

Large Language Model-Based Recommendation of Peptide Candidates for Age-Related Disease Research

Omisha Bandi

Department of Computer Science, George Mason University, Fairfax, VA, USA.

hellowmisha@gmu.edu

Abstract

Age-related diseases represent a complex and expanding burden on health systems, and peptide-based therapeutic candidates have emerged as a promising but difficult-to-navigate modality because of their structural diversity, uncertain bioavailability, immunogenicity, and context-dependent mechanisms. This paper examines the system-level design of large language model-based recommendation platforms for peptide candidates in age-related disease research. Rather than treating recommendation as a single generative act, the paper frames it as a socio-technical infrastructure problem involving knowledge integration, retrieval, validation, human oversight, deployment economics, and governance. The analysis considers how foundation models, protein structure prediction, molecular representation learning, and biomedical literature synthesis can be combined within modular architectures that support candidate triage without overstating predictive certainty. It discusses structural trade-offs among centralization and federation, latency and depth, model generality and domain specificity, and openness and intellectual property protection. The paper also addresses robustness under distribution shift, fairness in population and evidence representation, sustainability of compute-intensive workflows, and policy implications for regulatory review. A central argument is that recommendation quality depends less on model scale alone than on the surrounding evidence infrastructure, provenance controls, auditability, and appeal mechanisms that allow researchers to contest or refine outputs. Cross-domain comparisons with clinical decision support, drug repurposing, and materials discovery clarify both the promise and the limits of language model-driven peptide recommendation. The conclusion emphasizes that responsible deployment requires validated benchmarks, transparent model documentation, federated data strategies, and governance arrangements that align discovery acceleration with scientific accountability.

Keywords

large language models; peptide therapeutics; aging-related diseases; recommendation systems; biomedical knowledge infrastructure; governance; robustness; deployment.

1. Introduction

The study of age-related diseases has moved from a collection of isolated organ-specific conditions toward a more integrated understanding of biological aging as a shared risk architecture. This shift has created demand for computational systems that can synthesize heterogeneous evidence and propose plausible therapeutic candidates across molecular, cellular, and clinical scales. Peptides are especially attractive in this context because they can modulate protein-protein interactions, signaling pathways, and immune responses with high specificity, yet they are also difficult to design and prioritize because their activity depends on stability, delivery, target context, and patient heterogeneity. Large language models have entered this space as general-purpose reasoning and synthesis engines that can query literature,

generate candidate explanations, and connect molecular features with disease mechanisms. However, the translation from fluent text generation to reliable candidate recommendation is not automatic. It requires system-level thinking about data provenance, model adaptation, retrieval, evaluation, human decision-making, and the institutional conditions under which recommendations are used [1], [2], [3], [4].

A central premise of this paper is that peptide candidate recommendation for age-related disease research should be treated as a recommendation infrastructure rather than as a standalone chatbot or predictive model. Such an infrastructure must ingest structured and unstructured biomedical knowledge, represent peptide properties and aging mechanisms, rank candidates according to explicit and contestable criteria, and preserve enough provenance for experts to reconstruct why a candidate was surfaced. The recommender must also distinguish between novelty and plausibility, between mechanistic coherence and experimental validation, and between population-level evidence and individual-level inference. These distinctions are not merely technical; they determine whether the system accelerates discovery or amplifies unverified hypotheses. The discussion therefore focuses on architecture, governance, deployment, sustainability, fairness, and policy as first-order design concerns rather than downstream afterthoughts.

2. Aging-Related Disease Research and the Peptide Candidate Space

Aging-related disease research spans neurodegeneration, cardiovascular disease, metabolic dysfunction, musculoskeletal decline, immune senescence, and cancer, among other conditions. The hallmarks of aging framework organizes this space around processes such as genomic instability, epigenetic alterations, loss of proteostasis, disabled macroautophagy, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication, and chronic inflammation [11]. Geroscience extends this framework by linking aging biology to chronic disease susceptibility, arguing that interventions targeting fundamental aging processes may produce broad healthspan effects [12]. Drug discovery for aging has consequently become more systematic, with increasing attention to repurposing, combination therapies, and biomarkers that can serve as intermediate endpoints [13]. Peptides fit this landscape because they can act on many of these hallmarks, including inflammatory signaling, senescence-associated secretory phenotypes, mitochondrial regulation, and neurotrophic support.

