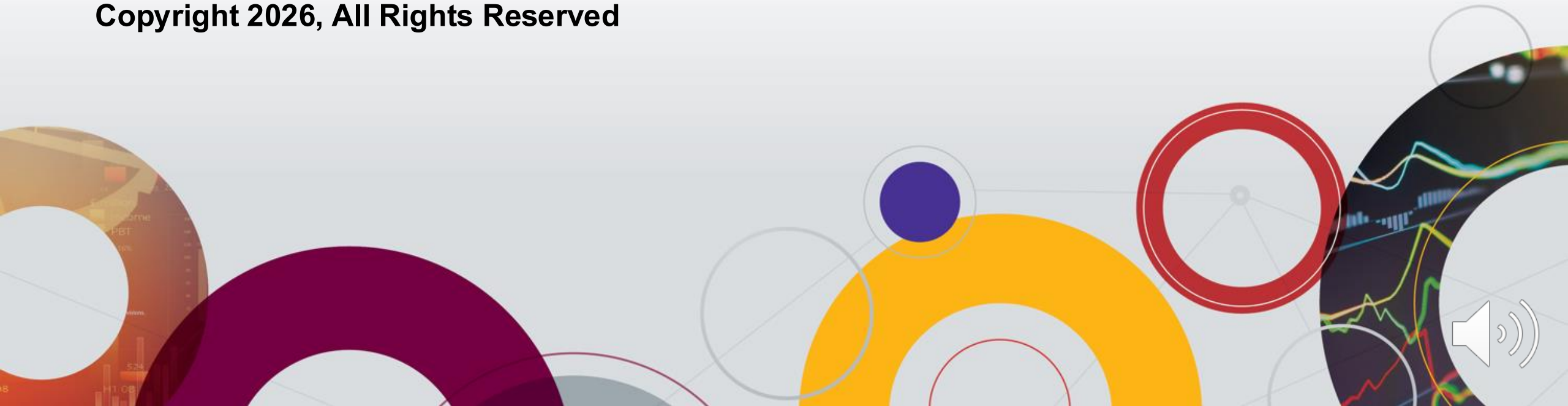


8.1b Process, Information, and Challenges

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



Open-access (OA) publishing

Scientists perform research, usually publicly (taxpayer) funded and then turn over copyright of report to publishers, limiting access (Frank, 2013; Wolpert, 2013)

Growing questions over what traditional publishing process adds to science publication, especially in era of Web (Van Noorden, 2012)

- Online-only costs lower cost of typesetting and printing
- OA publishing removes other costs

OA has caught on, with over half of all biomedical papers now published under model, but still concerns for fairness and sustainability (Brainard, 2024)

OA – definition

Publishing model where content is freely available and costs covered by some other means

- Rationale is that Web provides means to do so
- Usually “allowable” expenses on research grants, and estimated to cost just a fraction of typical research grant, e.g., 0.3% of research costs (Zerhouni, 2004)

Various models for OA publishing (Mager, 2022)

- Gold – “author pays” model, i.e., research funding cover costs
- Diamond – same as gold but someone other than author pays
- Green – author required to deposit manuscript in public repository, e.g., PubMed Central (PMC)
- Hybrid – model of many journals with some OA and others available only by subscription

OA – recent developments

Accumulating evidence shows value of OA, leading to advocacy for increased use (Spedding, 2016)

[Calls for better public access to research in US](#) as well

Has also been development of journals with business models built on OA

OA considered by some to be part of larger “open science” (National Academies, 2018), i.e.,

- Open data – all data collected in research
- Open source – all software code developed and used
- Open methodology – clear and detailed description; availability of all surveys and other tools
- Open peer review – all comments of peer reviewers

Some open access and related publishing initiatives

BioMed Central (BMC)

- Original family of gold OA journals, now over 200, including BMC Medical Informatics & Decision-Making, BMC Medicine, etc.
- Acquired by commercial publisher, Springer, who has maintained OA model
- Publishing charges vary by journal, from \$2-4K

Public Library of Science (PLoS)

- Smaller number of journals but higher profile, e.g., PLoS Medicine

Increasing numbers of journals (e.g., Springer, Oxford including JAMIA) have adopted hybrid approaches

- Also adding gold OA companions to regular journals, e.g., JAMA Network Open, JAMIA Open, etc.

