



**DIGITAL SERVICE AT CMS**

# **Open Source Program Office (OSPO) Briefing: Open Source & AI**

**AI Cross-Cutting Initiative // Feb 26th, 2024**

# CMS Open Source Program Office (OSPO)



**Natalia Luzuriaga**  
US Digital Corps  
Fellow '23



**Isaac Milarsky**  
US Digital Corps  
Fellow '23



**Dinne Kopelevich**  
US Digital Corps  
Fellow '24



**Sachin Panayil**  
US Digital Corps  
Fellow '24



**Remy DeCausemaker**  
Open Source Team Lead



**Marcus Koenig**  
CMS Intra-agency Rotation  
Data Engineer Q1 25



# OSPO Briefing

CMS AI CCI

- 1. Open Source & OSPO Intro**
2. Open Source AI Context
3. Open Source AI Takeaways
4. Open QnA



# Open Source Program Office (OSPO) Market Dynamics

96% of all commercial software contains open source software, representing \$8.8T of value if rewritten from scratch.

30% of Fortune 100 companies have implemented an OSPO, with 84% of OSPOs developing or managing GenAI infrastructure.

CMS is the first United States Federal Agency to implement this private-sector styled approach to an Open Source Strategy.

see: [linuxfoundation.org/research/ospo-2024](https://linuxfoundation.org/research/ospo-2024)

## THE 2024 STATE OF OSPOS AND OPEN SOURCE MANAGEMENT

**94%** of organizations **USE OPEN SOURCE SOFTWARE** in products or services.



**77%** of **LARGE** organizations **HAVE AN OSPO.**



**19%** of **SMALL**, and **33%** of **MEDIUM** organizations **HAVE AN OSPO.**



**105% GROWTH OF OSPOS IS EXPECTED FOR SMALL ORGANIZATIONS** over the next one to two years compared with just 12% for very large organizations.



**80%** of organizations report that OSPOs have a **MEANINGFUL POSITIVE IMPACT** on their ability to work with open source communities.



**#1** OSPO responsibility for small & medium organizations: **DEVELOP AND EXECUTE OPEN SOURCE STRATEGY.**



**#1** OSPO responsibility for large organizations: **ESTABLISH AND IMPROVE OSS POLICY AND PROCESSES.**



Top benefits reported from organizations having an OSPO are **BETTER LICENSE COMPLIANCE AND INCREASED TRANSPARENCY** of collaboration.



**#1** benefit anticipated from implementing an OSPO by organizations not having one is **MORE AWARENESS OF OSS USE AND DEPENDENCIES.**



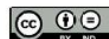
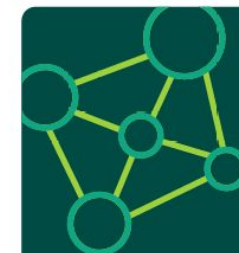
**#1** OSPO **CHALLENGE** reported from organizations having an OSPO is **INTERNAL AWARENESS** across teams of the program or initiative.



**91%** of OSPOs are involved in **MANAGING SECURITY ISSUES**



**84%** of OSPOs are involved in **DEVELOPING OR MANAGING GenAI INFRASTRUCTURE.**



## Early/Mid Career Talent Pipeline 2024: 9 employees across 5 programs

### 2 Local HS Interns - [CodeInTheSchools.org](https://codeintheschools.org) - 5/10 weeks, Aug-Oct

*“5-week summer experience for **Baltimore City residents** between the **ages of 14 and 21**, Operating in partnership with **Baltimore City YouthWorks Summer Jobs Program**, managed by the **Mayor's Office of Employment Development**.”*

### 3 Undergrad Fellows - [CodingItForward.com](https://codingitforward.com) - 10 weeks, July-Aug

*“10-week cohort based summer fellowship for **early-career technologists** in local, state, and **federal government** offices across the United States.”*

### 2 Recent Grad Fellows - [DigitalCorps.gsa.gov](https://digitalcorps.gsa.gov) - 2 years, September

***2 year tour of duty** for **early-career** technologists, eligible to convert to full-time, career positions in the competitive service at their agency. [GS-9 to 12](#), **50% recruitment Incentive**.*

### 1 Graduate Fellow - [PMF.gov](https://pmf.gov) - 6 weeks, September

***2 year full-time paid fellowship**, “Federal Government's premier leadership development program for advanced degree holders across all academic disciplines.”*

### 1 Participant - **CMS Intra-Agency Rotational Program** - 3 months, January-April

***3 month rotational program**, “Enriching development opportunity offering employees the chance to grow within CMS, expand their professional competencies and build networks”*

