
人工智能赋能生物武器时代的韧性生物安全

Jonathan Feldman

Georgia Institute of Technology

Atlanta, Georgia

jonathanfeldman@gatech.edu

Tal Feldman

Yale Law School

New Haven, CT

tal.feldman@yale.edu

ABSTRACT

近期生成生物学的进展使得设计新型蛋白质成为可能，为药物发现提供了重大机会，同时也引入了新的风险，包括潜在合成生物武器的发展。现有的生物安全措施主要依赖于推理时过滤器（如序列比对和蛋白质-蛋白质相互作用 (PPI) 预测）来检测危险输出。在这项研究中，我们评估了三个领先的 PPI 预测工具的表现：AlphaFold 3、AF3Complex 和 SpatialPPIv2。这些模型在一些特征明确的病毒与宿主之间的相互作用上进行了测试，包括乙型肝炎和 SARS-CoV-2 涉及的相互作用。尽管经过了许多相同的病毒训练，但这些模型未能检测到大量已知的相互作用。令人惊讶的是，没有一种工具能够成功识别出任何四种经实验验证且确认具有结合能力的 SARS-CoV-2 突变体。这些发现表明，目前的预测过滤器不足以可靠地标记甚至已知的生物威胁，并且更不可能检测到新的威胁。我们主张转向以响应为导向的基础设施建设，包括快速实验验证、可适应的生物制造和能够跟上 AI 驱动发展速度的监管框架。

Keywords 生物安全 · 政府 · 基础模型 · 国家安全 · 法律 · 规定

1 介绍

在近年来，生成式人工智能与分子生物学的融合产生了能够以令人惊叹的精度设计蛋白质的强大新系统。蛋白质语言模型 (PLMs) 以及结构预测器和设计工具如 RFdiffusion 正在通过快速研发治疗药物——从新型酶和疫苗到针对特定疾病的抗体候选物——革新医药领域 [1, 2, 3, 4, 5]。这些系统正在加速生物医学创新，缩短实验时间线，并将药物设计的范围扩展到以前难以解决的目标 [6, 7]。

然而，这种相同的能力引入了一个前所未有的双重用途困境。现在用于制造救命生物制剂的工具，在错误的人手中，可以被重新利用来生成具有增强传染性、免疫逃逸或毒性的病原体 [8]。PLMs 可以模拟病毒蛋白的进化轨迹，建议使它们更具传播性或更难中和的突变 [9, 10, 11]。结构感知设计工具可以在计算机上验证这些输出并提出新的支架来支持它们 [6, 12]。这个设计和验证循环——类似于治疗生物技术领域已经使用的流程——降低了制造合成威胁的门槛，其中一些可能新颖到完全逃避检测 [8]。

针对这一日益增长的担忧，人工智能和生物安全社区提出了技术对策 [13]。这些包括推理时过滤以阻止危险输出、红队测试模型以防滥用以及序列筛查以标记潜在的致病分子 [8, 13, 14]。然而，这些方法基于一个有问题的假设：我们可以可靠地通过计算识别有害蛋白质。

这一假设在仔细审查下会瓦解。一种蛋白质是否无害或有害取决于一些因素，即使是最先进的模型也难以理解这些因素——它的三级结构、细胞环境、结合特异性和免疫原性 [8, 14, 15]。大多数这些属性无法仅从序列中推断出来。事实上，正如我们在本文的技术部分所展示的那样，现有的蛋白质-蛋白质相互作用 (PPI) 预测

模型在应用于病毒蛋白时尤其表现不佳。它们缺乏数据多样性、上下文理解和生物学真实性，因此难以推广到狭窄且过度代表的领域之外。这一限制造成了盲点，可能会被恶意行为者利用，或者在善意部署中被忽视。

如果计算过滤器不可完全信赖，则只有实验才能提供最终的安全保障。然而，诸如结合测定、毒性筛选和感染模型等实验验证方法太慢、成本太高且资源密集，无法与模型输出 [16, 17] 扩展同步。理论上，标记的序列可以在实验室中进行测试。实际上，我们缺乏足够的基础设施来审查超过极小的一部分 [8, 18]。这使我们处于一种危险的不对称状态：生成模型可以大规模生产威胁，但我们验证它们的能力远远落后。

更糟糕的是，即使我们在从序列数据预测病原性方面取得了重大进展，也无法保证安全。开源模型在公共数据上进行了微调，在全球范围内均可获得。滥用的障碍正在降低：DNA 合成服务越来越容易获取，基础实验室工具广泛分布，所需的专业知识也不再局限于国家资助的项目 [8, 12, 19, 20, 21]。拥有足够资源的民族国家和非国家行为者可以私下运行模型，绕过安全机制，并在几乎没有监管的情况下合成输出结果。在这种威胁环境中，序列分类的技术成功并不能保证安全。我们不能仅仅依赖于预防性过滤器。

含义是明确的：目前的防御策略——专注于预测、预防和边界控制——在本质上与威胁的本质不匹配。无论我们的红队演练变得多么精细，无论我们部署多少推理时的过滤器，当我们核心工具缺乏真实数据且潜在对手不受约束时，我们就无法保证安全。

所需的是一个转变：从遏制的范式转向韧性。与其试图在生成时拦截每一个潜在危险的输出，我们必须建立能够在威胁出现后作出反应的系统——比那些威胁传播得更快。这意味着要开发用于治疗性响应的人工智能原生基础设施：训练模型以按需设计安全、稳定且可制造的对策；能够并行筛选数千个候选对象的高通量仿真和验证工具；以及一个全国性的生物制造管道，能够在数天内从和平时期的科研转变为紧急部署。

至关重要的是，这不仅仅是一个技术上的追求——它是一项国家安全的迫切需求。我们必须为一个合成病原体可能被生成并释放得比传统公共卫生系统检测或控制它们还快的世界做好准备。在这样一个世界中，韧性是唯一可行的防御措施。速度——而不是克制——成为战场。