PubMed Central (PMC) (Reznick, 2025)

Initiative of NIH for electronic distribution of life sciences research reports

- Not to be confused with PubMed (MEDLINE)

NIH has established barrier-free Web-based repository

- Searchable directly or via PubMed

Over 4000 journals participate fully or partially (e.g., BMC, BMJ) but many major journals do not (e.g., JAMA, NEJM)

NIH has also adopted policy that all authors of papers reporting research funded by NIH grants must submit article or last version prior to publication

- Resulting in additional journals “participating”

Concerns about OA publishing

Growth of “predatory” OA journals that offer inexpensive publishing and appointment to editorial boards, which exist mostly to make money (Haug, 2013; Xia, 2021; Khan, 2022)

- Some authors have exposed process, e.g., McCool (2017) – leading to calls for efforts to stop, e.g., (Talley, 2024; Laine, 2025)
- [Beall's List](#) of potential predatory journals and publishers

Has led to “corporate capture” by large publishers (Meagher, 2021), resulting

- Increasing page charges (Alperin, 2022)
- Systematic review of OA page charges found prominent use of gold (“author pays”) model and limited use of waivers (Borrego, 2023)
- “Read-and-publish” (Kember, 2023)

Challenges for knowledge-based information

Peer review

Misinformation and disinformation

Misconduct

Retractions

Publication bias

Peer review

Bedrock of scientific process ... or is it?

- True “peers” difficult to find and may not have needed time or expertise (Mastroianni, 2022)
- Affected by author prominence – blinding may help (Huber, 2022)

Growing numbers of submissions in face of researcher time constraints taxing system

- Estimated cost of \$1272/reviewer-year for >\$6B globally (LeBlanc, 2023)
- Should reviewers be paid, especially in light of profits of publishers (Cheah, 2023)?

Growing use of AI by peer reviewers with unknown extent or impact on quality (Cheng, 2024; Naddaf, 2025)

- 6.5-16.9% of text of peer reviews for AI-related conferences from LLMs (Liang, 2024)

New approaches needed? – posting preprints and facilitating open review (Dance, 2023)?

Misinformation and disinformation

Long a challenge in biomedicine but exacerbated in 21st century by (van der Linden, 2022)

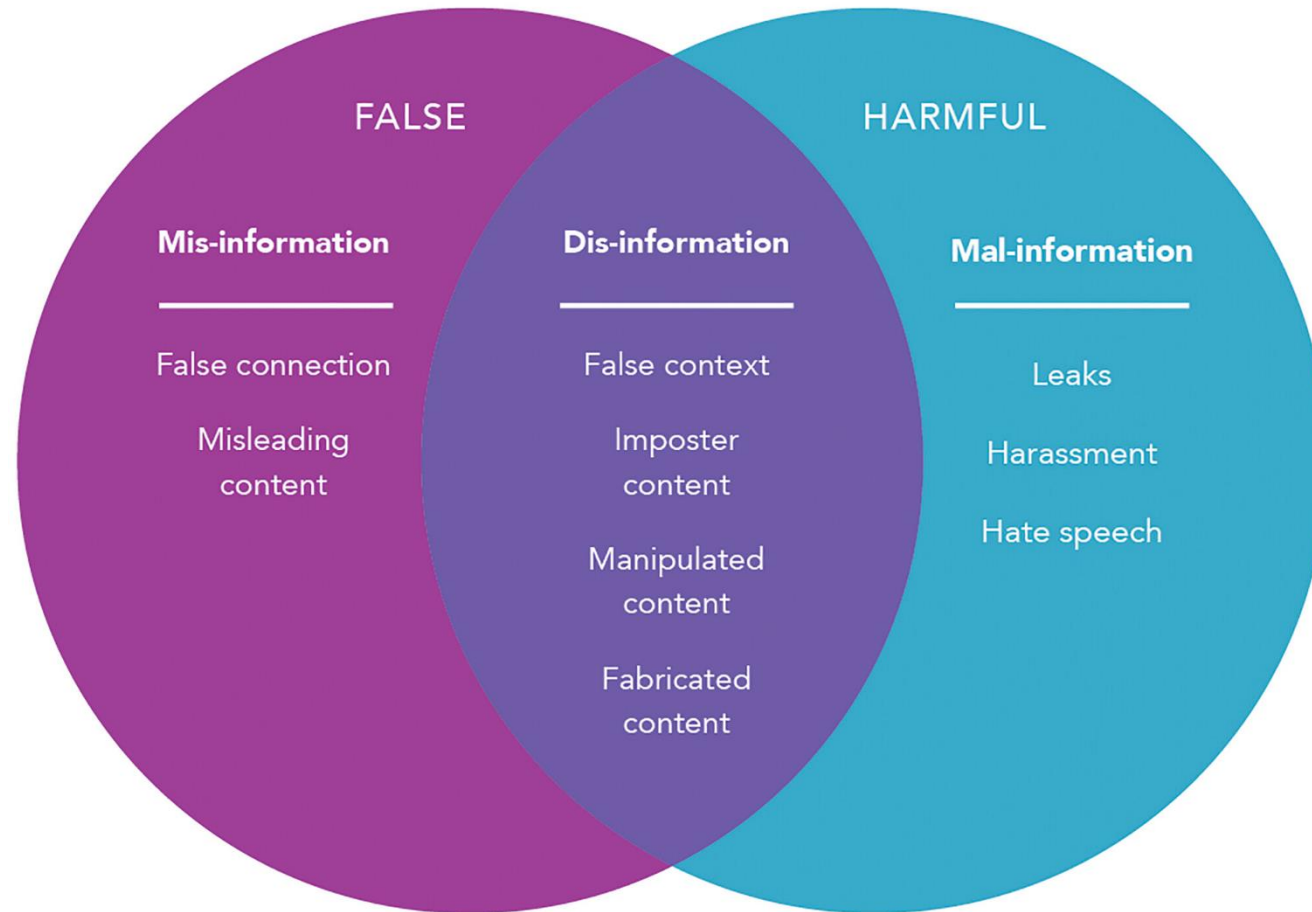
- Democratization of Internet and Web
- Interactive social media (Robertson, 2024), fueled by incentive to spread outrage online (McLaughlin, 2024)
- Global public health crises, e.g., COVID-19 (Vera, 2022)
- Political polarization (Levin, 2023; Perlis, 2023)

