International Center for Journalists AI Use Policy

Table of Contents

1. Overview

Journalists across the globe are scrambling to figure out how generative AI tools will impact their workplaces and communities. Whether it be used to hasten the spread of dis/misinformation or become a powerful research tool to uncover the truth, journalists need to know its implications. In order for the International Center for Journalists (ICFJ) to support the journalists in their efforts to adapt to AI, it first needs to create a policy that outlines how the organization will internally use generative AI.

2. Purpose

The purpose of this policy is to establish a standard for how ICFJ uses, adopts and engages with AI tools – especially when it comes to personal identification information.

3. Scope

The scope of this policy encompasses the internal use of AI tools and systems within the organization. This includes but is not limited to: reports to funders, internal memos, messages to donors, public-facing communications, and analyzing program data. This policy applies to all employees, interns, consultants, contractors, and any other individuals or entities working on behalf of or within ICFJ.

4. Policy

4.1 Best Practices

The successful use of AI within our organization hinges upon responsible use and human oversight. We recognize that some accessible generative AI tools are built using data obtained without consent, and that generative AI tools may produce content that is biased, inaccurate, and/or infringing of third party intellectual property rights. that have been evaluated by the Technology Governance and Information Security teams and that have been designed to reduce the risk of infringement and other harms. AI capabilities and risks may soon outpace the specific cases outlined below, but we should always abide with our core principles:

- **Do no harm**
- **Protect others' rights, privacy, and original work**
- **Use content or data with consent**

It is also important to keep in mind that content created by generative AI tools may not be protectable by copyright laws, which may prevent ICFJ from obtaining a copyright registration for such content and/or make it difficult for ICFJ to prevent third parties from using that content. It is therefore important to be transparent about your use of generative AI content, as outline in section 4.2.1 below.

By following the best practices outlined below, we can use AI to enhance our operations, all while upholding our commitment to the protection of data privacy and our organization's values.

The organization can use AI and generative AI tools:

- To generate draft text and ideas for proposals, programs, social media posts, fundraising campaigns or stories published on either ICFJ's or IJNet's websites.
- To automate repetitive administrative tasks
- As a data analytics tool to gain insights on program specific data or organizational data
- To create slide presentations
- To create images or illustrations using our own content or open source content. (Adobe Firefly is an example of a consentful image generating tool.)

The organization will not:

- Use material solely generated and edited by AI without a staffer editing and verifying that it is accurate, transparent and free from bias
- Use text generated from AI without being first edited by an ICFJ employee
- **Upload any participant, applicant or donor personal identifiable information (PII) into natural language processing tools or systems stored outside of ICFJ's proprietary database.** PII is any data that can be used to identify someone such as their name, address, phone number, passport id, and social security numbers

4.2 Ethical Deployment

The ethical use of AI tools and systems within the organization is paramount. To ensure that such tools are used responsibility, the organization's staff will adhere to the guardrails outlined within this section.

4.2.1 Transparency

ICFJ will be transparent about its uses of AI and explain the goals or objectives associated with those uses by following these guidelines:

- **Internally Inform Staff Members:** Be transparent about your use of AI with other staff members. Staff are to disclose any work products or other efforts that have used AI in a way that might be perceived as their original work. You must fully fact-check and edit any product using AI-generated material.
- **Producing Public-Facing Communications:** ICFJ staffers will not present published content (articles, social media posts, newsletters, program reports)that have been created with AI tools to the public in a way that purports to be their original work without fully fact-checking, editing and customizing the material.
- **Data Transparency with AI tools:** When the organization is storing, collecting or using participant or donor data within AI tools, it will communicate to the relevant stakeholders how the data is being used and protected.

4.2.2 Accuracy and Verification

AI tools may be used to improve our content through automated editing and filtering tools or other appropriate methods. However, until further notice, we recommend individuals, as defined in Section 3,

do not use generative-AI tools for fact-checking because of the tools' propensity to hallucinate and fabricate. Anytime AI tools are used, the information derived from AI should be cross-verified for accuracy and reliability by a human.

4.2.3 Bias and Fairness

The organization should be aware of and work to mitigate potential biases towards historically marginalized communities that are inherent within AI tools. This includes biases in the algorithms themselves and in the data used to train them. Individuals will adhere to the organization's Diversity, Equity and Inclusion Policy when working with AI tools.

