Open Science Fair 2025



Contribution ID: 177 Contribution code: -177

Type: Poster

The Double-Edged sword of Open Science: Empowering AI through ethical breaches within repositories

With the rapid expansion of AI and its applications in recent years, a new challenge has emerged: how to sustainably improve the quality of the data on which AI builds its knowledge. One solution is leveraging the benefits of Open Science, with its open infrastructure and repositories, as one of valuable sources of such knowledge. However, as global demand for data to advance AI systems increases, it has led to a rise in bot and crawler activity that overwhelms servers, amplifies system overhead, degrades service quality, and restricts access for legitimate users.

The primary challenge for repositories that follow the principles of Open Science is to find ways to manage crawler-generated traffic in a manner that ensures continued access for legitimate users, upholds ethical standards for information exchange over the Internet, and strives to serve all users without significant delays. Our goal was to identify a solution capable of recognizing damaging crawler patterns and guiding their actions to comply with ethical standards.

The University of Belgrade and its Computer Centre have recently faced a series of challenges due to traffic overload in their repositories, including incidents such as DDoS attacks. These issues compromised the daily operations of repository administrators and researchers who actively contribute to the repositories. Since it affected multiple production instances, the overall infrastructure was endangered. As a result, we developed a new generalized methodology to identify malicious crawlers and guide them to operate within ethical behavior principles.

Tagline

Developing AI-driven tools to support scholarly communication and Open Science, with a focus on accessibility, innovation, and community impact

Keywords

Web crawlers, Internet bots, DDoS attack, Pattern detection

Authors: KOSANOVIĆ, Biljana (University of Belgrade - Computer Center); STIJOVIĆ, Dušan (University of Belgrade - Computer Center); OTAŠEVIĆ, Vladimir (University of Belgrade, School of Electrical Engineering)

Session Classification: Poster & Demos Sessions

Track Classification: Building the Digital Backbone: Open Science Infrastructures