. A system that generates thousands of poor candidates may increase rather than decrease resource consumption. Sustainable design includes model compression, caching, active learning, and early stopping based on uncertainty. It also includes transparent reporting of compute and energy use so that institutions can make informed trade-offs. Policy mechanisms may

eventually require environmental impact statements for large biomedical models, similar to other high-performance computing infrastructures.

Policy implications extend to regulation, liability, and post-market surveillance. Regulatory agencies currently evaluate drugs and devices, but they are less equipped to evaluate continuously learning discovery platforms that generate candidates and predictions. A model that updates after deployment may change its behavior without a new clinical trial. Policy should therefore distinguish between discovery tools that produce hypotheses and clinical tools that directly influence patient care. For discovery tools, transparency, reproducibility, and biosafety review may be the primary requirements. For clinical tools, prospective validation, monitoring, and liability frameworks are necessary. Post-market surveillance should track not only adverse events but also model drift, inequitable access, and failure modes in underrepresented populations. International coordination is needed because protein and chemical data cross borders, while privacy and biosafety rules remain national. A sustainable policy environment must encourage innovation while preserving public trust.

6. Evaluation, Translation, and Clinical Integration

Evaluation of joint protein generation and compound interaction prediction cannot rely on a single benchmark. Computational metrics such as structural accuracy, binding affinity correlation, and molecular validity are necessary but insufficient. A generated protein structure may score well on a geometric metric while representing a biologically irrelevant conformation. A compound may show high predicted affinity while failing ADMET, synthesis, or selectivity criteria. Evaluation should therefore be multimodal and multi-stage. It should include retrospective validation on held-out assays, prospective wet-lab testing, and comparison against established baselines. It should also include negative controls that test whether the model can distinguish active from inactive compounds under realistic conditions. The antibiotic discovery pipeline that used deep learning to identify new candidates illustrates how computational predictions must be coupled to experimental validation to establish biological relevance [3]. Similarly, generative models for kinase inhibitors have shown that rapid identification is possible, but translation still requires medicinal chemistry and clinical development [4].

Translation from computational discovery to clinical use involves a sequence of evidence requirements. First, the platform must demonstrate technical validity for the intended target and patient context. Second, it must show that generated candidates have acceptable safety and pharmacokinetic profiles. Third, it must establish clinical utility through trials or real-world evidence. These stages are expensive and slow, and most generated candidates will fail. The value of multimodal AI is therefore not guaranteed by more candidates; it depends on improving the probability of success per candidate and reducing the cost of learning from failures. A well-designed platform should capture negative results and feed them back into training. This requires data infrastructure that records failed syntheses, inactive compounds, and off-target effects. Without such feedback, the system will repeat mistakes and overfit to published positive results.

Clinical integration also requires workflow alignment. A discovery platform that produces candidates faster than laboratories can test them may create bottlenecks rather than value. A platform that generates recommendations without explaining uncertainty may be ignored or misused by clinicians. Human factors research and participatory design are therefore essential. Clinicians, medicinal chemists, regulatory experts, and patients should be involved in defining system requirements and evaluation criteria. Explainability can support this process

by showing which structural features or patient variables influenced a prediction, but explanations must be actionable [13]. The platform should also support auditability so that external reviewers can reconstruct how a candidate was generated and why it was prioritized. Translation is not a one-time handoff from computation to clinic. It is an iterative sociotechnical process in which models, workflows, and governance co-evolve.

7. Conclusion

Multimodal AI for personalized drug discovery through joint protein structure generation and compound interaction prediction represents both a technical opportunity and an institutional challenge. The technical opportunity lies in coupling generative design with predictive grounding so that protein conformations and chemical candidates are evaluated together rather than in isolation. Advances in protein structure prediction, protein language modeling, geometric deep learning, and unified multimodal generation provide the necessary foundations [1], [6], [9], [11]. The institutional challenge lies in building data infrastructure, governance, evaluation, and policy frameworks that make these systems trustworthy, equitable, and sustainable. Without such frameworks, multimodal models may produce impressive demonstrations but fail in clinical translation.

The paper has argued that architecture should be designed around uncertainty, modularity, and human review. Data infrastructure should support provenance, privacy, and federated learning. Governance should include datasheets, model cards, access controls, and post-market surveillance. Robustness and fairness should be evaluated under distribution shift and across underrepresented populations. Sustainability should be considered in training, inference, and experimental design. Policy should distinguish between hypothesis-generating discovery tools and clinical decision tools, with appropriate evidence requirements for each. These recommendations are not separate from algorithmic innovation; they are conditions for it. A personalized discovery platform that cannot be audited, updated, or governed will not be adopted at scale, regardless of its benchmark performance.

The path forward requires interdisciplinary collaboration among computer scientists, structural biologists, chemists, clinicians, ethicists, regulators, and patients. The most valuable systems will not simply generate more molecules or structures. They will generate better questions, prioritize more promising experiments, and learn from clinical outcomes in ways that respect privacy and equity. If architecture, infrastructure, and governance co-evolve, multimodal AI can contribute to a more precise and accountable era of drug discovery. If they do not, the field risks reproducing old inequities with new computational confidence.

References

1. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, 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.

2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

3. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe,

A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), 688-702.

4. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., ... Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology, 37(9), 1038-1040.

5. Sanchez-Lengeling, B., & Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: Generative models for matter engineering. Science, 361(6400), 360-365.

6. Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences, 118(15), e2016239118.

7. Hie, B., Zhong, E. D., Berger, B., & Bryson, B. (2021). Learning the language of viral evolution and escape. Science, 371(6526), 284-288.

8. Rao, R., Meier, J., Sercu, T., Ovchinnikov, S., & Rives, A. (2021). Transformer protein language models are unsupervised structure learners. International Conference on Learning Representations.

9. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871-876.

10. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.

11. Wu, H., Song, Y., Zhang, Z., Zhang, Z., Zhou, H., Ma, W. Y., & Liu, J. (2026). Rationalized All-Atom Protein Design with Unified Multi-Modal Bayesian Flow. Advances in Neural Information Processing Systems, 38, 3019-3050.

12. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463-477.

13. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2(10), 573-584.

14. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

15. 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.

16. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.

17. Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37-43.

18. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 28, 2503-2511.

19. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

20. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

21. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229.

22. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.