# What is an Open Source Program Office? (OSPO)

*An open source program office (OSPO) serves as the center of competency for an organization's open source operations and structure. It is responsible for defining and implementing **strategies**, **programs**, and **policies** to guide these efforts.*

## CMS OSPO Functional Statement:

*“Establishes and maintains guidance, policies, practices, and talent pipelines that advance equity, **build trust**, and **amplify impact** across CMS, HHS, and Federal Open Source Ecosystems by working and sharing openly.”*





## What is expected of an OSPO?



Save us Money



Save us Time



Reduce Duplicate Work



Reduce Duplicate Costs



Reduce Security Risk



Reduce Continuity Risk



Accountability for Contract Performance



Engine for Talent

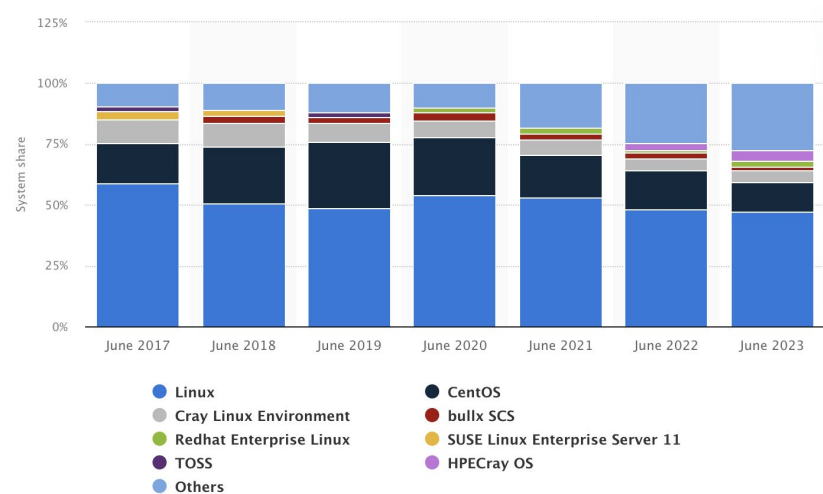
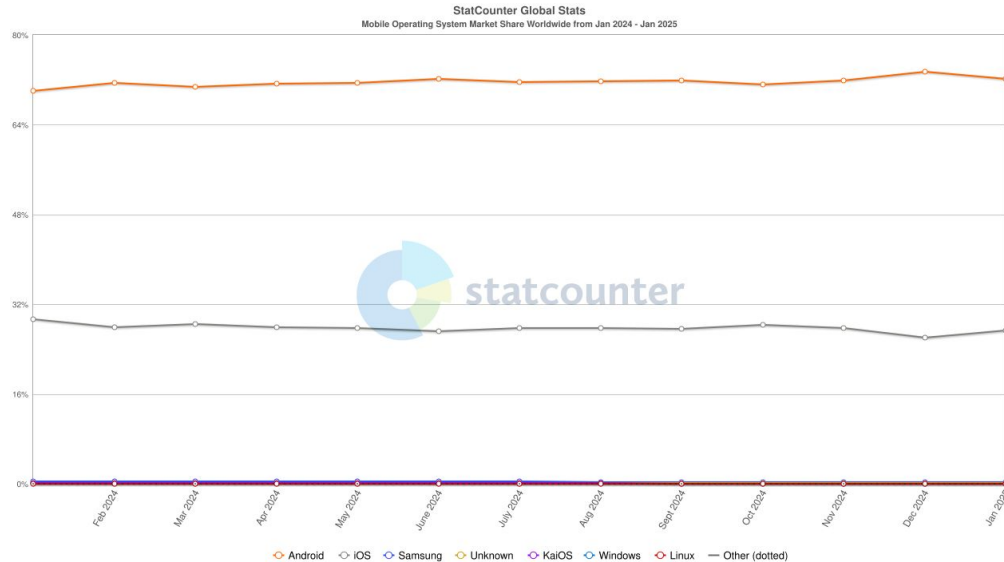
# What is Open Source Software?

Open Source means source code that anyone can **inspect**, **modify**, **enhance**, and **share**.