The peptide candidate space is nevertheless difficult to navigate. Peptides can be natural, synthetic, modified, cyclic, stapled, or conjugated, and their therapeutic behavior depends on sequence, structure, charge, half-life, permeability, and immunogenicity. Animal-derived peptides have received attention for their potential effects on aging-related diseases, but the evidence base is heterogeneous and often requires careful interpretation across species, preparation methods, and disease models [18]. Computational recommendation must therefore avoid collapsing diverse evidence types into a single relevance score. It must represent mechanistic hypotheses, experimental conditions, model organisms, dosage, delivery route, and outcome measures. Databases such as the Human Ageing Genomic Resources provide structured aging-related knowledge that can anchor recommendations [14], while deep learning in biomedicine and molecular representation learning offer ways to encode chemical and biological relationships [15], [16]. The challenge is to integrate these resources without losing the contextual nuance that determines whether a peptide candidate is genuinely promising.

3. Large Language Models as Scientific Recommendation Infrastructure

Large language models are built on transformer architectures that model long-range dependencies in text and can be adapted to many tasks through prompting, fine-tuning, or retrieval augmentation [2]. Few-shot learning demonstrated that sufficiently large models can perform new tasks from limited examples, which encouraged their use in scientific literature synthesis and hypothesis generation [3]. Foundation model research has since emphasized both the generality of these systems and their risks, including brittleness, opacity, and concentration of power [4]. In medicine, large language models have been shown to encode clinical knowledge and to perform well on some medical question-answering tasks, but they also exhibit limitations in reasoning, calibration, and safety [5], [6]. These findings matter for peptide recommendation because biomedical decisions require more than plausible text; they require traceable evidence, uncertainty communication, and accountability.

Earlier expert systems such as Watson illustrated both the ambition and the difficulty of automating biomedical reasoning at scale [7]. Contemporary language models improve fluency and flexibility, but they do not remove the need for structured domain knowledge. Machine learning in chemoinformatics and drug discovery has long combined molecular descriptors, graph representations, and predictive models to prioritize compounds [8], [9]. Deep learning has enabled rapid identification of potent inhibitors in specific discovery campaigns [10], and molecular representation methods such as contrastive graph learning improve the ability to relate chemical structure to function [16]. A peptide recommender should therefore treat the language model as one component in a larger inference pipeline: it can interpret queries, summarize evidence, propose connections, and generate explanations, while specialized models evaluate structure, binding, stability, and synthesizability. This division of labor reduces the risk of relying on linguistic fluency as a proxy for experimental validity.

4. System Architecture for Peptide Candidate Recommendation

A robust system architecture for peptide candidate recommendation should be modular, auditable, and capable of incorporating heterogeneous evidence. At the ingestion layer, the system must collect biomedical literature, clinical trial records, patent data, protein interaction databases, peptide repositories, aging-related ontologies, and internal experimental results. At the representation layer, it must normalize entities such as genes, proteins, pathways, diseases, peptides, modifications, and assays. At the retrieval layer, it must connect natural language queries to structured records and document passages, often through retrieval-augmented generation, so that the language model can ground its outputs in source material rather than relying solely on parametric memory. At the inference layer, specialized modules can predict structure, estimate binding affinity, assess stability, flag immunogenic motifs, and evaluate drug-likeness. At the recommendation layer, candidates can be ranked by multiple criteria, including mechanistic fit, novelty, feasibility, safety risk, and evidence strength. At the oversight layer, human experts must be able to inspect provenance, adjust weights, and record decisions.