Long-standing methods for communicating science may contribute to problem (Boyle, 2022)

- Deliberate pace of science
- Literature more focused on communication with other scientists than public

Misinformation definitions (Waldrop, 2023)

Disinformation and misinformation has led to infodemic



Misinformation has long been challenge for Web

Early systematic review of Web information quality studies found problems in many clinical areas (Eysenbach, 2002)

- Self-care expert Ferguson (2002) retorted that savvy consumers search multiple sites and are able to better discern quality than we might believe
- Varies by topic but continues to be a problem for Web information (Sbaffi, 2017)

Problem not limited to consumer information – journals engage in “spin” about what they publish (Yavchitz, 2012)

US hospital Web sites may tout benefits more than risks of procedures, e.g., for transcatheter aortic valve replacement only one-quarter reported any known risks (Kincaid, 2015)

Came to head in COVID-19 pandemic

Modern communications led early in pandemic to

- “Toxic legacy of poor-quality research, media hype, lax regulatory oversight, and vicious partisanship” (Lenzer, 2020)
- Proliferation of pseudoscience (Caulfield, 2020) and conspiracy theories (Allen, 2020; Neil, 2020)
- Misuse of real-world data (Dolgin, 2020)

Exacerbated by some advances in open science (Smalley, 2021), including preprints (Majumder, 2020; Flanagan, 2020; Fraser, 2021)

Spread by small number of people – 12 responsible for almost two-thirds of anti-vaccine content on social media (Center for Countering Digital Hate, 2021)

How do we balance free speech with professional responsibility (Rubin, 2022; Sage, 2022)?

Misinformation and disinformation (cont.)

Search engines can be exploited by “data voids” (Boyd, 2019)

- Breaking news – from legitimate sites takes longer, may also be blocked by paywalls
- Strategic new terms – can be amplified
- Outdated terms – still used to maintain information availability

Even more exacerbated in modern times by social media, e.g., X/Twitter and Facebook (Meserole, 2018; Khullar, 2022)

- Proliferated by automated means, e.g., bots (Himelein-Wachowiak, 2021)
- May be countered with Community Notes on X (Allen, 2024)

Searching to counter misinformation may lead to increasing it if searchers fall into data voids (Aslett, 2024)

Opportunities for countering misinformation

Need means to support countering by healthcare and public health professionals (Bautista, 2021; Arora, 2022; Sharfstein, 2022)

Should there be sanctions for physicians, e.g., via medical boards or other authoritative bodies, who spread (Baron, 2022; Baron, 2023)?