# Open Source Market Dynamics



- For smartphones and other mobile devices, Android has **73%** market share, and Apple's iOS has **26%**.<sup>[2]</sup>
- For the **top 500 most powerful supercomputers**, Linux distributions have had **100%** of the marketshare since 2017.
- The global server operating system marketshare has Linux leading with a **62.7%** marketshare, followed by Windows, **Unix** and other operating systems.<sup>[6]</sup>

See: [en.wikipedia.org/wiki/Usage share of operating systems](https://en.wikipedia.org/wiki/Usage_share_of_operating_systems)

<https://gs.statcounter.com/os-market-share/mobile/worldwide>

# CMS Open Source Footprint

**1,000's**

open source contributors

(As of 02/2025)

**10,000's**

source code repositories

(As of 02/2025)

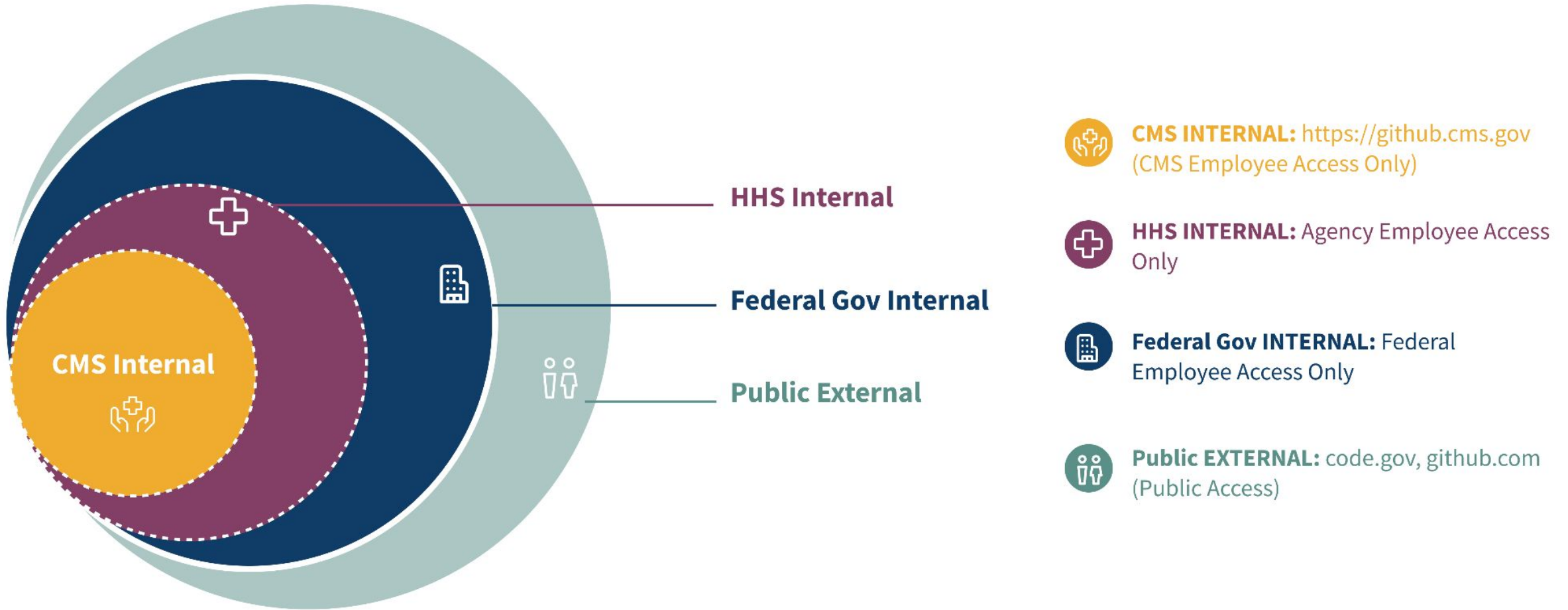
**1,000,000's**

Lines of Code

(As of 02/2025)

<https://dsacms.github.io/metrics>

# Open Source is a Community of Communities



# Open Source APIs at CMS: Developer.cms.gov



**CMS Blue Button 2.0**



**Beneficiary  
Claims  
Data API**



**Data at the  
Point of Care**



# CMS Source Code Stewardship Taskforce Strategy



**We know where  
our repos are.**

Inventory of all the  
components using GitHub



**We know who is  
working on them.**

Understand who is  
maintaining and  
contributing to projects, and  
how often.