本文结构如下：在下一节中，我们将展示当前用于预测病毒蛋白相互作用的机器学习模型，特别是那些依赖大规模序列和结构建模的模型，在泛化方面表现不佳。这一点在低数据量和高差异性场景中尤为明显。这些缺点突显了现有安全框架的局限性，并强调迫切需要将研究重点转向响应驱动的能力。然后，我们提出了一种基于韧性（resilience）的框架来解决现有生物安全努力中的空白。最终，我们的目标不是消除生成生物学，而是引导它——确保今天开发的工具能够与未来的需求保持一致。

2 当前推理时间过滤方法的局限性

识别潜在致病蛋白质，特别是那些可能经过合成工程或基因改造的蛋白质，在推理时间过滤时需要强大的可靠工具。在生物安全的背景下，这些工具必须能够预测给定的蛋白质序列或结构是否对人类有害。目前，用于预测致病性的两种主要方法是序列相似性比较和 PPI 预测机制 [14, 22, 23, 24]。尽管这些工具已被证明很有用，但它们在适用性和准确性方面从根本上存在局限性，尤其是在检测新型或工程化蛋白质时。

序列相似性搜索被认为是识别已知病原蛋白最直接和有效的工具。通过将输入的蛋白质序列与大量已知序列数据库进行比较，研究人员可以迅速识别来自自己建立病毒株的同源蛋白质 [25]。该方法特别适用于识别经历了微小突变或源自先前已知病毒蛋白的病原蛋白。诸如 BLAST 之类的工具促进了与 UniProt 和 NCBI Virus 等包含数千种病毒序列的数据库之间的蛋白质序列比较 [26, 27, 28]。这些比较允许确定输入蛋白质与已知病毒序列之间的同源性程度，提供了检测病原性的相对简单且高度可扩展的方法 [26, 29]。当同源性程度超过预定义阈值时，该蛋白可以被标记为潜在的病原体。

尽管其有效性，序列相似性搜索从根本上受到依赖现有数据库的限制。由于这些数据库由已知序列组成，它们本质上无法检测与以前编目内容足够不同的新型病毒蛋白 [14]。这意味着虽然序列相似性搜索可以有效识别已知病原体的变体，但它们无法检测完全新颖的蛋白质，如使用先进计算技术（例如 RFdiffusion 或 ESM3）

设计的那些，这些技术生成的新颖蛋白质序列不存在于任何现有数据库中 [6, 12]。因此，尽管序列相似性搜索仍然是检测已知病原体的重要工具，但它们不能依赖来识别全新的威胁，特别是那些经过精心设计的。

为了解决这一限制，研究人员转向了 PPI 预测方法 [30, 31]。PPI 在许多病毒的致病性中起着核心作用，因为病毒蛋白通常需要与人类细胞受体相互作用以进入宿主 [32, 33, 34]。虽然大多数病毒结合标准受体，但有些还会与宿主细胞表面的糖蛋白和糖脂相互作用 [34]。PPI 预测工具的研发旨在识别这些相互作用，从而提供一种补充序列相似性搜索的方法。理论上，PPI 预测工具可用于确定病毒蛋白是否可能结合到人类细胞上的受体，作为致病性的合理替代指标。这将有助于识别出那些可能会与人类受体发生交互的蛋白质，即使它们与任何已知病原体没有密切关系。

然而，在实践中，PPI 预测工具——特别是基于人工智能的那些——往往无法提供高风险应用（如生物安全 [35]）所需的准确性。这些模型的固有限制在考虑到不准确性的潜在后果时尤其令人担忧。一个假阴性结果，即一种危险的病原蛋白未被检测到，可能会对公共健康产生严重后果。正因为如此，PPI 预测模型的准确性是一个必须解决的关键问题，才能使这些工具可靠地用于生物安全目的。

为了评估最先进的蛋白质-蛋白质相互作用 (PPI) 预测模型的性能，我们使用了十三个特征明确的病毒蛋白-人类受体相互作用进行了基准研究。这些相互作用是从结构完全或部分解析并在蛋白质数据库 (PDB) 中可用的病毒中选择的，例如 HIV、麻疹和 SARS-CoV-2 [36]。选择了三种广泛使用的 PPI 预测工具进行测试：AlphaFold 3、AF3Complex 和 SpatialPPIv2。AlphaFold 3 是由 Google DeepMind 开发的一种基于深度学习的模型，普遍认为是目前可用于预测蛋白质结构的最佳工具 [37]。AF3Complex 是 AlphaFold 3 的衍生产品，专门设计用于通过更准确地建模这些相互作用来改进 PPI 预测 [38]。SpatialPPIv2 是一种基于图神经网络的模型，代表了 PPI 预测方法的最新进展 [23]。

每个模型都提供了病毒结合蛋白及其相应的人类受体的蛋白质序列。AlphaFold 3 和 AF3Complex 使用基于序列的输入来预测单个蛋白质的结构，而 SpatialPPIv2 则使用由 AlphaFold 3 [23, 37, 38] 预先计算生成的结构。SpatialPPIv2 与 AlphaFold 3 和 AF3Complex 的区别在于它专门训练用于通过结构输入而不是序列信息来预测 PPIs [23]。AlphaFold 3 依赖其内部生成的接口预测 TM 分数 (ipTM) 来评估预测的蛋白质复合体结构的信心；先前的研究使用 ipTM 分数为 0.6 或更高作为有意义的 PPI 的代理指标 [37, 39]。相比之下，AF3Complex 使用单独的指标，即预测的界面相似性分数 (pIS)，分数达到 0.38 及以上表示可能存在的 PPI [38, 40, 41]。SpatialPPIv2 则输出输入蛋白质之间存在有意义相互作用的概率，使用阈值为 0.5 或更高来指示预测的 PPI。

