

Disseminating biological AI models

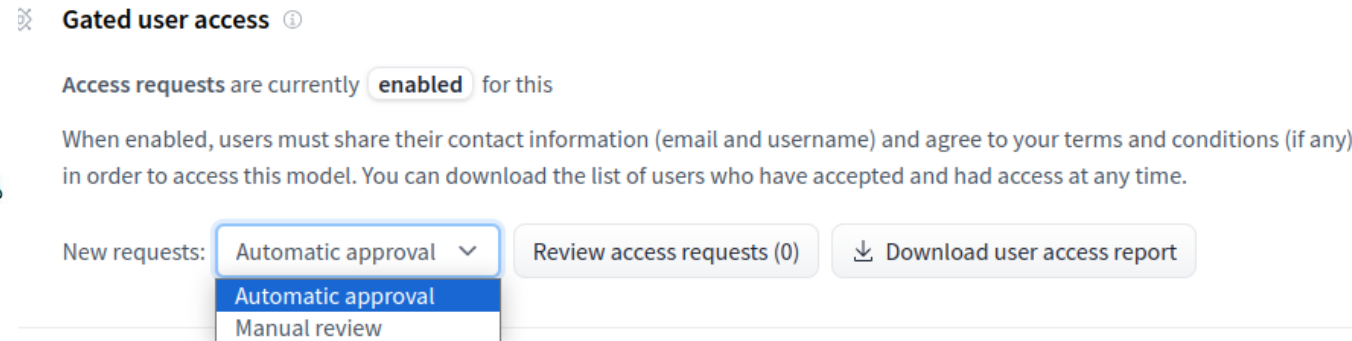
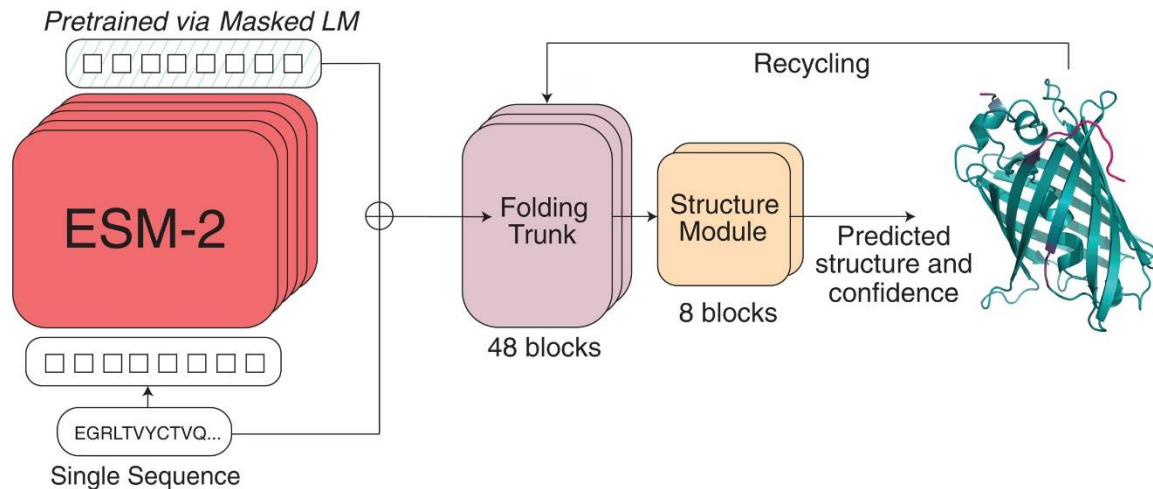
Anthony Gitter

University of Wisconsin-Madison

Morgridge Institute for Research

NASEM Benefits and Risks of Publishing
Studies of In Silico Modeling Workshop

April 3, 2025



NATIONAL ACADEMIES
Sciences
Engineering
Medicine

The Age of AI in the Life Sciences

Benefits and Biosecurity Considerations

Consensus Study Report



Post

Reply



National Academies
@theNASEM



#AI is transforming #biology, enabling rapid innovation — but with progress comes risk.

