

Transparency FTW! LLMs, OpenBoxScience and GetFreeCopy

Scott Edmunds 

Published May 30, 2024

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/68nbe-jsx12>

Keywords

Open Access, Education, LLMs, Machine Learning
Feature Image

Copyright

Copyright © Scott Edmunds 2024. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

As an Open Science publisher we've pushed for transparency and access in the research that we disseminate, and in GigaByte journal we've just [published a new](#) open-source software tool "[GetFreeCopy](#)" that is demonstrative and addresses many features of this. To tell us more we have a [Q&A](#) with lead author Kuan-lin Huang, an Assistant Professor of Genetics and Genomics & Artificial Intelligence and Human Health at Icahn School of Medicine at Mount Sinai, and co-founder of [OpenBoxScience](#) (OBS), an NGO for Open Science & Education.

Open Access For the Win!

Researchers from the Mount Sinai Center for Transformative Disease Modeling in New York have just published a [new paper](#) presenting [GetFreeCopy](#), a web-based platform designed to streamline the search for biomedical literature across major repositories and preprint servers such as arXiv, bioRxiv, medRxiv, and PubMed Central. Addressing challenges posed by paywalls and fragmented databases, it offers a unified interface for efficient retrieval of free, legitimate copies of biomedical literature. The *GigaByte* paper promoting the use of this tool to the wider community to make access easier, [the code](#) is also open-source and can be built upon and adapted by others. While there are existing tools and plugins such as [Unpaywall](#) designed to facilitate easier access to articles, this new tool has been updated to support to growth of preprint publications, is completely open, and is accessible in a wider number of ways to broaden the use across different research contexts.

GetFreeCopy from OpenBoxScience

Screenshot of the different GetFreeCopy outputs (figure adapted from the paper).

****Welcome to the Machine. Publishing in the ChatGPT Era.**

****On top of promoting Open Access, GigaScience Press has [pushed transparency in research](#) since we [launched GigaScience](#) in 2012, promoting [creative commons licensing](#) and [data mining](#) of our content, [transparent Open Peer Review](#), [usage of preprints](#) to throw light further up the research cycle, and mandating that all data and code supporting our papers is publicly available under open licenses. Large Language Models (LLMs) such as ChatGPT have had a significant impact in how research is now being carried out, but these technologies are particularly challenging to trace the provenance of, and there has been a lot of [controversy](#) and [challenges](#) with their usage in scientific publishing. As these technologies are here to stay, at GigaScience Press we have continued to follow our "sunlight is the best disinfectant" approach, updating our [Editorial policies](#) to allow the use of AI tools and technologies in paper writing as long as the human authors carefully check and take all the credit and responsibility for the AI's work (see the [COPE position statement](#) on this). Any usage in creating the paper then needs to be declared in a section at the end of the paper, and the outputs from the LLM are archived somewhere publicly (such as [in our GigaDB repository](#)) and cited in the paper. This follows the approach proposed by [Mohammad Hosseini and colleagues](#) in disclosing the use of AI-tools in writing scholarly manuscripts. This transparency-first approach is demonstrated in this new paper by Kuan-lin Huang, in which on top of assisting in the writing of the paper, ChatGPT4 also assisted in the coding of the tool.**

Kuan from OpenBoxScience

Kuan-lin Huang

In this [author Q&A](#) Kuan-lin Huang tells us about his rationale for building the [GetFreeCopy](#) tool, his Open Science and Education NGO [OpenBoxScience](#), and his usage of ChatGPT in his work..

Before you created GetFreeCopy what was your workflow for finding papers?

Before creating [GetFreeCopy](#), finding research papers could be a hassle. I would often start by searching on PubMed, which often has free full-text articles through PubMed Central. However, if the given journal was not open access, I frequently had trouble accessing the papers. When off-campus or if my institution didn't subscribe to a specific journal, I would then have to search preprint servers like bioRxiv, medRxiv, and arXiv. I often find myself opening multiple tabs and searching each repository individually, which could be time-consuming and frustrating.

What inspired you to automate this process with the tool?

The whole chore of searching across multiple repositories simultaneously is what inspired me to automate this process with [GetFreeCopy](#). By consolidating the search into a single tool, I could streamline the process of finding freely available research papers—using a one-stop shop!

You are the co-founder of [OpenBoxScience](#), so can you tell us a bit more about that?

Sure. OpenBoxScience (OBS: [OpenBoxScience.org](#)) is a 501(c)(3) nonprofit organization that I co-founded, dedicated to promoting open science and democratizing scientific training. Scientists everywhere are free to join the OBS community, where we come together to organize free virtual seminars that are open to a worldwide audience. OBS started during the COVID-19 pandemic lockdown in 2020. Back then, I recruited a few friends, and we started hosting these free, virtual seminars where first authors and early career scientists presented their latest work. Instead of keeping it exclusive, we opened it up to audiences worldwide. It was very cool to have people tuning in from over 70 countries. Since then, we've already organized over 200 of these talks. We're breaking down barriers and connecting people across disciplines and borders. It's been an incredible journey spreading knowledge and democratizing science education.

Get Free Copy is supported by [OpenBoxScience](#), so how does it fit in and assist the work you are doing, particularly as a nonprofit dedicated to promoting open science and democratizing scientific training?

GetFreeCopy is a perfect fit for the open science mission we're pushing at OBS. Many OBS community members come from under-resourced countries or institutions where they can't always access research papers behind expensive paywalls. With [GetFreeCopy](#), these researchers and students have an easy way to find legitimate, free copies of papers across different repositories. It's like a one-stop shop for open access literature.

The tool complements the [free virtual seminars](#) hosted by the OBS [see the example below of [GigaScience author](#) Shing Wan Choi [presenting a tutorial on PRS analysis](#)]. Attendees can use it to track down any papers discussed during the talks or other papers of their interest. It helps create a more comprehensive learning experience that can work for everyone. At the end of the day, [GetFreeCopy](#) aligns with our core values of breaking down barriers, promoting open science, and making scientific knowledge accessible to everyone, everywhere.

Large language models (ChatGPT 4) were used in the initial drafts of coding, literature review, and writing of this work. How have LLMs changed the way you carry out and share research?

LLMs like [ChatGPT](#) have been a total game-changer for how I do research. To be honest, GetFreeCopy started as a fun little 2023 New Year's coding project to see if I could build something useful with GPT's help, despite having minimal web dev experience. Thankfully, with GPT's assistance and the leadership of Nodir who got it deployed, we were able to make it happen.

GigaScience Press has attempted to make this process more transparent by asking for this process to be transparent, and the outputs to be shared and cited. How do you find this approach?

While LLMs have made so many tasks way more efficient, I'm all for the transparency approach that GigaScience Press is advocating. It's important to be upfront about when LLMs were involved. At the end of the day, whether the work was done by humans or machines (or most likely, a combination), we scientists need to carefully review everything to ensure accuracy to the best of our ability. Having reproducible experiments is the only way we can accumulate meaningful evidence and advance knowledge in science.

Access GetFreeCopy here: <https://getfreecopy.com/>

References

Hosseini M., Resnik DB., & Holmes K. (2023). The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*, 19(4), 449-465. <https://doi.org/10.1177/17470161231180449>

Kosimkhujaev N., & Huang KL. (2024) Get Free Copy: a multi-repository search platform for biomedical publications, *GigaByte* <https://doi.org/10.46471/gigabyte.126>

The post [Transparency FTW! LLMs, OpenBoxScience and GetFreeCopy](#) appeared first on [GigaBlog](#).