基准测试研究的结果，汇总在表 1 中，揭示了所有三种蛋白质-蛋白质相互作用 (PPI) 预测模型在一致识别文献中记录良好的病毒-宿主相互作用方面存在显著失败。值得注意的是，AlphaFold 3 和 SpatialPPIv2 都未能正确识别 SARS-CoV-2 刺突蛋白与人类 ACE2 受体之间的相互作用——这一相互作用对于病毒感染人体细胞的能力至关重要，并且是 COVID-19 大流行的核心 [9, 42]。总体而言，SpatialPPIv2 错误地识别了大约 40% 的 PPIs，而 AlphaFold 3 错误地识别了约 50%，AF3Complex 错误地识别了约 30%。这些结果表明，即使是最先进的模型在预测病毒-宿主相互作用方面也远不可靠，这削弱了这些模型用于识别计算设计的病原体蛋白质的能力。此外，基准测试研究中包含的许多 PPIs 都存在于所有三个模型的训练数据集中 [37, 38, 43]。这意味着这些模型在训练过程中已经接触到了这些相互作用，但在测试阶段仍未能准确预测它们。这是一个特别令人担忧的结果，因为它表明这些模型并不能很好地推广到现实世界的 PPI 预测中，即使当这些相互作用是众所周知且特征明确时也是如此。如果这些模型连最确切的病毒 PPIs 也无法准确预测，那么它们检测新型或工程化的病毒-宿主相互作用的可能性变得更加渺茫。

作为辅助实验，我们评估了四种 SARS-CoV-2 刺突受体结合域的变异体，这种病毒蛋白负责介导 COVID-19 感染。这些变异源自先前的一项研究 [44]，该研究实验测量了它们对刺突 RBD 与人类 ACE2 受体之间结合亲和力的影响，这是蛋白质间相互作用强度的关键决定因素。在所选的变异体中，两种显示出对 ACE2 结合亲和力的最大增强，增加超过十倍，而另外两种则表现出最显著的亲和力降低但仍保持功能性相互作用。

我们使用之前评估过的三种基于结构的模型——SpatialPPIv2、AlphaFold3 和 AF3Complex——来评估这四种刺突突变体，以确定任何刺突与 ACE2 之间的蛋白质相互作用是否被计算预测。

病毒	空间蛋白质相互作用版本 2	AlphaFold3	AF3 复数
Adenovirus	✓	✓	✓
Chikungunya	×	×	×
Ebola	✓	×	✓
Epstein-Barr	✓	✓	✓
Hepatitis B	×	×	×
HIV	✓	✓	✓
HSV (HVEM Receptor)	✓	×	×
HSV (Nectin-1 Receptor)	✓	✓	✓
Measles	×	×	×
Measles (Edmonston strain)	×	✓	✓
MERS	✓	✓	✓
Nipah	✓	✓	✓
SARS-CoV-2	×	×	✓

表 1: 病毒-宿主蛋白质-蛋白质相互作用预测性能在三种最先进的计算模型中的表现。每一行代表不同的病毒株或受体变异，勾号（✓）表示成功识别 PPI，叉号（×）表示预测失败。这些模型有不同的接受置信度阈值：SpatialPPIv2 (0.5)、AlphaFold3 (0.6) 和 AF3Complex (0.38)。

结果见表 2。

令人惊讶的是，所有三个模型都未能检测到任何变体的有意义相互作用，包括那些实验确认具有高亲和力结合的变体。这揭示了对组合突变效应的强大不敏感性，强调了一个显著的盲点：在没序列相似性搜索的情况下，这些模型可能无法识别致病病毒变体。这引发了对其在验证病毒致病性方面效用的严重担忧。

更加令人担忧的是模型信心变化的方向。对于有益和有害的突变，SpatialPPIv2 和 AlphaFold3 增加了它们的预测信心，而 AF3Complex 减少了其自身信心，无论真正的生物学影响如何。这种不稳定的行爲調了這些模型在面對發生突變的病毒蛋白時的局限性——這是一個更加致命或更具逃逸性的病原体可能出現的重要途徑 [8]。

这些失败的主要原因之一是基础数据不足，现有的训练数据集中缺乏高质量的病毒-宿主蛋白质相互作用示例 [45, 46]。尽管像 PDB 这样的资源包含数十万个蛋白质结构，但其中绝大多数代表的是非病毒蛋白质，而病毒蛋白，特别是与宿主相互作用相关的病毒蛋白只占很小一部分 [36, 46]。这种针对性数据的缺乏限制了模型学习特定于病毒蛋白的相关互动特征的能力。

然而，这种数据不平衡只是挑战的一部分。由于蛋白质与蛋白质相互作用的复杂生物物理性质 [45]，泛化到新奇或工程化的蛋白质仍然本质上是困难的。即使有了改进的病毒数据集，预测涉及合成或之前未见过的蛋白质（如现代蛋白质设计工具生成的那些）的相互作用仍然是一个重大障碍。这些人工蛋白质可能与已知的

病毒蛋白不相似，并且可能会以非常规的方式与宿主受体相互作用，使得训练于天然蛋白质的模型特别难以检测它们 [6, 12]。

亲和力变化 ($\Delta \log_{10} K_D$)	空间 PPIv2	AlphaFold3	AF3 复数
-1.04	0.27 (+0.08)	0.16 (+0.03)	0.11 (-0.47)
+2.14	0.19 (+0.00)	0.24 (+0.11)	0.06 (-0.52)
-1.03	0.26 (+0.07)	0.14 (+0.01)	0.06 (-0.52)
+2.11	0.24 (+0.05)	0.15 (+0.02)	0.05 (-0.53)

表 2: Sars-Cov-2 刺突蛋白变异体上 PPI 预测模型的性能比较。每一行代表一个不同的突变型及其相应的结合亲和力变化 ($\Delta \log_{10} K_D$)，负值表示结合亲和力增加，正值则表示相反的情况。括号中的数值显示了各模型对野生型 Sars-Cov-2 刺突蛋白与人 ACE2 PPI 的基准置信度分数的差异。SpatialPPIv2 的 PPI 识别阈值为 0.5，而 AlphaFold 3 和 AF3Complex 的分别为 0.6 和 0.38。这三个模型都没有正确预测任何一个蛋白质相互作用。