**We know what is  
inside them.**

Identify FISMA labels,  
repository topics, maturity  
models, security  
vulnerabilities, and library  
composition

# CMS Open Source Policy Update V1.1 (First in 6 years!)

## Changes include:

- Incorporates OSPO's baseline tools and processes
  - Outbound Review Process
  - repo-scaffolder for repo creation
  - metrics for reports on OSS projects
- Provides guidance and best practices on CMS public OSS repositories
- Reflects current state of software development
  - List of active OSS projects
  - Usage of TLC framework
- Fixes broken links
- Credited more agencies and organizations in acknowledgement section

## OSS Implementation & Infrastructure

Any CMS project team that wishes to publish their code as OSS must set a clear expectation of their level of involvement in sustaining that project. The project team shall define an OSS release process that begins with a determination of if the intended software can indeed be released as OSS, considering any security and other CMS policy restrictions. Depending on the nature of the OSS and associated licensing model, the project team shall adequately allocate resources to be able to sustain and flourish the project in the open source community.

The project team shall utilize an existing public-facing website to convey information about the project and provide a link to the project's GitHub repository. The project team should implement tools to support the community around an open source project, such as mailing lists, message forums, a version control, wiki, and tracking mechanisms, such as Kanban boards to track issues and bugs on the GitHub repository.

For every iteration of the code release, the project team must ensure that the software code is adequately peer reviewed ~~and is free of~~ security vulnerabilities that can be exploited by malicious actors and that any discovered vulnerabilities are removed prior to release. Until the software code is adequately reviewed, it should be either 1) maintained in an internal code repository that replicates the intended public repository or 2) checked by publicly available services providing the same functions on all code check-ins to the public source code repository. The software should also contain automated unit tests, and build scripts and should be checked for software vulnerabilities, code quality and code coverage using available standard CMS tools and guidance. The code should be built with CMS's standard continuous integration (CI) ~~server~~servers. The project team should include ample documentation with the software code for increased adoption and modification by the community. The documentation should provide the information on project's mission, philosophy, goal, design, decision-making process, product roadmap and instructions on how to submit issues, feature requests and how to contribute towards a fix or enhancements. The documentation should be accessible to the Open Source community via the repository.

<https://github.com/CMSgov/cms-open-source-policy>





# Open Source in the News: SHARE IT Act of 2024

Source Code Harmonization and Reuse in Information Technology Act

Passed by Congress and signed into law on December 23rd!

One Hundred Eighteenth Congress  
of the  
United States of America

AT THE SECOND SESSION

*Begun and held at the City of Washington on Wednesday,  
the third day of January, two thousand and twenty-four*

## An Act

To require governmentwide source code sharing, and for other purposes.

*Be it enacted by the Senate and House of Representatives of  
the United States of America in Congress assembled,*

### SECTION 1. SHORT TITLE.

This Act may be cited as the “Source code Harmonization  
And Reuse in Information Technology Act” or the “SHARE IT  
Act”.

### SEC. 2. DEFINITIONS.

In this Act:

(1) AGENCY.—The term “agency” has the meaning given  
that term in section 3502 of title 44, United States Code.

[HR9566 PDF](#)

## tl;dr

- OMB and Agency CIOs are responsible to ensure ALL custom source code, Fed or Contractor, to be:
  - In at least one public or private repo
  - Contain metadata about the repo
  - Update Agency Acquisition policies
  - Report by Comptroller General to OPM and Congress one year after enactment annually

[HR9566 Summary Congress.gov](#)

[FedScoop Article](#)



# OSPO Briefing

CMS AI CCI

1. CMS OSPO Intro
2. **Open Source AI Context**
3. Open Source AI Takeaways
4. Open QnA



# What is an AI Model?

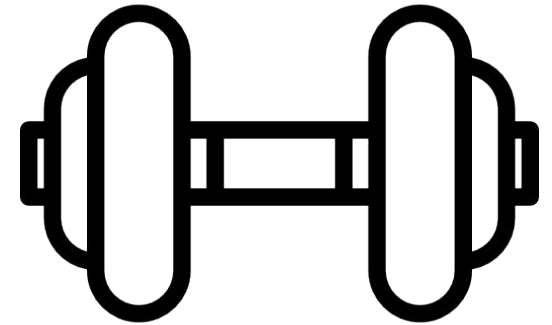
It has at least three parts:



**Training data**



**Source Code**



**Model Weights**

# Is there an Open Source AI “Definition”?

## Opensource.org (OSI) Open Source AI Definition

<https://opensource.org/ai/open-source-ai-definition>

“...an AI system made available under terms and in a way that grant the freedoms to:

- **Use** the system for any purpose and without having to ask for permission.
- **Study** how the system works and inspect its components.
- **Modify** the system for any purpose, including to change its output.
- **Share** the system for others to use with or without modifications, for any purpose.