A new National Academies report examines how AI-driven biological tools could impact #biosecurity, from potential misuse to their role in risk mitigation.

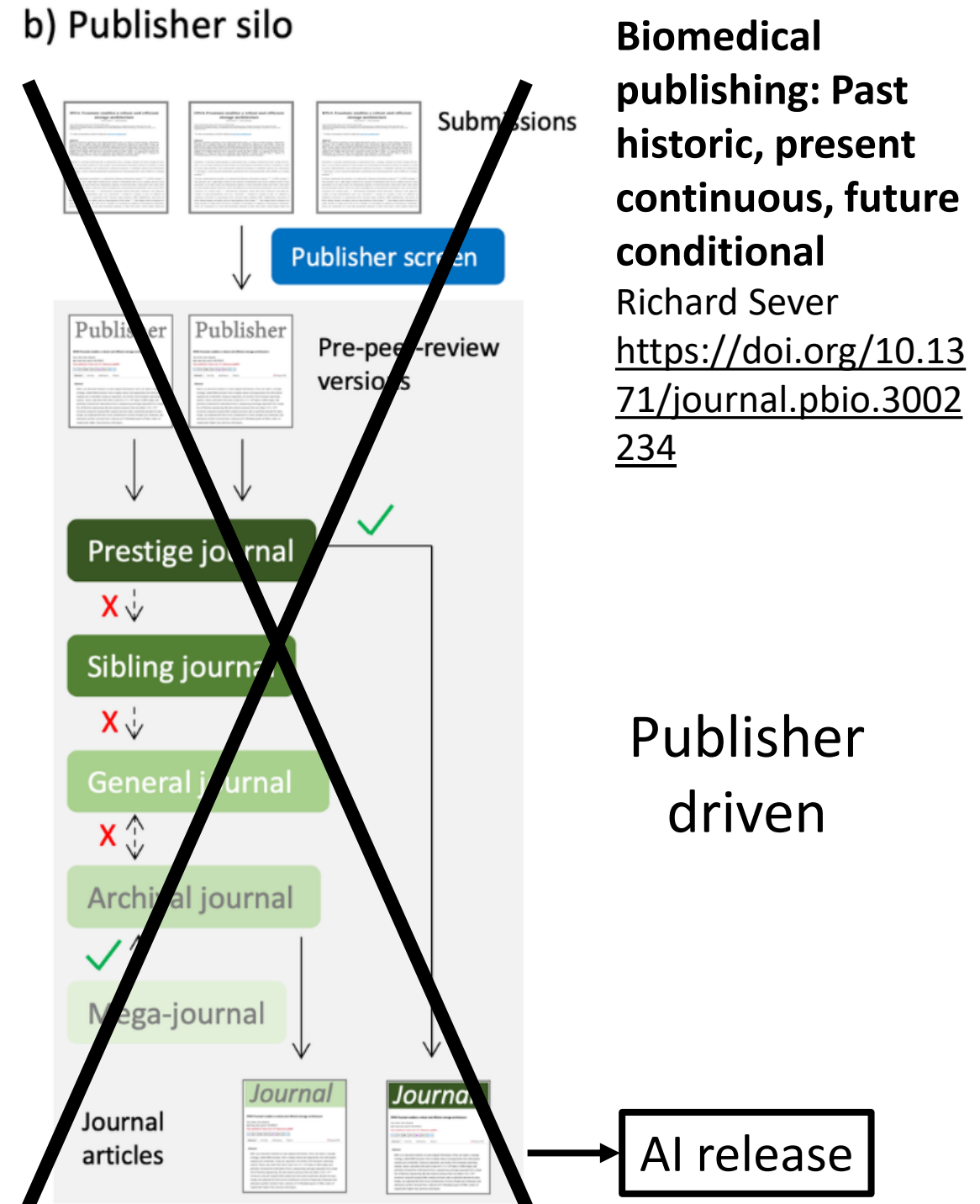
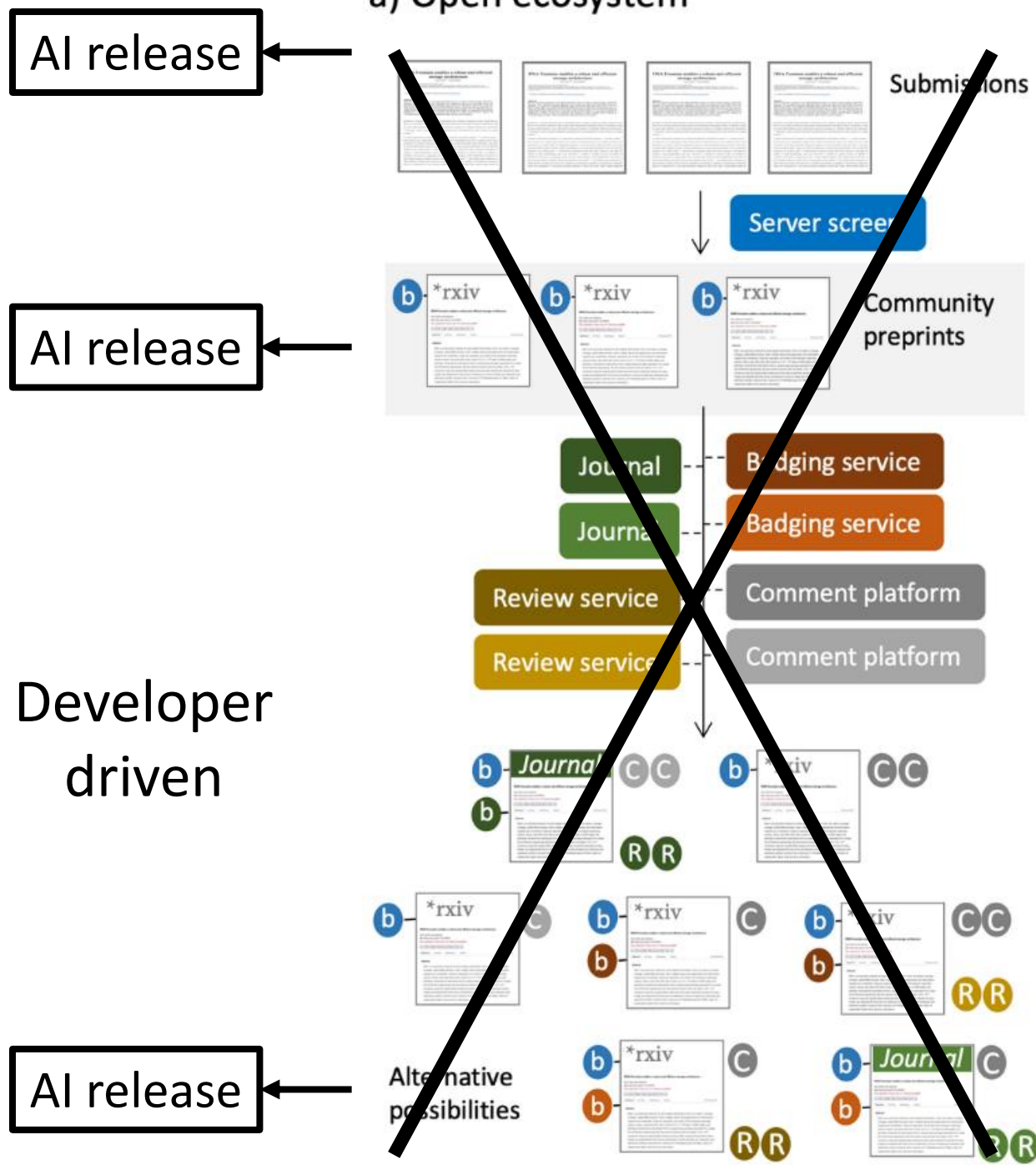
Read: ow.ly/65Bc50Vhxhs



11:02 AM · Mar 14, 2025 · 13.1K Views

128 pages

<https://doi.org/10.17226/28868>



Biomedical publishing: Past historic, present continuous, future conditional
 Richard Sever
<https://doi.org/10.1371/journal.pbio.3002234>

Dissemination outside the publication stream



A multimodal foundation
model for controlling
generation and representation
learning



ESM Cambrian: Revealing
of proteins
using learning

AI developers make
decisions about how,
when, where to
disseminate

Arc Institute

Genome modeling and

vo 2

Pro-1

Michael Hla



Pro-1 is a fully open source reasoning model (8b and 70b) trained using GRPO towards a
physics based reward function for protein stability.