这一差距凸显了当前蛋白质相互作用预测方法的更广泛限制。当模型应用于熟悉的蛋白质类型时，可以表现出合理的性能，但在需要泛化的高风险场景中，如检测具有潜在危害的工程蛋白时，则表现不佳。随着生成工具变得越来越易于访问和功能强大，无法可靠评估新型蛋白质的风险成为一个日益严重的生物安全问题。

3 重新定向研究以应对基础设施准备和快速响应

当前的推理时间过滤器和蛋白质-蛋白质相互作用预测器不足以捕捉危险的合成蛋白。随着生成性生物模型的进步与普及，很明显生物安全不能仅依赖静态防护措施。相反，重点必须转向构建一个能够以机器时间尺度设计、验证、制造和部署医疗对策的弹性基础设施。这意味着联邦机构的领导力以及大学、研究实验室和工业界在研究优先事项上的广泛调整。

到目前为止，大多数用于治疗设计的生成式 AI 投资集中在制药行业和生物技术初创公司 [47, 48]。这些努力自然侧重于具有强大商业激励和已知生物学目标的常见疾病。但这留下了一个空白：对于与生物防御最相关的罕见、快速演变或经过工程改造的病原体，研究很少 [8, 49, 50]。学术研究必须填补这一空白。大学和联盟非常适合研究行业研发通常忽视的边缘案例。通过针对新型病毒、未知蛋白质结构以及当今生物学数据之外的人工序列，它们可以帮助构建对对齐和安全至关重要的工具。

这些努力如果不解决更深层次的问题将会失败：训练和评估对策模型所需的关键数据是分散且孤立的。病原体进化、免疫反应、无效疗法以及合成构建物的信息分布在互不连接的学术界、企业及政府系统中 [51]。整合这些数据至关重要，但访问控制也同样重要。在国家安全背景下，广泛共享是有风险的。应将访问权限限制在经过认证的研究实验室和在明确保障措施下运营的可信赖合作伙伴。我们需要持续投资共享基础设施以支持这一点：已知和假设病毒株的存储库、免疫原性和毒性的标准化测量以及透明的临床性能数据。资助机构应该优先考虑这些努力，学术机构必须愿意建立并维护安全平台，实现合作而不会损害生物安全。COVID-19 应对措施显示了快速协调的数据共享可以带来的成果 [52]；现在生防需要一个既追求速度又强调控制的版本。

超越数据和建模，学术界有责任重新构建对蛋白质生成设计的概念化方式。围绕 PLMs 和其他生成工具的大部分讨论都集中在投机性风险上，特别是它们可能被用于制造有害生物制剂的风险 [8, 13, 49, 50]。但由生成蛋白模型带来的风险已不再仅仅是理论上的，研究界需要相应地对待这些风险。随着这些系统能力的增长，尤其是在设计新颖或工程化蛋白质方面，其潜在的滥用成为一个具体关注的问题。应对这一问题需要的不仅

仅是表面的安全措施或临时过滤。它要求结构性解决方案。将生物安全视为次要或事后问题已不再可行。这些挑战处于负责任模型开发的核心，并必须纳入核心研究议程。

尽管学术界可以帮助定义准备工作的技术基础，实际部署的责任仍然牢牢掌握在联邦政府手中。人工智能生成的治疗药物的湿实验室验证、临床测试和大规模制造将需要协调、供应链弹性以及紧急基础设施，这些都是任何单一实验室或机构无法承担的任务。生物学高级研究与发展管理局（BARDA）、食品药品监督管理局（FDA）和战略准备与响应管理局（ASPR）等机构必须在建立将模型输出转化为可部署对策的国家基础设施方面发挥主导作用。

这包括建立实验筛选平台，可以并行测试数千种蛋白质设计的结合、毒性、可制造性和免疫反应。这些系统应该是快速的、模块化的和自动化的，以跟上模型输出的速度。但仅仅筛选是不够的。联邦政府还应支持一个分布式的生物制造站点网络，准备在危机中生产对策。这些设施应该保持待命状态，储存关键材料，并能够迅速从常规工作转换到应急响应。冗余和地理上的分散对于避免供应链故障和确保韧性至关重要。

现代化监管基础设施同样重要。现有的审批流程，特别是 FDA 内部的流程，并未调整到生物紧急情况所需的时间表 [53, 54]。监管路径必须适应 AI 生成的治疗药物，通过建立预授权快速通道机制、附条件紧急批准并带有部署后监控以及整合实时安全数据的自适应临床试验来实现。这些系统必须比那些为传统药物开发设计的速度更快。此外，监管改革还必须伴随操作灵活性。公共卫生机构内部预先定位的决策权必须被赋予在紧急情况下迅速行动的能力，避免出现威胁以数字速度发展时可能导致致命后果的官僚延迟。

为了确保系统不仅在原则上而且在实践中能够正常工作，整个对策流程必须定期进行压力测试。由国家安全委员会、BARDA、ASPR、国防部及其他相关部门协调的跨机构模拟应演练从生物威胁序列检测到治疗部署的全过程。这些模拟应包括实操练习以及桌面模型，并且必须评估不仅是技术性能还包括制度反应能力。学术机构和研究型医院可以通过提供实验能力、验证工具和情景建模参与这些模拟。

同时，联邦政府必须创建长期的资金机制来维持这一基础设施。私营部门不太可能建设和维护那些可能会不经常启用的准备平台，并且商业激励措施与公共卫生准备工作并不对齐。诸如 NIH、NSF 和 DARPA 这样的机构必须为 AI 治疗设计系统、筛选平台和生物制造能力提供稳定支持，要求具备双重用途部署——在和平时期可用于研究，在危机时可随时用于防御。具有兼容性的公私合作将至关重要。研究补助资格、快速监管路径以及采购合同应以符合生物安全、网络安全和审计标准为条件。在高后果情况下，对紧急部署有限责任保护可能是适当的，前提是保持透明度和安全性。