This architecture involves structural trade-offs. A centralized platform can enforce consistent data standards and simplify auditing, but it may create bottlenecks, concentrate sensitive data, and disadvantage smaller research groups. A federated architecture can preserve institutional control and support privacy-preserving collaboration, but it complicates model training, evaluation, and versioning. An end-to-end model may offer convenience and smooth user experience, but it obscures where errors originate and makes it difficult to replace outdated components. A modular pipeline may be more transparent and maintainable, but it imposes

integration costs and can fragment responsibility. Similarly, a system optimized for low latency may rely on cached or approximate retrieval, while a system optimized for depth may run extensive simulations and literature synthesis. The appropriate balance depends on whether the recommender is used for exploratory hypothesis generation, preclinical prioritization, or regulatory-facing evidence assembly. In all cases, the system should expose uncertainty, cite sources, and distinguish between generated hypotheses and validated findings. Protein structure prediction tools such as AlphaFold can support structural plausibility assessment [1], while molecular representation models and chemoinformatics methods can help evaluate candidate properties [8], [16]. Instruction-tuned language models can improve adherence to task constraints, but they still require guardrails and evaluation [17].

5. Data, Knowledge Integration, and Evidence Governance

The quality of peptide recommendation depends heavily on the underlying knowledge infrastructure. Aging research spans decades of literature, diverse model organisms, inconsistent outcome measures, and heterogeneous assay protocols. Integrating this material requires ontologies, controlled vocabularies, provenance metadata, and versioned datasets. Structured resources such as aging genomic databases provide valuable anchors, but they must be linked to peptide-specific data, disease ontologies, and experimental evidence [14]. Natural language processing can extract relationships from publications, but extraction errors can propagate into recommendations. Therefore, evidence governance must include source tracking, confidence scoring, conflict detection, and periodic revalidation. When the system surfaces animal-derived peptides or other natural product candidates, it should indicate the strength and limitations of the supporting evidence rather than presenting a uniform recommendation [18].

Governance also concerns access, privacy, and intellectual property. Biomedical data may be subject to patient privacy regulations, institutional review board constraints, material transfer agreements, and commercial licensing. A recommendation platform that aggregates such data must define who can query which records, how outputs are stored, and whether generated hypotheses can be patented or published. Model documentation practices such as model cards can improve transparency by specifying intended use, training data, performance characteristics, and limitations [20]. However, documentation alone is insufficient if the system cannot be audited or contested. Evidence governance should therefore include mechanisms for expert review, error reporting, version rollback, and dispute resolution. These mechanisms transform the recommender from an opaque oracle into a negotiated scientific instrument.

6. Model Adaptation, Evaluation, and Robustness

Adapting large language models to peptide recommendation can take several forms. Prompt engineering can shape outputs without changing model weights, but it is sensitive to phrasing and may not enforce domain constraints. Retrieval augmentation can ground outputs in current literature, but it depends on index quality and retrieval recall. Supervised fine-tuning can teach the model to follow domain-specific formats and reasoning patterns, but it requires curated examples and can overfit to narrow datasets. Reinforcement learning from human feedback can align outputs with expert preferences, but it may encode the biases of the annotators and can reward stylistic fluency over scientific correctness [17]. A hybrid approach is often preferable: retrieval for grounding, fine-tuning for domain adaptation, specialized predictors for molecular properties, and human feedback for iterative refinement.

Evaluation must go beyond text similarity or user satisfaction. It should assess whether recommended peptides are chemically valid, synthetically feasible, mechanistically plausible, and supported by evidence. It should also measure calibration, because a recommender that is confidently wrong is more dangerous than one that expresses uncertainty. Robustness testing should examine performance under distribution shift, such as when the model is applied to a new disease area, a different species, or a novel peptide modification. It should include adversarial prompts that attempt to elicit unsafe or unsupported recommendations, and it should test whether the system can abstain when evidence is insufficient. Comparisons with clinical language model evaluations show that domain knowledge does not guarantee safe reasoning [5], [6]. The stochastic parrots critique further warns that large models can reproduce patterns without understanding, which makes superficial evaluation especially risky [19]. For peptide recommendation, robustness means maintaining scientific humility while still providing useful prioritization.

7. Deployment, Sustainability, and Scalability

Deployment choices shape who can use the system and under what conditions. Cloud-based deployment offers scalability, elastic compute, and managed services, but it raises concerns about data sovereignty, vendor lock-in, and recurring costs. On-premises or hybrid deployment can protect sensitive data and support institutional control, but it requires significant expertise and capital investment. Academic consortia may prefer federated or shared services that allow querying across institutions without moving raw data. The deployment model also affects sustainability. Large language models require substantial energy for training and inference, and biomedical workflows may add structure prediction, molecular simulation, and database queries. A system that recommends thousands of candidates without filtering may waste compute and researcher attention. Sustainable design therefore favors tiered evaluation, caching, smaller domain-specific models where appropriate, and clear stopping rules for candidate exploration.