**Sufficiently detailed information about the data** used to train the system so that a skilled person can build a substantially equivalent system.

The **complete source code** used to train and run the system. The Code shall represent the full specification of how the data was processed and filtered, and how the training was done.

The **model parameters, such as weights or other configuration settings.** Parameters shall be made available under OSI-approved terms.

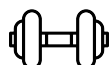
Training data



Source Code



Weights



## UN/Digital Public Goods Alliance Open Source AI Definition

<https://www.youtube.com/watch?v=9k6aVS-gc5s>

To qualify as a Digital Public Good, *open source AI* systems must provide:

The **dataset(s)** used to train the system should be open.

### All source code

including:

- Data pre-processing
- Training, validation, and testing
- Inference
- Supporting libraries and tools

**Model architecture:** the type of model, layers, and structure

*Available under an OSI-approved license*

**Model parameters:** weights, optimizers, coefficients, and other applicable hyperparameters

*Available under conformant licenses to the Open Definition*

*Available under OSI-approved license*

*Available under an OSD-conformant terms*

Training data





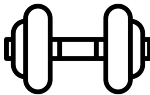






Source Code



Weights



# What is Open Source AI?

Definition	Training Data <div></div>	Source Code <div></div>	Weights <div></div>
Open Source Initiative (OSI)	<div></div> <p>Sufficiently detailed information about the data used to train the system so that a skilled person can build a substantially equivalent system</p>	<div></div>	<div></div>
United Nations/Digital Public Goods Alliance	<div></div> <p>Open Training Data</p>	<div></div>	<div></div>

**Open Source Initiative:** For their purposes, there only needs to be **"sufficiently detailed information about the data used to train the system"** rather than the full dataset itself. This approach aims to balance transparency with practical and legal considerations such as copyright and private medical data.

**UN/Digital Public Goods Alliance:** “This proposal, which includes **the requirement to open training data is a fundamental change to the Standard**, that directly affects other indicators such as Platform Independence and although it may restrict the number of AI solutions that meet the DPG Standard initially, we believe it reinforces the DPG’s core values, including transparency, equity, and ethical development. This requirement ensures that AI systems contributing to the public good do so with a commitment to openness that supports accountability, safety, and meaningful societal benefits.”



# How robust is the Open Source AI Community?

The open source AI community has seen remarkable growth, with **100,000+** new projects on GitHub *in the past year alone*.

**68 new projects** are now being hosted by the LF AI Foundation

MEMBERS

Premier

aws

databricks

HUAWEI

intel

IBM

Microsoft

Neural Magic

NVIDIA

ZTE

General

AMD

Alibaba

蚂蚁集团 ANT GROUP

CLOUDERA

CloudGeometry

Datastrato

数元灵 DMATOSUI

FUJITSU

FURIOSA

FUTUREWEI Technologies

GitCode

neo4j

Network Optix

new relic

NOKIA

Okahu

posit

precisely

Red Hat

sas

snyk

TikTok

union

zilliz

The LF AI & Data Foundation has had contributions from over **100,000 developers**

PROJECTS AND PRODUCTS

Machine Learning

Library

1ML

DeepCausality

IREE

Recommenders

SapientML

AutoDL

Cartosat

FlareFlight

MediPro

mlpack

OpenCV

Accord

Microsoft Azure

Mapping Data Center

ML.NET

RAY

ZenML

Platform

Angel

ForestFlow

Open Platform for Enterprise AI

Alibaba AIlink

Ax

Cortex

H2O

Kubeflow

mlflow

SELDON

Deep Learning

Framework

DeepRec

RVOMA

ShaderNN

PyTorch

TensorFlow

Tool

BeyondML

BoTorch

Intel Distiller

Pytorch

PyTorch

PyTorch

As of October 2024 the LF AI & Data Foundation was comprised of **77 companies** North America, Europe and Asia.

COMPANIES HOSTING PROJECTS

Hosting Companies

蚂蚁集团 ANT GROUP

amazon

AT&T

BAI 百度

Baidu 百度

BASIC AI

Bloomberg

data.world

databricks

datakin

DiDi

数元灵 DMATOSUI

DREAMWORKS

EMET-LABS

ERICSSON

FUJITSU

GRAVITI

Herron Tech

IBM

intel

jine

LabeliaLabs

LinkedIn

lyft

Meta

Microsoft

Okahu

gojek

OWKIN

QuantumBlack AI by McKinsey

R Studio

SAMSUNG

Tech Mahindra

Tencent 腾讯

ETL

Uber

Berkeley

wework

zilliz

ZTE





# The Rise of Open Source AI

The AI market has shifted as open source large language models have grown to challenge proprietary models, and proven to be as useful and efficient, without drawbacks in Cost, Performance, and Quality.