Public policy advocacy, e.g., (Sell, Johns Hopkins Center for Health Security, 2021)

- Intervene against false and damaging content as well as sources propagating it
- Promote and ensure abundant presence and dissemination of factual information
- Increase public's resilience to misinformation and disinformation
- Ensure whole-of-nation response through multisector and multiagency collaboration

Suggested criteria for evaluating quality of Web information

Most widely known came from Silberg, 1997

Cannot judge “truth” but can assess other criteria

- Authorship – names, affiliations, and credentials
- Attribution – references, sources, and (where appropriate) copyright
- Disclosure – potential and real conflicts of interest
- Currency – dates content posted and updated

Many new approaches over the years but still many challenges (Sun, 2019)

- Important approach is to steer to known trustworthy sites (Oxman, 2019)

Misconduct in science

History over 40 years but exacerbated by recent technology and pandemic (Flier, 2021)

Honor system of science allows fraud to flourish, especially when there are incentives (Economist, 2023)

Manipulation of images and references (Mehregan, 2022); may get worse with ability of AI to manipulate images (Gu, 2022)

“Zombie science” that appears in literature but does not answer scientific questions (Pasternak, 2022)

Non-disclosure of conflict of interest (Baraldi, 2022; VanDeMark, 2022)

May be facilitated by generative AI, e.g., “formulaic” papers using public datasets such as NHANES (Suchak, 2025)

Downstream problems of misconduct

Large numbers of publications put “strain” on publishing enterprise (Hanson, 2024)

- Including on preprint servers, e.g., ArXiv, leading site to put restrictions on who can post (Jones, 2026)

Flawed papers

- Cited by others (Suelzer, 2019; Wadhaw, 2021; Marcus, 2022; Peterson, 2022; Cabenac, 2024)
- Included in systematic reviews and meta-analyses (Kataoka, 2022; Wilkinson, 2024; Graña Possamai, 2025)

Paper mills and fake papers

Paper mills sell fabricated data, papers, and more (Van Noorden, 2023; Parker, 2024; Abalkina, 2025)

- Fueled in part by pressure on academics to publish (Ro, 2025), especially early-career researchers (Tomlinson, 2023)
- “Entities enabling scientific fraud at scale are large, resilient, and growing rapidly” (Richardson, 2025)

Peer review fraud (Day, 2022)

Up to 20% of articles published in journals indexed in PubMed may contain faked data produced by paper mills (Sabel, 2023)

Has resulted in growing number of retractions of scientific papers

Reasons for retractions (Koo, 2024)

[Retraction Watch](#) maintains (Oransky, 2022)

- Database of retracted papers
- List of “hijacked” journals
- “Leaderboard” of most retracted authors

NLM and journals have means to designate on Web but can be difficult to remove

- May be harder in era of generative AI with instances of ChatGPT referencing retracted papers (Thelwall, 2025)

Retraction Reason	Count (%)
Data and Results Issues	6160 (28.8)
Plagiarism and Duplication	4085 (19.1)
Investigations and Findings	3588 (16.8)
Authorship and Ethical Concerns	1918 (9.0)
Misconduct and Fraud	1288 (6.0)
Image Manipulation and Fabrication	1092 (5.1)
Peer Review and Editorial Issues	985 (4.6)
Withdrawal and Retraction Notices	763 (3.6)
Institutional and Policy Issues	643 (3.0)
Complaints and Objections	418 (2.0)
Miscellaneous	412 (1.9)
Procedural and Legal Issues	70 (0.3)
Total	21422 (100)

Possible solutions going forward

Publishers must be vigilant, e.g., large open-access publisher Hindawi (Flintoft, 2023)

[PubPeer](#) is online forum devoted to concerns about integrity of published research (Chawla, 2024)

- <https://pubpeer.com/>
- Browser plug-in alerts for studies listed in PubPeer

Systematic reviewers must identify and eliminate them from scientific literature (Bero, 2025)

[Collection of Open Science Integrity Guides](#)

Publication bias

Studies with positive results more likely to be published (Dwan, 2013)

Originally called “file drawer” problem (Rosenthal, 1979)

Failure to report science is a form of misconduct (Chalmers, 1990)

Even a problem in informatics (Vawdrey, 2013)

Still persists in many biomedical fields (Rivero-de-Aguilar, 2023)