Scalability is not only a matter of hardware. It includes the ability to update knowledge bases, retrain or recalibrate models, manage versions, and support growing user communities. A recommender that works for a single laboratory may fail in a multi-institutional setting where vocabulary, data access, and regulatory requirements differ. Interoperability standards and modular interfaces can reduce integration costs. Model cards and transparent documentation can help downstream users understand the system's scope and limitations [20]. Foundation model research highlights the risks of concentrating capability in a few large systems, so deployment should consider pluralism, portability, and the ability of smaller institutions to participate [4]. Sustainability also requires long-term funding for maintenance, because biomedical knowledge decays quickly and an unmaintained recommender can become misleading.

8. Fairness, Policy, and Regulatory Implications

Fairness in peptide recommendation has multiple dimensions. Representation fairness concerns whether training data and knowledge bases adequately cover diverse populations, diseases, and experimental models. Methodological fairness concerns whether the ranking criteria systematically disadvantage certain peptide classes, delivery routes, or research communities. Access fairness concerns whether the benefits of the system are available only to well-resourced institutions. In aging research, these issues are especially salient because aging-related diseases affect populations unevenly, and clinical translation often depends on regulatory pathways that vary across jurisdictions. A recommender that prioritizes

commercializable peptides or well-studied targets may reinforce existing inequalities in research funding and therapeutic development. Governance should therefore include explicit criteria for prioritization, periodic bias audits, and mechanisms for underrepresented groups to influence research agendas.

Policy and regulatory implications are equally significant. If a language model recommends a peptide candidate that later enters preclinical or clinical development, questions arise about evidence provenance, reproducibility, and responsibility. Regulators may not directly approve a recommendation system, but they may scrutinize the data and methods used to support investigational new drug applications. Transparency requirements, such as model documentation and audit trails, can help [20]. However, overly prescriptive regulation may slow innovation or favor large firms that can afford compliance. A balanced policy approach would encourage disclosure of training data sources, evaluation results, uncertainty estimates, and conflict-of-interest arrangements without requiring proprietary model weights to be publicly released. The dangers of large language models, including bias and environmental cost, also suggest that policy should address dual-use risks and the dissemination of potentially harmful peptide candidates [19]. Ethical review boards, institutional biosafety committees, and funding agencies all have roles in setting expectations for responsible use.

9. Case Illustrations and Cross-Domain Comparisons

Several cross-domain examples clarify the opportunities and limits of peptide recommendation. Protein structure prediction demonstrated that deep learning can transform a core biological problem when paired with large curated datasets and rigorous evaluation [1]. Drug discovery campaigns using deep learning have identified potent molecules in specific target contexts, showing that computational prioritization can accelerate early discovery [10]. Expert systems such as Watson showed that integrating knowledge at scale is valuable but also difficult to maintain and validate [7]. These cases suggest that peptide recommendation should be judged by its ability to improve decision-making under uncertainty, not by its ability to generate impressive text. In aging research, a recommender might help triage peptides that modulate senescence, inflammation, or proteostasis, but it must account for the fact that aging is multifactorial and that interventions may have context-dependent effects. Evidence on animal-derived peptides illustrates both the promise and the interpretive complexity of this domain, because findings may depend on preparation, species, and endpoint selection [18].

Cross-domain comparison also reveals institutional differences. Clinical decision support operates under strict safety and liability constraints, while early discovery is more tolerant of exploratory hypotheses. Materials discovery uses simulation and experimental feedback loops that can validate recommendations relatively quickly, whereas aging research often requires long time horizons and complex in vivo models. A peptide recommender for aging should therefore adopt a staged approach: broad hypothesis generation, structured filtering, expert review, and targeted experimental validation. This staged model aligns with the way research organizations already manage risk, but it requires the computational system to support handoffs between stages rather than attempting to automate the entire pipeline.