总体而言，从预防转向韧性需要科学生态系统各个层面的贡献。联邦机构必须在基础设施、制造、验证和监管方面发挥领导作用。学术机构必须推进技术基础建设、构建数据基础设施，并推动治疗一致性。私营部门必须被激励将安全性和准备性整合到商业创新中。生成生物学带来的风险是真实的，但机会也同样存在。如果我们迅速行动，那些可能加速生物威胁的系统也可以成为我们最强大的应对工具。

4 结论

生成式 AI 工具能够设计新颖的生物序列，这一迅速出现的技术为生物医学发现带来了新的机遇，同时也引入了误用的新途径。这些模型可以产生没有明显进化先例的蛋白质候选物，这使得它们潜在的功能变得模糊，并且使下游风险评估复杂化。正如我们的案例研究所示，依赖于事后过滤器——如序列相似性搜索或 PPI 预测——来评估生成序列的安全性是不够的。基于序列的经验法则通常无法捕捉功能的新颖性，而 PPI 预测器仍然过于不准确且校准不足，无法可靠地区分良性与危险的输出。

这些限制暴露了一个更广泛的结构性问题：推理时的干预措施本质上是反应性的和推测性的。在没有可靠、可扩展的实验验证的情况下，仅靠过滤无法保证安全。更糟糕的是，过度依赖这些技术可能会产生一种虚假的安全感，从而推迟对实质性防护措施的投资。

为应对这些挑战，我们建议将生物安全政策的重点从通过预测进行预防转向通过响应增强韧性。这包括对快速治疗设计、高通量筛选、可扩展生物制造和敏捷监管监督的基础设施进行投资。目标不是消除所有生成时的风险，而是确保新兴威胁能够被实时识别和缓解。

生成模型无疑将在塑造生物学的未来中发挥越来越重要的作用。确保其益处得以实现而不损害安全，需要对当前的技术限制进行现实评估，并建立以稳健性而非单纯预测为基础的政策框架。随着这些技术的发展，研究人员和决策者都必须构建不仅能够预见威胁，而且能够在威胁出现时迅速有效应对的系统。

5 数据和模型可用性

用于本文分析的 SpatialPPIv2 模型可在此处访问：GitHub 仓库。同样，所使用的 AF3Complex 模型也可在此处访问：GitHub 仓库。AlphaFold 3 模型由 Google DeepMind 托管在此处：网页服务器。十三个病毒-受体相互作用案例研究的 PDB 参考结构包括已解析的蛋白质相互作用和相关序列。它们如下：对于 SARS-CoV-2，参考结构是 6M0J。对于 MERS-CoV，结构是 4L72。HIV 相互作用由 1GC1 表示，腺病毒相互作用则用 1F5W 捕捉。单纯疱疹病毒的相互作用由两个结构代表：HVEM 受体使用 1JMA，Nectin-1 受体使用 3SKU。乙型肝炎病毒由 8HRX 表示。Epstein-Barr 病毒的相互作用通过结构 8SM0 描述。两种不同的麻疹病毒结构被包含在内：Edmonston 株用 3INB 表示，另一种株则用 4GJT 表示。埃博拉病毒由结构 5F1B 代表，尼帕病毒用 2VSM 表示，基孔肯雅热病毒用 6JO8 表示。实验验证的 SARS-CoV-2 突变体序列，参考自之前的一项研究 [44]，在此处公开可用：GitHub 仓库。

参考文献

- [1] Jeffrey A. Ruffolo and Ali Madani. Designing proteins with language models. *Nature Biotechnology*, 42(2): 200–202, February 2024. ISSN 1546-1696. doi: 10.1038/s41587-024-02123-4. URL <https://doi.org/10.1038/s41587-024-02123-4>.
- [2] Lin Chen, Zehong Zhang, Zhenghao Li, Rui Li, Ruifeng Huo, Lifan Chen, Dingyan Wang, Xiaomin Luo, Kaixian Chen, Cangsong Liao, and Mingyue Zheng. Learning protein fitness landscapes with deep mutational scanning data from multiple sources. *Cell Systems*, 14(8):706–721.e5, 2023. ISSN 2405-4712. doi: <https://doi.org/10.1016/j.cels.2023.07.003>. URL <https://www.sciencedirect.com/science/article/pii/S2405471223002107>.
- [3] Yijia Xiao, Wanjia Zhao, Junkai Zhang, Yiqiao Jin, Han Zhang, Zhicheng Ren, Renliang Sun, Haixin Wang, Guancheng Wan, Pan Lu, Xiao Luo, Yu Zhang, James Zou, Yizhou Sun, and Wei Wang. Protein large language models: A comprehensive survey. 2025. URL <https://arxiv.org/abs/2502.17504>.
- [4] Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alex Rives. Language models enable zero-shot prediction of the effects of mutations on protein function. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 29287–29303. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/f51338d736f95dd42427296047067694-Paper.pdf.
- [5] Brian L. Hie, Kevin K. Yang, and Peter S. Kim. Evolutionary velocity with protein language models predicts evolutionary dynamics of diverse proteins. *Cell Systems*, 13(4):274–285.e6, April 2022. ISSN 2405-4720. doi: 10.1016/j.cels.2022.01.003.
- [6] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Coubet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion.