AI release: components and venues

Software

- Code: GitHub, GitLab
- Executable program
- Web server
- Application programming interface
- Cloud services
- Container: DockerHub

Datasets

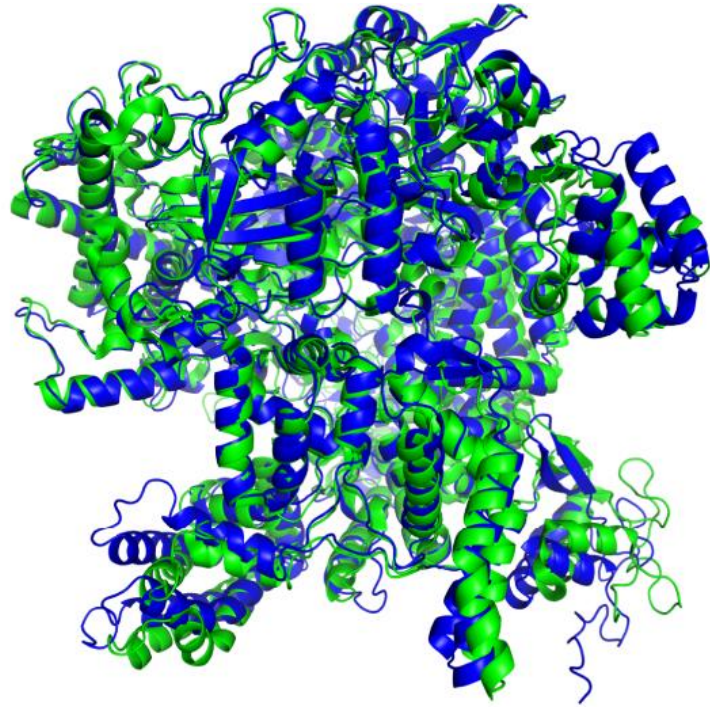
Protein Data Bank,
Sequencing Read Archive,
database of Genotypes
and Phenotypes

Model weights

uggingFace
enodo, figshare, etc.
oogle Drive
itHub, GitLab
ustom websites
orrents (!)

**Dissemination options
are broad even if
some are more
popular**

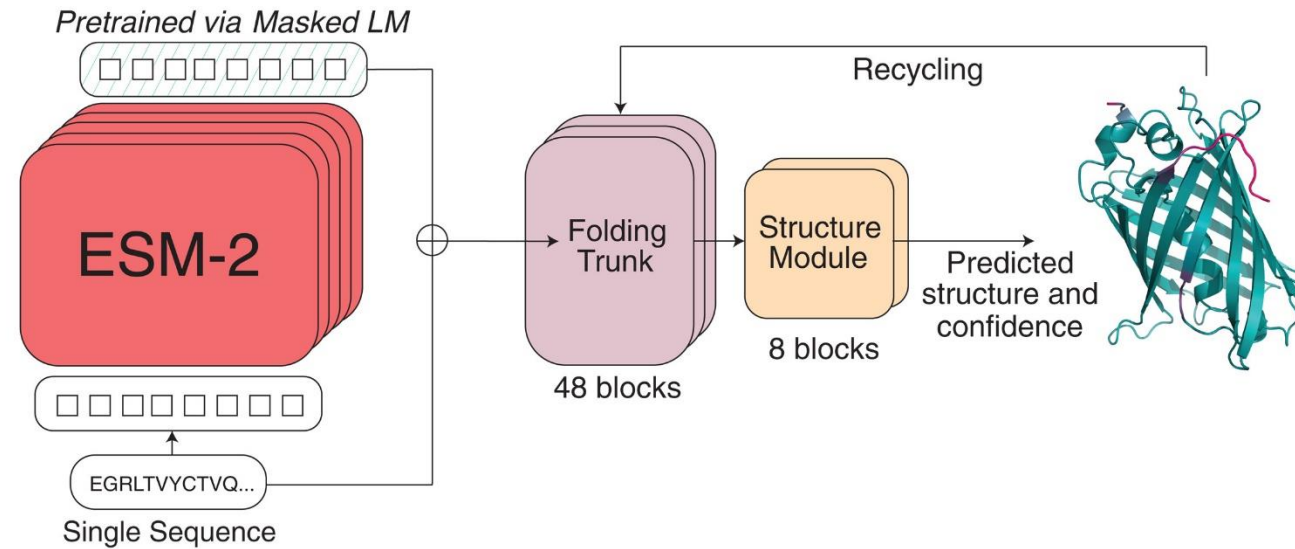
Open science to realize the benefits of biological AI



