IT **.**

🇬🇧   **Subscribe ▼**

| TRENDING | Enhancing Business Security | Artificial intelligence in business | The best business servers for 2024 |

Software > Open Source

New definition of open source AI is "flawed", experts say

News   By George Fitzmaurice published 2 days ago

The terms compel open source AI projects to reveal information around training data – but there are a few key caveats

f X P

When you purchase through links on our site, we may earn an affiliate commission. Here's how it works.
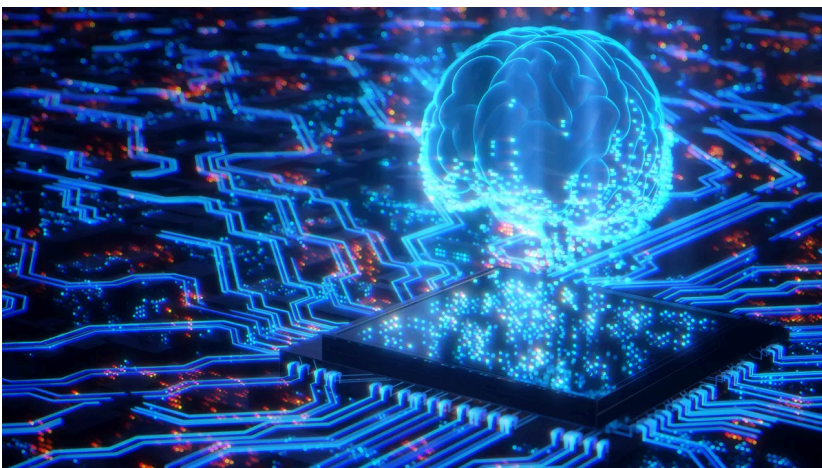
(Image credit: Getty Images)

A new definition for 'Open Source AI' has been launched by the Open Source Initiative (OSI), though experts have told *ITPro* that the terms lack nuance and may be under the wrong management.

The new definition has AI training data as its focus, clarifying that training data must be shared and disclosed. It also states that code must be

complete to the extent that recipients understand how the training was done.

"In my opinion, OSI continues with its flawed 'one size fits all' approach rather than helping to better define the 'spectrum' for open source and AI." Peter Zaitsev, founder of Percona, told *ITPro*.

The OSI's new definition considers four types when it comes to training data - open, public, obtainable, or unshareable. It noted that, while the legal requirements are different for each, all must be shared in a form allowed by law to adhere to the new terms.

LATEST VIDEOS FROM ITPRO

It highlights two key features, the first of which demands that the code used to train and process data in AI development must be complete to the extent that open source recipients understand how the training was done.

Training is where the innovation is taking place, the OSI said, so transparency around the code used in training is necessary to allow open source users to study and modify AI systems.

Another feature of the definition acknowledges that requirements of 'copyleft-like terms' are admissible. This is where the training code and a dataset are bundled together in a legal sense. The OSI's new definition is at the 'release candidate' stage, meaning no new features will be added going forward, only bug fixes.

Get the ITPro. daily newsletter

Receive our latest news, industry updates, featured resources and more. Sign up today to receive our FREE report on AI cyber crime & security - newly updated for 2024.

| Your Email Address | **SIGN ME UP** |

☐ Contact me with news and offers from other Future brands

☐ Receive email from us on behalf of our trusted partners or sponsors

By submitting your information you agree to the Terms & Conditions and Privacy Policy and are aged 16 or over.

LATEST ARTICLES
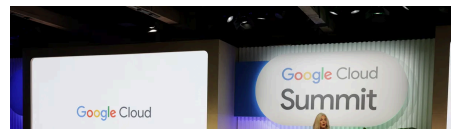
1 Google Cloud announces data residency for machine learning in the UK

2 North, south, east, west-- Securing your network with zero trust

3 UK Government launches competition to find future cybersecurity professionals

A contentious definition

Zaitsev takes issue with several terms in the definition, particularly the triaging of data types into obtainable and unshareable. In a linked FAQ, the OSI clarified that obtainable data can be revealed for a cost while unshareable data can only be revealed in the form of a detailed description.

"While it makes a lot of difference for actual users, if training data is not freely available for everyone, it is not the same as 'open source'," Zaitsev said.

"I think it would make sense for OSI to lead the effort to properly define the standard classification for these free-to-use models which, in my opinion, and in particular due to massive training costs, is where potential value for competition will be massive," he added.

Amanda Brock, CEO of OpenUK, told *ITPro* that the issues here are even broader and that the problem is more deeply entrenched in the OSI's position as an institution.

RELATED WHITEPAPER

- *Optimize your security by adopting AI*

"This is not only concern about the content of any definition, but whether there should be an 'open source AI definition,' and, if there is one, whether the OSI is the right organization to create it and whether the broader open source software community support its changed role as the custodian of two definitions," Brock said.

"The OSI's stated purpose is around open source software - yes, advocating for open source principles is a part of that purpose, but it's questionable whether managing a whole new definition in AI falls under that purpose," she added.

Brock's thinking is that the OSI should maintain its focus on open source software, more than enough work for one small organization to manage, in her opinion. The OSI's role as a guardian of the Open Source Definition (OSD) is critical, she added.

"The open source software community is being split and fractured by the new Open Source AI definition," Brock said.

George Fitzmaurice

Staff Writer

George Fitzmaurice is a staff writer at *ITPro*, C*hannelPro*, and *CloudPro*, with a particular interest in AI regulation, data legislation, and market development. After graduating from the University of Oxford with a degree in English Language and Literature, he undertook an internship at the New Statesman before starting at ITPro. Outside of the office, George is both an aspiring musician and an avid reader.

MORE ABOUT OPEN SOURCE

LATEST
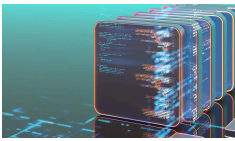
Redis insists license changes were the "only way to compete with Amazon and Google" — now it could face a user exodus ▶
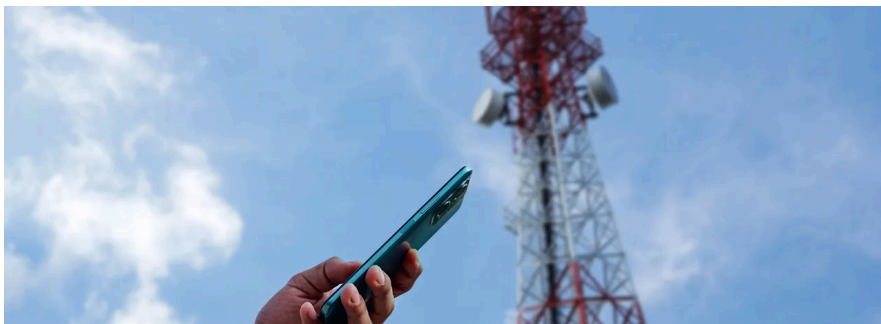
Elastic returns to open source, but can it regain the community's trust? Some industry players aren't holding their breath ▶

Three secrets to success for the MSSP ▶

SEE MORE LATEST ▶

MOST POPULAR

UK still lags behind other countries in 5G connectivity

Cybersecurity is the fastest growing tech occupation in the UK – but it's still not enough to dent the growing industry skills shortage