- Nature*, 620(7976):1089–1100, August 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL <https://www.nature.com/articles/s41586-023-06415-8>. Publisher: Nature Publishing Group.
- [7] Weronika Bielska, Igor Jaszczyszyn, Pawel Dudzic, Bartosz Janusz, Dawid Chomicz, Sonia Wrobel, Victor Greiff, Ryan Feehan, Jared Adolf-Bryfogle, and Konrad Krawczyk. Applying computational protein design to therapeutic antibody discovery – current state and perspectives, 2025. URL <https://arxiv.org/abs/2503.00913>.
- [8] Dianzhuo Wang, Marian Huot, Zechen Zhang, Kaiyi Jiang, Eugene I Shakhnovich, and Kevin M Esvelt. Without Safeguards, AI-Biology Integration Risks Accelerating Future Pandemics. 2025. doi: 10.13140/RG.2.2.29765.15849. URL <https://rgdoi.net/10.13140/RG.2.2.29765.15849>. Publisher: Unpublished.
- [9] Marian Huot, Dianzhuo Wang, Jiacheng Liu, and Eugene Shakhnovich. Few-Shot Viral Variant Detection via Bayesian Active Learning and Biophysics, March 2025. URL <https://www.biorxiv.org/content/10.1101/2025.03.12.642881v1>. Pages: 2025.03.12.642881 Section: New Results.
- [10] Noor Youssef, Sarah Gurev, Fadi Ghantous, Kelly P. Brock, Javier A. Jaimes, Nicole N. Thadani, Ann Dauphin, Amy C. Sherman, Leonid Yurkovetskiy, Daria Soto, Ralph Estanboulieh, Ben Kotzen, Pascal Notin, Aaron W. Kollasch, Alexander A. Cohen, Sandra E. Dross, Jesse Erasmus, Deborah H. Fuller, Pamela J. Bjorkman, Jacob E. Lemieux, Jeremy Luban, Michael S. Seaman, and Debora S. Marks. Computationally designed proteins mimic antibody immune evasion in viral evolution. *Immunity*, 58(6):1411–1421.e6, June 2025. ISSN 1074-7613. doi: 10.1016/j.immuni.2025.04.015. URL [https://www.cell.com/immunity/abstract/S1074-7613\(25\)00178-5](https://www.cell.com/immunity/abstract/S1074-7613(25)00178-5). Publisher: Elsevier.
- [11] Peng Cheng, Cong Mao, Jin Tang, Sen Yang, Yu Cheng, Wuke Wang, Qiuxi Gu, Wei Han, Hao Chen, Sihan Li, Yaofeng Chen, Jianglin Zhou, Wuju Li, Aimin Pan, Suwen Zhao, Xingxu Huang, Shiqiang Zhu, Jun Zhang, Wenjie Shu, and Shengqi Wang. Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. *Cell Research*, 34(9):630–647, September 2024. ISSN 1748-7838. doi: 10.1038/s41422-024-00989-2. URL <https://www.nature.com/articles/s41422-024-00989-2>. Publisher: Nature Publishing Group.
- [12] Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q. Tran, Jonathan Deaton, Marius Wiggert, Rohil Badkundri, Irhum Shafkat, Jun Gong, Alexander Derry, Raul S. Molina, Neil Thomas, Yousuf Khan, Chetan Mishra, Carolyn Kim, Liam J. Bartie, Matthew Nemeth, Patrick D. Hsu, Tom Sercu, Salvatore Candido, and Alexander Rives. Simulating 500 million years of evolution with a language model, July 2024. URL <https://www.biorxiv.org/content/10.1101/2024.07.01.600583v1>. Pages: 2024.07.01.600583 Section: New Results.
- [13] Mengdi Wang, Zaixi Zhang, Amrit Singh Bedi, Alvaro Velasquez, Stephanie Guerra, Sheng Lin-Gibson, Le Cong, Yuanhao Qu, Souradip Chakraborty, Megan Blewett, Jian Ma, Eric Xing, and George Church. A call for built-in biosecurity safeguards for generative AI tools. *Nature Biotechnology*, 43(6):845–847, June 2025. ISSN 1546-1696. doi: 10.1038/s41587-025-02650-8.
- [14] David Baker and George Church. Protein design meets biosecurity. *Science*, 383(6681):349–349, 2024. doi: 10.1126/science.ado1671. URL <https://www.science.org/doi/abs/10.1126/science.ado1671>.
- [15] Constance J. Jeffery. Current successes and remaining challenges in protein function prediction. *Frontiers in Bioinformatics*, 3, July 2023. ISSN 2673-7647. doi: 10.3389/fbinf.2023.1222182. URL <https://www.frontiersin.org/journals/bioinformatics/articles/10.3389/fbinf.2023.1222182/full>. Publisher: Frontiers.
- [16] Douglas M. Fowler and Stanley Fields. Deep mutational scanning: a new style of protein science. *Nature Methods*, 11(8):801–807, August 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3027. URL <https://www.nature.com/articles/nmeth.3027>. Publisher: Nature Publishing Group.