AlphaFold Experiment
r.m.s.d.₉₅ = 2.2 Å; TM-score = 0.96

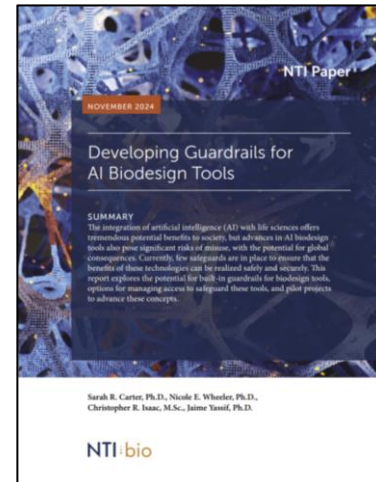
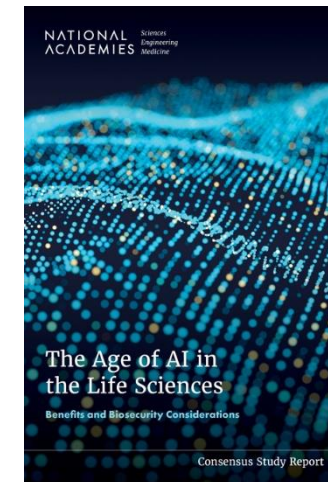
AlphaFold2

Jumper 2021 *Nature* doi:10.1038/s41586-021-03819-2



ESM-2

Lin 2023 *Science* 10.1126/science.ade2574



A renewed call for open artificial intelligence in biomedicine

<https://doi.org/10.31219/osf.io/2xh3w>

AUTHORS

Anthony Gitter, James S. Fraser, Tamir Gonen, Robert Patro, Hannah K. Wayment-Steele, Alicia Williams, Benjamin Haibe-Kains, Roland L. Dunbrack, Cameron Cook, Anshul Kundaje, and 4 more ▼

Importance of open science by default unless specific biosecurity risks are identified

Controlled-access repositories are then suitable for sharing AI data and models

“If a hypothetical AI model carried such great risk that it could not be shared even through a controlled-access repository, **researchers must anticipate these risks in advance and not execute research where the harms outweigh the benefits.**”

Intertwining biosecurity risks and commercial interests is detrimental to the assessment of biosecurity risks

Opinions on controlled-access

There are some existing resources

HuggingFace gated models and gated datasets are quite good

Valuable to more
systematically state
desired features and
assess existing repositories

Synapse supports ~~gated models, gated datasets, and gated challenges~~



Zenodo has restricted records but lack desired features

Public Restricted

The files of this record are restricted.

 **Public with restricted files**
The record is publicly accessible. The files can only be accessed by users specified in the permissions.

Opinions on controlled-access

- Importance of equitable access and persistence
- Best practices in open source require little to no active maintenance
- Many options for controlled-access will require more active maintenance

Thank you