10. Limitations and Future Directions

Current large language models cannot replace experimental validation, and peptide recommendation remains limited by data scarcity, measurement heterogeneity, and incomplete mechanistic understanding. The system may overrepresent well-studied targets

and underrepresent neglected diseases or non-Western herbal traditions. It may also confuse correlation with causation when synthesizing literature. Future work should develop benchmarks that reflect real discovery decisions, including negative results and failed candidates. Federated learning and privacy-preserving data sharing could expand access to proprietary or sensitive datasets without centralizing them. Multi-modal integration of text, structure, omics, imaging, and assay data could improve candidate ranking, but it also increases complexity and governance burden. Agentic workflows that plan searches and run simulations may increase efficiency, yet they require stronger oversight to prevent uncontrolled action. Finally, long-term field studies are needed to determine whether recommender systems improve research outcomes, reduce bias, or simply accelerate the production of plausible but unvalidated hypotheses [4], [19], [20].

11. Conclusion

Large language model-based recommendation of peptide candidates for age-related disease research is best understood as an infrastructure challenge rather than a single-model problem. The promise lies in synthesizing vast biomedical knowledge, connecting aging mechanisms with peptide properties, and helping researchers prioritize candidates under uncertainty. The risk lies in confusing fluent generation with scientific validity, embedding hidden biases, and deploying systems without adequate provenance, evaluation, or accountability. A responsible architecture should combine retrieval, domain-specific predictors, expert oversight, and transparent documentation. Governance should address fairness, intellectual property, privacy, regulatory expectations, and sustainability. As aging research becomes more data-intensive and computationally mediated, recommender systems will increasingly shape which hypotheses are pursued and which are ignored. Their design must therefore reflect scientific values such as reproducibility, contestability, and equitable access, not merely technical performance. The most credible systems will be those that assist expert judgment while remaining honest about their limits.

References

1. 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>
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. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
4. 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. S., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://doi.org/10.48550/arXiv.2108.07258>

5. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole, E., Schulman, J., Wang, P., Gottesman, S., Ewalds, T., Khosla, S., Huang, K., Goyal, N., Dandala, B., Sood, M., Fiedler, T., ... Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, 620(7972), 172–180. <https://doi.org/10.1038/s41586-023-06291-2>
6. Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930–1940. <https://doi.org/10.1038/s41591-023-02448-8>
7. Ferrucci, D. A. (2012). Introduction to "This is Watson". *IBM Journal of Research and Development*, 56(3/4), 1:1–1:15. <https://doi.org/10.1147/JRD.2012.2184356>
8. Lo, Y.-C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today*, 23(8), 1538–1546. <https://doi.org/10.1016/j.drudis.2018.05.010>
9. 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. <https://doi.org/10.1038/s41573-019-0024-5>
10. 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., Asetov, A. S., Veselov, A. S., ... Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), 1038–1040. <https://doi.org/10.1038/s41587-019-0224-x>
11. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2023). Hallmarks of aging: An expanding universe. *Cell*, 186(2), 243–278. <https://doi.org/10.1016/j.cell.2022.11.001>
12. 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. <https://doi.org/10.1016/j.cell.2014.10.039>
13. Partridge, L., Fuentealba, M., & Kennedy, B. K. (2020). The quest to slow ageing through drug discovery. *Nature Reviews Drug Discovery*, 19(8), 513–532. <https://doi.org/10.1038/s41573-020-0067-7>
14. de Magalhães, J. P., Budovsky, A., Lehmann, G., Costa, J., Li, Y., Fraifeld, V., & Church, G. M. (2009). The Human Ageing Genomic Resources: Online databases and tools for biogerontologists. *Aging Cell*, 8(1), 65–72. <https://doi.org/10.1111/j.1474-9726.2008.00442.x>
15. Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, 13(5), 1445–1454. <https://doi.org/10.1021/acs.molpharmaceut.5b00982>
16. Wang, Y., Wang, J., Cao, Z., & Barati Farimani, A. (2022). Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3), 279–287. <https://doi.org/10.1038/s42256-022-00447-x>

17. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Aspell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
18. 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.
19. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
<https://doi.org/10.1145/3442188.3445922>
20. 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.
<https://doi.org/10.1145/3287560.3287596>