4.2.4 Intention of Use

The use of AI tools and systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses. These harms include safety risks that make the organization vulnerable to attack, exposing data or information in a way that may become a security risk, awareness of bias inherent in AI models, privacy protection, and protection of intellectual property rights, including copyright and trademark. Special care should be taken not to over-rely on AI tools in hiring processes, employee performance evaluations, program selection, and recruitment, which are areas that demand human judgment and context awareness. ICFJ staffers are accountable for all material that they produce, even when incorporating AI tools.

4.2.5 Risk Management Assessment Approach

The approach outlined below has been designed to help staff mitigate risks associated with using AI tools.

- **Address:** Before the decision is made to use an AI tool within your work, first identify the harms this tool could have to ICFJ and its participants. These harms could be but are not limited to:
    - Opening a security breach within the organization's network
    - Inputting participant's personal identifiable information into a database that is not controlled by the organization
    - Skewing organization materials towards bias or discrimination against marginalized populations
- **Validate:** After identifying the potential harms an AI tool may have, ensure that the tool you'll be using is performing consistently as intended and adheres to this policy's best practices and ethical standards.
- **Deploy:** The decision to deploy an AI tool should only be done after the tool has been identified as safe and responsible.
- **Evaluate:** Staff will regularly evaluate the performance, benefits, and potential drawbacks of the AI tools they deploy to ensure they are used effectively and responsibly.

4.2.6 Program Application Reviewal Process

Final selection(s) of program participants, reporting grantees, fellowships and award recipients will be made by humans. There will be no exceptions.

4.4 Security

Individuals associated with ICFJ, as defined in section 3, are expected to adhere to the following security best practices when using AI tools:

- **Evaluation of AI tools: Before individuals begin using a new AI tool they first must submit a request to the Technology and Data Operations Manager (TDOM).** The request must include how the individual intends to use the tool, what type of information the tool will link to the tool's primary website. The TDOM will evaluate the security of the AI tool by reviewing the tool's security features, terms of service, and privacy policy. They will also check the reputation of the tool developer and any third-party services used by the tool. If the tool is approved, staff will be able to use it.
- **Review Terms of Use and Agreements:** The terms of use and any contractual agreements governing use of the generative AI tools should be carefully reviewed, including by legal counsel as appropriate. In particular, review terms and agreements to determine whether ICFJ data, prompts, or outputs will be used to train the AI models, and whether the service provider offers any contractual protections, such as indemnity against third party intellectual property infringement claims.
- **Protection of confidential data:** Individuals will not upload any participant, applicant or donor personal identifiable information into natural language processing tools.
- **Compliance with security policies:** Individuals will adhere to security best practices outlined in the organization's Information Security Policies.

5. Training and Collaboration:

Staff members should be adequately trained in the use of AI tools, and there should be human oversight of AI to prevent misuse and to ensure that the technology is serving its intended purpose.

5.1 Training

The Technology Governance team in accordance with the Information Security team will periodically prepare training materials on how to responsibly use AI tools on an as needed basis. This can be but not limited to sessions on:

- How to use generative AI tools
- Security risks AI presents to the organization
- Best practices on how to incorporate AI tools within your workflow
- Understanding the potential harms of AI
- Understanding the intellectual property implications of using generative AI tools.

5.2 Collaboration and Partnership

ICFJ will continually engage with the journalism community and other relevant stakeholders to address the challenges related to the use of AI in journalism.

7. Policy Compliance:

7.1 Compliance Measurement

The Infosec team will verify compliance to this policy through various methods, including but not limited to, business tool reports, internal and external audits. Additionally, the Infosec team will periodically revisit the policy with relevant stakeholders to ensure that the policy remains effective in achieving its intended objectives.

7.2 Exceptions

Any exception to the policy must be approved by the Infosec team in advance.

7.3 Non-Compliance

An employee found to have violated this policy may be subject to disciplinary action, up to and including termination of employment.

7.4 Review

This policy will be reviewed on a quarterly basis by the Technology Governance team. Updates will be communicated to all users and appropriate training will be provided.