- [17] Takafumi Ueki and Masahito Ohue. Antibody complementarity-determining region design using AlphaFold2 and DDG predictor. *The Journal of Supercomputing*, 80(9):11989–12002, June 2024. ISSN 1573-0484. doi: 10.1007/s11227-023-05887-9. URL <https://doi.org/10.1007/s11227-023-05887-9>.
- [18] Lin Zhu, Yi Fang, Shuting Liu, Hong-Bin Shen, Wesley De Neve, and Xiaoyong Pan. ToxDL 2.0: Protein toxicity prediction using a pretrained language model and graph neural networks. *Computational and Structural Biotechnology Journal*, 27:1538–1549, April 2025. ISSN 2001-0370. doi: 10.1016/j.csbj.2025.04.002. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12018212/>.
- [19] Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Towards Accurate and Efficient Binding Affinity Prediction.
- [20] Sriram Kosuri and George M. Church. Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*, 11(5):499–507, May 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2918. URL <https://www.nature.com/articles/nmeth.2918>. Publisher: Nature Publishing Group.
- [21] Po-E Li, Chien-Chi Lo, Joseph J. Anderson, Karen W. Davenport, Kimberly A. Bishop-Lilly, Yan Xu, Sanaa Ahmed, Shihai Feng, Vishwesh P. Mokashi, and Patrick S.G. Chain. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Research*, 45(1):67–80, 11 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw1027. URL <https://doi.org/10.1093/nar/gkw1027>.
- [22] Mark Edward M Gonzales, Jennifer C Ureta, and Anish M S Shrestha. PHIStruct: improving phage – host interaction prediction at low sequence similarity settings using structure-aware protein embeddings. *Bioinformatics*, 41(1):btaf016, January 2025. ISSN 1367-4811. doi: 10.1093/bioinformatics/btaf016. URL <https://doi.org/10.1093/bioinformatics/btaf016>.
- [23] Wenxing Hu and Masahito Ohue. SpatialPPIv2: Enhancing protein – protein interaction prediction through graph neural networks with protein language models. *Computational and Structural Biotechnology Journal*, 27:508–518, January 2025. ISSN 2001-0370. doi: 10.1016/j.csbj.2025.01.022. URL <https://www.sciencedirect.com/science/article/pii/S2001037025000224>.
- [24] Durga B. Kuchibhatla, Westley A. Sherman, Betty Y. W. Chung, Shelley Cook, Georg Schneider, Birgit Eisenhaber, and David G. Karlin. Powerful sequence similarity search methods and in-depth manual analyses can identify remote homologs in many apparently “orphan” viral proteins. *Journal of Virology*, 88(1):10–20, 2014. doi: 10.1128/jvi.02595-13. URL <https://journals.asm.org/doi/abs/10.1128/jvi.02595-13>.
- [25] Martin Steinegger and Johannes Söding. Mmseqs2: sensitive protein sequence searching for the analysis of massive data sets. *bioRxiv*, 2017. doi: 10.1101/079681. URL <https://www.biorxiv.org/content/early/2017/06/07/079681>.
- [26] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990. ISSN 0022-2836. doi: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2). URL <https://www.sciencedirect.com/science/article/pii/S0022283605803602>.
- [27] J. Rodney Brister, Danso Ako-adjei, Yiming Bao, and Olga Blinkova. Ncbi viral genomes resource. *Nucleic Acids Research*, 43(D1):D571–D577, 11 2014. ISSN 0305-1048. doi: 10.1093/nar/gku1207. URL <https://doi.org/10.1093/nar/gku1207>.
- [28] The UniProt Consortium. The universal protein resource (uniprot) in 2010. *Nucleic Acids Research*, 38(suppl_1): D142–D148, 10 2009. ISSN 0305-1048. doi: 10.1093/nar/gkp846. URL <https://doi.org/10.1093/nar/gkp846>.

- [29] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022. ISSN 1548-7105. doi: 10.1038/s41592-022-01488-1. URL <https://www.nature.com/articles/s41592-022-01488-1>. Publisher: Nature Publishing Group.
- [30] Somaye Hashemifar, Behnam Neyshabur, Aly A Khan, and Jinbo Xu. Predicting protein – protein interactions through sequence-based deep learning. *Bioinformatics*, 34(17):i802–i810, September 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty573. URL <https://doi.org/10.1093/bioinformatics/bty573>.
- [31] Rahul Nikam, Kumar Yugandhar, and M. Michael Gromiha. Deep learning-based method for predicting and classifying the binding affinity of protein-protein complexes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1871(6):140948, November 2023. ISSN 1570-9639. doi: 10.1016/j.bbapap.2023.140948. URL <https://www.sciencedirect.com/science/article/pii/S1570963923000626>.
- [32] Mu Gao and Jeffrey Skolnick. Improved deep learning prediction of antigen-antibody interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 121(41):e2410529121, October 2024. ISSN 1091-6490. doi: 10.1073/pnas.2410529121.
- [33] Nicole N. Thadani, Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Daniel Ritter, Chris Sander, Yarin Gal, and Debora S. Marks. Learning from prepandemic data to forecast viral escape. *Nature*, 622(7984): 818–825, October 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06617-0. URL <https://www.nature.com/articles/s41586-023-06617-0>. Publisher: Nature Publishing Group.
- [34] Melissa S. Maginnis. Virus – receptor interactions: The key to cellular invasion. *Journal of Molecular Biology*, 430(17):2590–2611, 2018. ISSN 0022-2836. doi: <https://doi.org/10.1016/j.jmb.2018.06.024>. URL <https://www.sciencedirect.com/science/article/pii/S0022283618306302>.
- [35] Loïc Lannelongue and Michael Inouye. Pitfalls of machine learning models for protein – protein interaction networks. *Bioinformatics*, 40(2):btac012, 01 2024. ISSN 1367-4811. doi: 10.1093/bioinformatics/btac012. URL <https://doi.org/10.1093/bioinformatics/btac012>.
- [36] Stephen K Burley, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, Ken Dalenberg, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sutapa Ghosh, David S Goodsell, Rachel K Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Tara Kalro, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Periskova, Andreas Prlić, Chris Randle, Alexander Rose, Peter Rose, Raul Sala, Monica Sekharan, Chenghua Shao, Lihua Tan, Yi-Ping Tao, Yana Valasatava, Maria Voigt, John Westbrook, Jesse Woo, Huanwang Yang, Jasmine Young, Marina Zhuravleva, and Christine Zardecki. Rcsb protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(D1):D464–D474, 10 2018. ISSN 0305-1048. doi: 10.1093/nar/gky1004. URL <https://doi.org/10.1093/nar/gky1004>.
- [37] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O’Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w. URL <https://www.nature.com/articles/s41586-024-07487-w>.