## Cost

---

- Locally hosted AI
- Reduced infrastructure costs
- Reduced hardware costs

## Performance

---

- LoRA gives cost efficient adaptation to OS models
- Quantization reduces memory size so low-end hardware can run LLMs

## Quality

---

- Flexibility of model allows for use case fine tuning
- Same or better quality as proprietary models

# How are Open Source AI/LLMs shaping the Market?

## Cost

- Rather than paying high API fees for hosted proprietary models, users can fine tune and deploy their own open source models *significantly* reducing costs. Many cloud providers are optimizing for open source AI hosting as a key revenue source... Kudos to chat.cms.gov for adapting!

## Performance

- Fine tuning on *domain specific data such as medical or legal data* allows purpose built open models to compete with proprietary general models. Multiple techniques has been developed that make fine tuning open models easier and perform just as effectively as proprietary.

## Quality

- Flexibility is the biggest quality advantage when comparing to proprietary models. Users can tailor a model to specific use cases depending on their needs, which puts model quality on par with the biggest proprietary services out there.

# What new techniques makes Open Source more efficient?

## Low-rank adaptation (LoRA)

- Fine tuning method that allows new model adaptation without modifying original models weights.
- *This reduces the level of hardware needed to build LLMs while also making it more cost effective.*

## Quantization

- Reduces model weight precision from 16-bit floats to 8-bit and sometimes 4-bit integers. This makes models much smaller by reducing storage size, while maintaining a similar levels of accuracy.
- *This allows for open source models to be run on lower-end hardware, making AI much more accessible.*

## Distillation

- A method that trains a smaller model based on the behaviors of a larger model.
- *This allows for compact models to be run on lower-end hardware while keeping similar performance.*

## AI Market Disruptions

Dynamics in the AI space were recently disrupted by the **Deepseek R1 LLM model**. This open weight model has shown that AI can be produced that has better results, at a lower cost, while using far fewer resources than other proprietary LLM models.

Deepseek has not released full numbers, but cost of development has been estimated to be **under \$6 million dollars**.

OpenAI's ChatGPT cost of development has been estimated to be **hundreds of millions of dollars** just for their o1 model alone...

Though some countries and organizations around the world have already banned DeepSeek, citing ethics, privacy and security issues within the company, their more open approach has already impacted the market significantly.



# OSPO Briefing

CMS AI CCI

1. CMS OSPO Intro
2. Open Source AI Context
- 3. Open Source AI Takeaways**
4. Open QnA



# What are our take-aways at the OSPO?

Open means **Reproducible**.

- Yes access to the source code, but also access to training data and model weights.

Open means **no paywalls**

- Minimizing financial costs beyond necessary infra for the agency and our users.

Open means **no platform or vendor lock-in**

- Digital Public Goods are platform agnostic, and competition keeps prices lower.

Open means **open to contributions**

- Not just one-way sharing code, but enabling two-way participation by welcoming and encouraging bug reports, issues, and pull-requests from our users.



## ● What does this mean for CMS?

Within the Open Source AI community, we can be more than just consumers, we can be contributors, even leaders!

*If CMS isn't the most authoritative upstream source for Training Data, Source Code, and Models for Medicare and Medicaid Services, then who should be???*

### Immediate contributions CMS can make to Open Source AI:

- Kaggle: Provide de-facto standard **training datasets** from CMS
- HuggingFace: Share our **models** where researchers are innovating
- GitHub: Continue sharing our **source code** and publish more inventories
- JSON-LD: Provide **metadata** that optimizes for search *AND* for AI training



—• **Thank You AI CCI!**



**DIGITAL  
SERVICE  
AT CMS**



## **Questions or Comments?**

<https://go.cms.gov/ospo>

**Open Source Questions?**  
[opensource@cms.hhs.gov](mailto:opensource@cms.hhs.gov)

**Digital Service Questions?**  
[DigitalService@cms.hhs.gov](mailto:DigitalService@cms.hhs.gov)

## **Help Answer The Call!**

**Digital Service at CMS.gov**  
<https://cms.gov/digital-service>

**DigitalCorps Fellowships**  
<https://digitalcorps.gsa.gov>

**CodingItForward Summer Internships**  
<https://codingitforward.com>

