U.S. HOUSE OF REPRESENTATIVES COMMITTEE ON
SCIENCE, SPACE, & TECHNOLOGY
Opening Statement

Chairman Bill Foster (D-IL)
of the Subcommittee on Investigations and Oversight

Investigations & Oversight Subcommittee Hearing:
Paper Mills and Research Misconduct: Facing the Challenges of Scientific Publishing
July 20, 2022

Good morning and welcome to our Members and witnesses.

For today's hearing, I'm proud to announce that Representative Perlmutter and I have, in our spare time, been conducting experiments on a groundbreaking topic in nuclear physics. We're excited to share the results of that effort today as a preprint. We plan to submit it to *Reviews of Modern Physics*.

Just kidding. What I'm actually holding is a cheap rip-off of a seminal paper called "*Neutron Production and Absorption in Uranium*," which was published in the journal *Physical Review* in 1939. Its author was Enrico Fermi. We took Dr. Fermi's paper and ran it through a free online text generator that uses artificial intelligence to disguise plagiarism. This took 15 seconds. We then took five minutes to tweak a few sentences to disguise their true source a little better. Once it was ready, we ran this paper through two well-respected plagiarism checkers. We even sent it over to the Inspector General at the National Science Foundation. Each of these detectors found our fake paper was, and I quote -- "100% unique, 0% plagiarism."

Now, any real physicist peer reviewer for a journal or an NSF grant would notice immediately that this paper uses silly technical jargon and plagiarizes from a very famous paper. They would also find it unconventional that the report was authored by two sitting Congressmen and includes an acknowledgement to Ranking Member Jay Obernolte. But you can imagine how bad actors might use tools as we did to sneak plagiarized content past journal editors and peer reviewers.

The AI-assisted plagiarism tool we used to make the fake paper is only one of many in the arsenals of "paper mills." These are criminal enterprises that sell authorship credits for the fraudulent papers they place in academic journals. Scientific disciplines such as the life sciences, which rely heavily on images to communicate the results of experiments, are popular targets for fraud because of how easy it is to manipulate images. Now, with the advent of sophisticated natural language processing software, it is becoming just as easy to churn out fraudulent but coherent text. Add in a few basic templates and the creation of hundreds of papers – complete

with figures and citations – becomes the work of an afternoon, much to the disgust of real scientists who might spend months on a single paper.

The scientific community must rise to meet this challenge, and it is already taking the first steps. Journals are looking for new ways to collaborate in detecting fraud during the review process. One recent effort is the STM Integrity Hub, which will serve as a platform for journals to share dedicated fraud detection tools. The first tool under construction will flag the simultaneous submission of papers to multiple journals, a strong indicator of paper mill activity.

There is also a strong international community of researchers unaffiliated with publishers, many of them volunteers, who work to identify fraudulent papers following publication. Just as automation is enabling those committing fraud, it is also being used by these researchers to combat it. Next generation plagiarism checkers don't just compare text to text, but intelligently scan for indicators of AI-generated text. Other tools detect manipulated images or identify erroneous science within the text of the paper itself. The automation arms race is upon us. We are here today to discuss how researchers and publishers can develop tools and policies that will keep them ahead of the paper mills.

As we discuss scientific misconduct today, one of the most important things to keep in mind is the scale of this problem. Hundreds of papers with signs of fraud are indeed a serious concern. However, according to NSF, a whopping 2.9 million papers were published last year alone. The number of cases of fraud must be viewed within that context. Creating and maintaining a body of scientific literature without flaws of any kind is a quixotic quest.

But published scientific literature remains the greatest body of human knowledge about the world, and it is our responsibility to look after its integrity. This effort begins with a public dialogue. As Dr. Fermi said famously –

> *"Whatever Nature has in store for mankind, unpleasant as it may be, men must accept, for ignorance is never better than knowledge."*

I look forward to earning some more knowledge today with the help of our esteemed witnesses.

I now yield to Ranking Member Obernolte for his opening statement.