- [38] Jonathan Feldman and Jeffrey Skolnick. AF3Complex Yields Improved Structural Predictions of Protein Complexes, March 2025. URL <https://www.biorxiv.org/content/10.1101/2025.02.27.640585v1>. Pages: 2025.02.27.640585 Section: New Results.
- [39] JunJie Wee and Guo-Wei Wei. Benchmarking AlphaFold3’s protein-protein complex accuracy and machine learning prediction reliability for binding free energy changes upon mutation. *ArXiv*, page arXiv:2406.03979v1, June 2024. ISSN 2331-8422. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11177964/>.
- [40] H. Zhao, C. Velez, A. Navarene, A. Saha, J. Feldman, J. Skolnick, D. Murray, and B. Honig. Combining structural modeling and deep learning to calculate the e. coli protein interactome and functional networks. *bioRxiv*, 2025. doi: 10.1101/2025.05.07.652715. URL <https://www.biorxiv.org/content/early/2025/05/12/2025.05.07.652715>.
- [41] Mu Gao, Davi Nakajima An, Jerry M. Parks, and Jeffrey Skolnick. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature Communications*, 13(1):1744, April 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29394-2. URL <https://www.nature.com/articles/s41467-022-29394-2>.
- [42] Guangyu Wang, Xiaohong Liu, Kai Wang, Yuanxu Gao, Gen Li, Daniel T. Baptista-Hon, Xiaohong Helena Yang, Kanmin Xue, Wa Hou Tai, Zeyu Jiang, Linling Cheng, Manson Fok, Johnson Yiu-Nam Lau, Shengyong Yang, Ligong Lu, Ping Zhang, and Kang Zhang. Deep-learning-enabled protein – protein interaction analysis for prediction of SARS-CoV-2 infectivity and variant evolution. *Nature Medicine*, 29(8):2007–2018, August 2023. ISSN 1546-170X. doi: 10.1038/s41591-023-02483-5. URL <https://www.nature.com/articles/s41591-023-02483-5>. Publisher: Nature Publishing Group.
- [43] Daniel Kovtun, Mehmet Akdel, Alexander Goncarencu, Guoqing Zhou, Graham Holt, David Baugher, Dejun Lin, Yusuf Adeshina, Thomas Castiglione, Xiaoyun Wang, Céline Marquet, Matt McPartlon, Tomas Geffner, Gabriele Corso, Hannes Stärk, Zachary Carpenter, Emine Kucukbenli, Michael Bronstein, and Luca Naef. Pinder: The protein interaction dataset and evaluation resource. *bioRxiv*, 2024. doi: 10.1101/2024.07.17.603980. URL <https://www.biorxiv.org/content/early/2024/07/20/2024.07.17.603980>.
- [44] Alief Moulana, Thomas Dupic, Angela M. Phillips, Jeffrey Chang, Serafina Nieves, Anne A. Roffler, Allison J. Greaney, Tyler N. Starr, Jesse D. Bloom, and Michael M. Desai. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nature Communications*, 13(1):7011, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34506-z. URL <https://www.nature.com/articles/s41467-022-34506-z>. Publisher: Nature Publishing Group.
- [45] Rongqing Yuan, Jing Zhang, Jian Zhou, and Qian Cong. Recent progress and future challenges in structure-based protein-protein interaction prediction. *Molecular Therapy*, 33(5):2252–2268, 2025. ISSN 1525-0016. doi: <https://doi.org/10.1016/j.ymthe.2025.04.003>. URL <https://www.sciencedirect.com/science/article/pii/S1525001625002771>.
- [46] Jason Nomburg, Erin E. Doherty, Nathan Price, Daniel Bellieny-Rabelo, Yong K. Zhu, and Jennifer A. Doudna. Birth of protein folds and functions in the virome. *Nature*, 633(8030):710–717, September 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07809-y. URL <https://www.nature.com/articles/s41586-024-07809-y>. Publisher: Nature Publishing Group.
- [47] Tao Huang and Yixue Li. Current progress, challenges, and future perspectives of language models for protein representation and protein design. *The Innovation*, 4(4):100446, 2023. ISSN 2666-6758. doi: <https://doi.org/10.1016/j.xinn.2023.100446>. URL <https://www.sciencedirect.com/science/article/pii/S2666675823000747>.
- [48] Sara Reardon. Five protein-design questions that still challenge AI. *Nature*, 635(8037):246–248, November 2024. doi: 10.1038/d41586-024-03595-9. URL <https://www.nature.com/articles/d41586-024-03595-9>.

Bandiera_abtest: a Cg_type: Technology Feature Publisher: Nature Publishing Group Subject_term: Technology, Computational biology and bioinformatics, Machine learning, Synthetic biology.

- [49] Renan Chaves de Lima, Lucas Sinclair, Ricardo Megger, Magno Alessandro Guedes Maciel, Pedro Fernando da Costa Vasconcelos, and Juarez Antônio Simões Quaresma. Artificial intelligence challenges in the face of biological threats: emerging catastrophic risks for public health. *Frontiers in Artificial Intelligence*, 7, May 2024. ISSN 2624-8212. doi: 10.3389/frai.2024.1382356. URL <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1382356/full>. Publisher: Frontiers.
- [50] Leyma P. De Haro. Biosecurity Risk Assessment for the Use of Artificial Intelligence in Synthetic Biology. *Applied Biosafety: Journal of the American Biological Safety Association*, 29(2):96–107, June 2024. ISSN 1535-6760. doi: 10.1089/apb.2023.0031. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11313549/>.
- [51] Alissa M. Hummer, Constantin Schneider, Lewis Chinery, and Charlotte M. Deane. Investigating the volume and diversity of data needed for generalizable antibody – antigen G prediction. *Nature Computational Science*, pages 1–13, July 2025. ISSN 2662-8457. doi: 10.1038/s43588-025-00823-8. URL <https://www.nature.com/articles/s43588-025-00823-8>. Publisher: Nature Publishing Group.
- [52] Jade Butterworth, David Smerdon, Roy Baumeister, and William von Hippel. Cooperation in the time of covid. *Perspectives on Psychological Science*, 19(4):640–651, 2024. doi: 10.1177/17456916231178719. URL <https://doi.org/10.1177/17456916231178719>. PMID: 37384624.
- [53] Cynthia G. McCormick. Regulatory challenges for new formulations of controlled substances in today’s environment. *Drug and Alcohol Dependence*, 83:S63–S67, 2006. ISSN 0376-8716. doi: <https://doi.org/10.1016/j.drugalcdep.2006.02.001>. URL <https://www.sciencedirect.com/science/article/pii/S0376871606000548>. Drug Formulation and Abuse Liability.
- [54] Leonard V. Sacks, Hala H. Shamsuddin, Yuliya I. Yasinskaya, Khaled Bouri, Michael L. Lanthier, and Rachel E. Sherman. Scientific and regulatory reasons for delay and denial of fda approval of initial applications for new drugs, 2000-2012. *JAMA*, 311(4):378–384, 01 2014. ISSN 0098-7484. doi: 10.1001/jama.2013.282542. URL <https://doi.org/10.1001/jama.2013.282542>.