

AI Principles

April 2023

The News/Media Alliance (NMA) represents the most trusted publishers in print and digital media based in the United States, from small, local outlets to national and international publications read around the world.

Every day, these publishers invest in producing high-quality creative content that is engaging, informative, trustworthy, accurate and reliable. In doing so, they not only make significant economic contributions, but they also play a crucial role in educating, upskilling and informing our communities, building our democracy and economy, and furthering America's economic, security and political interests abroad.

Introduction

As generative artificial intelligence (GAI) technologies become more prevalent, our membership believes these new tools must only be developed respecting journalistic and creative content, in accordance with principles that protect publishers' intellectual property (IP), brands, reader relationships, and investments. **The unlicensed use of content created by our companies and journalists by GAI systems is an intellectual property infringement: GAI systems are using proprietary content without permission.** It's also critical to acknowledge the societal risks associated with the proliferation of mis- and dis-information through GAI, which high-quality, original content, produced by skilled humans and trusted brands, can help to combat. GAI developers and deployers must negotiate with publishers for the right to use their content in any of the following manners:

- Training: Including publishers' content in datasets and using it for GAI system training and testing.
- Surfacing: The serving of publishers' content in response to user inputs, possibly including a cover note generated by the GAI system of what is contained in the surfaced content.
- Synthesizing: Summaries, explanations, analyses etc. of source content in response to a query.

This document highlights the overarching principles that must guide the development and use of GAI systems as well as the policies and regulations governing them. These principles are founded on our understanding of these systems and technologies as they are currently used – and may therefore be amended as these technologies and uses develop – and apply equally to all publisher content, whether in text, image, audiovisual or any other format.

Principles

INTELLECTUAL PROPERTY

Developers and deployers of GAI must respect creators' rights to their content. These rights include copyright and all other legal protections afforded to content creators and owners, as well as contractual restrictions or limitations imposed by publishers for the access and use of their content (including through their on-line terms of service). Developers and deployers of GAI systems—as well as legislators, regulators and other parties involved in drafting laws and policies regarding GAI—must maintain an unwavering respect for these rights and recognize the value of creators' proprietary content. ***GAI developers and deployers should not use publisher IP without permission, and publishers should have the right to negotiate for fair compensation for use of their IP by these developers.*** Professional journalism is particularly valuable due to its reliability, accuracy, coherency and timeliness, enhancing GAI system outputs and improving perceptions of system quality. Absent permission and specific licenses, GAI systems are not simply using publishers' content, they are stealing it.



Use of publishers' IP requires explicit permission. Use of publisher content by GAI systems for training, surfacing and synthesizing is not authorized by most publishers' terms and conditions, and authorization for search should not be construed as an authorization for uses such as training GAI systems or displaying more content than contemplated for or as used in traditional search. GAI system developers and deployers should not be crawling, ingesting or using publishers' proprietary content without express authorization; requiring publishers to opt out is not acceptable. ***Negotiating written, formal agreements is therefore necessary.*** Industry standards should be developed to allow for automatic detection of permissions that distinguish among potential uses of crawled or scraped content. These standards and usage agreements can also address other issues such as attribution, monetization, responsibility, and derivative uses.

Compensation agreements must account for harms GAI systems may cause publishers and the public. GAI system surfacing and synthesizing are providing much more proprietary content and information from the original sources than traditional search and often provide little or no attribution, and will exacerbate the growing trend toward zero-click, reducing or even eliminating value for publishers. GAI systems use publishers' proprietary content to generate outputs that may replace their role in the consumer/information provider relationship. In addition to reducing traffic, this harms publisher brands that have taken years, decades, or even centuries to build.

Copyright laws must protect, not harm, content creators. The fair use doctrine does not justify the unauthorized use of publisher content, archives and databases for and by GAI systems. Any previous or existing use of such content without express permission is a violation of copyright law. The Section 1201 triennial rulemaking process should not be used to allow for the bypassing of content protections for GAI development purposes. Exceptions to copyright protections for text and data mining (TDM) should be narrowly tailored to limited nonprofit and research purposes that do not damage publishers or become pathways for unauthorized uses that would otherwise require permission. The U.S. also has made international law commitments in this area that protect its IP-based businesses across multiple sectors and these must be upheld in its approach to AI.

There is an existing market for licensing publishers' news content. Valuing publishers' legitimate IP interests need not impede GAI innovation because compensation frameworks (for example, licensing) already exist to permit use in return for payment. GAI innovation should not come at the expense of publishers, but rather at the expense of developers and deployers. Publishers encourage the use of efficient ways to license through standard-setting organizations that can facilitate efficient training of GAI systems.

TRANSPARENCY

GAI systems should be transparent to publishers. Publishers have a right to know who copied our content and what they are using it for. We call for strong regulations and policies imposing transparency requirements to the extent necessary for publishers to enforce their rights. Publishers have a legitimate interest in determining what content of theirs has been and is used in GAI systems. Using datasets or applications developed by non-profit, research, or educational third parties to power commercial GAI systems must be clearly disclosed and not used to evade transparency obligations or copyright liability.

GAI systems should be transparent to users. Direct relationships between users and publishers are critical for the sustainability of the news media and informational content sector. Surfaced and synthesized outputs should connect, not disintermediate, users with publishers. Members of the public should know the source of information that may affect them. Generative outputs should include clear and prominent attributions in a way that identifies to users the original sources of the output and encourages users to easily and directly navigate to those products, as well as to let them know when content is generated by GAI. Transparency into GAI systems can also help prevent misuse and the spread of mis- and dis-information. Similarly, it enables the evaluation of GAI systems for unintended bias to avoid discriminatory outcomes.



ACCOUNTABILITY

Deployers of GAI systems should be held accountable for system outputs. GAI systems pose risks for competition, the integrity of news and creative content, and for public trust in the journalistic and creative content. This is aggravated by the ability of AI applications to devalue publisher brands by generating content that attributes false or inaccurate information to publishers who have not published the information and who have processes in place to prevent such publication in the first place. Accordingly, deployers of GAI systems should not be shielded from liability for their outputs—to do so would be to provide deployers of GAI systems with an unfair advantage against which traditional publishers cannot compete and increase the danger to the public and institutions from the unchecked power of this technology.

FAIRNESS

GAI systems should not create, or risk creating, unfair market or competition outcomes. Regulators should be attuned to ensuring GAI systems are designed, deployed, and used in a way that is compliant with competition laws and principles. Developers and deployers should also use their best efforts to ensure that GAI models are not used for anti-competitive purposes. The use of publisher content for GAI purposes without express permission from content owners by firms that have market power in online content distribution should be considered evidence of a violation of competition laws. Regulators should be vigilant for other anti-competitive uses of GAI systems.

SAFETY

GAI systems should be safe and avoid privacy risks. GAI systems, including GAI models, should be designed to respect the privacy of users who interact with them. Early indications are that GAI tools will exacerbate trends towards digital platforms collecting large volumes of user data. The collection and use of personal data in GAI system design, training and use should be minimal and should be disclosed to users in an easily understandable manner so that users can make informed judgments about how their data is used in exchange for the GAI service. Users should be informed about, and should have the right to prevent, the use of their interactions with GAI systems for the purposes of training or collection of personal data. Systems should also be designed in a way that means paywalled and otherwise protected content cannot be exposed (including but not limited to, for example, by membership inference methods).

DESIGN

All of the principles discussed above should be incorporated in the very design of GAI systems, as significant elements of the design, and not considered as an afterthought or a minor concern to be addressed when convenient or when a third party brings a claim.