



PROGRAM MATERIALS

Program #36208

September 22, 2026

Open Source AI: The Unstoppable Rise and Legal Issues You Must Know How to Manage

Copyright ©2026 by

- **James Gatto, Esq. - Sheppard**

All Rights Reserved.

Licensed to Celesq®, Inc.

Celesq® AttorneysEd Center

www.celesq.com

**5301 North Federal Highway, Suite 150, Boca Raton, FL 33487
Phone 561-241-1919**



Open-Source AI: The Unstoppable Rise and Legal Issues You Must Know How to Manage

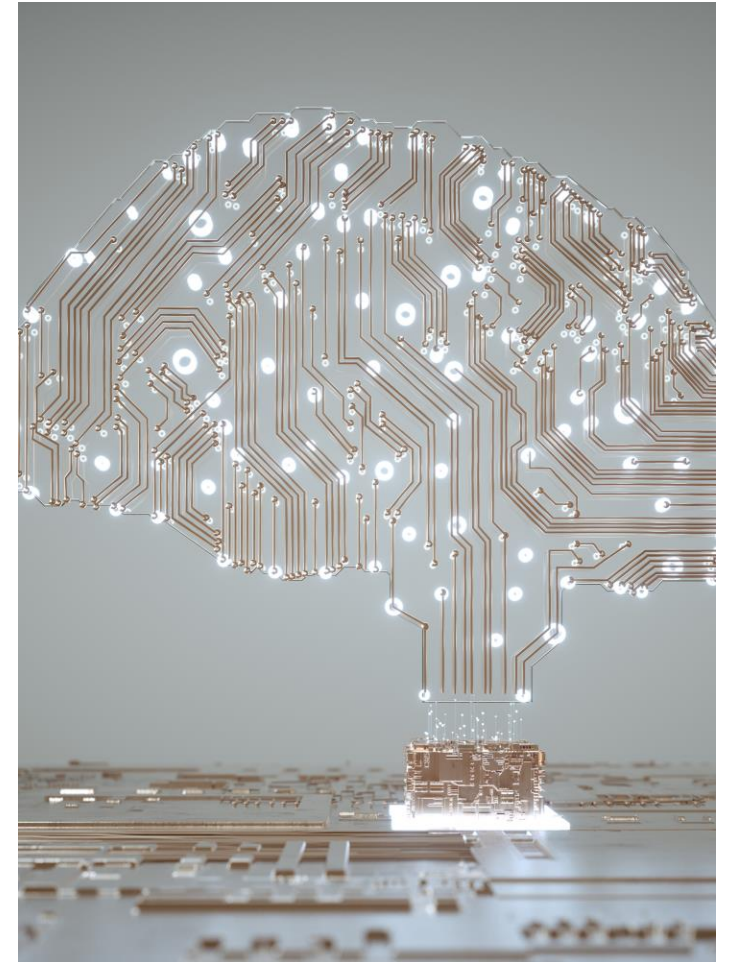
Presented by:



James G. Gatto
Partner | Washington, DC
jgatto@sheppard.com
+1.202.747.1945

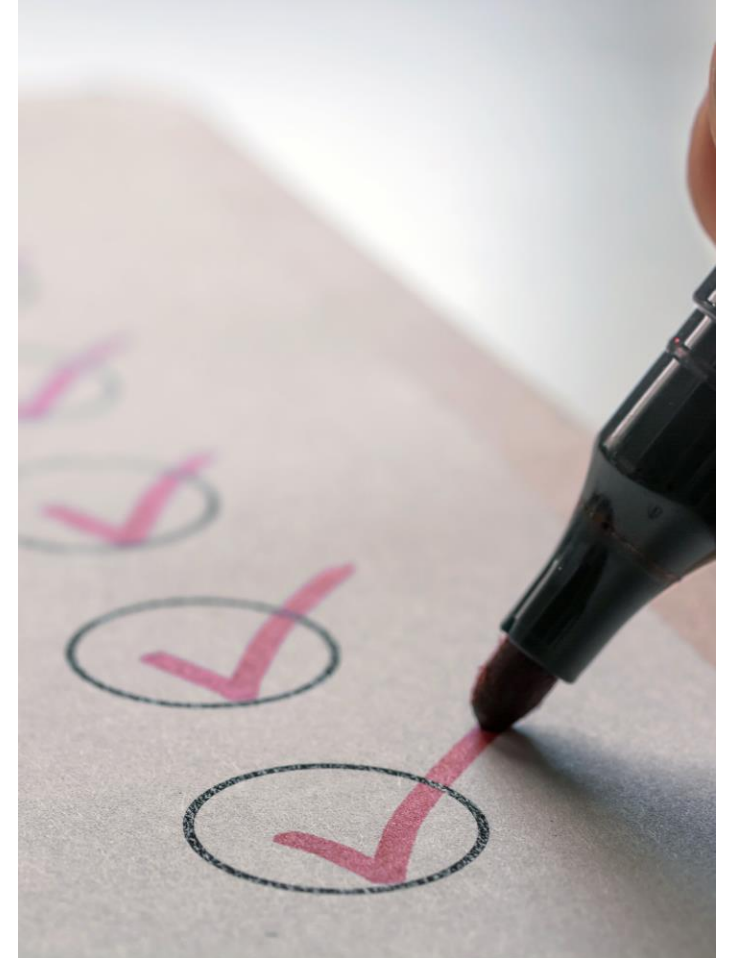
Learning Objectives

- Develop an Understanding of OS AI and Why its Use is Surging
- Learn About OS AI Licenses, Their Compliance Obligations and How to Develop Usage Policies
- Comprehend and Learn How to Manage Legal and Regulatory Risks with OS AI



Topics to Cover

- Why is OS AI surging?
- Overview of OS, AI and OS AI
- Some Advantages and Disadvantages of OS AI
- OS License Issues
- Sample lawsuits
 - *Stockfish v. Chess Base*
 - *Doe v. GitHub*
- OS Policies



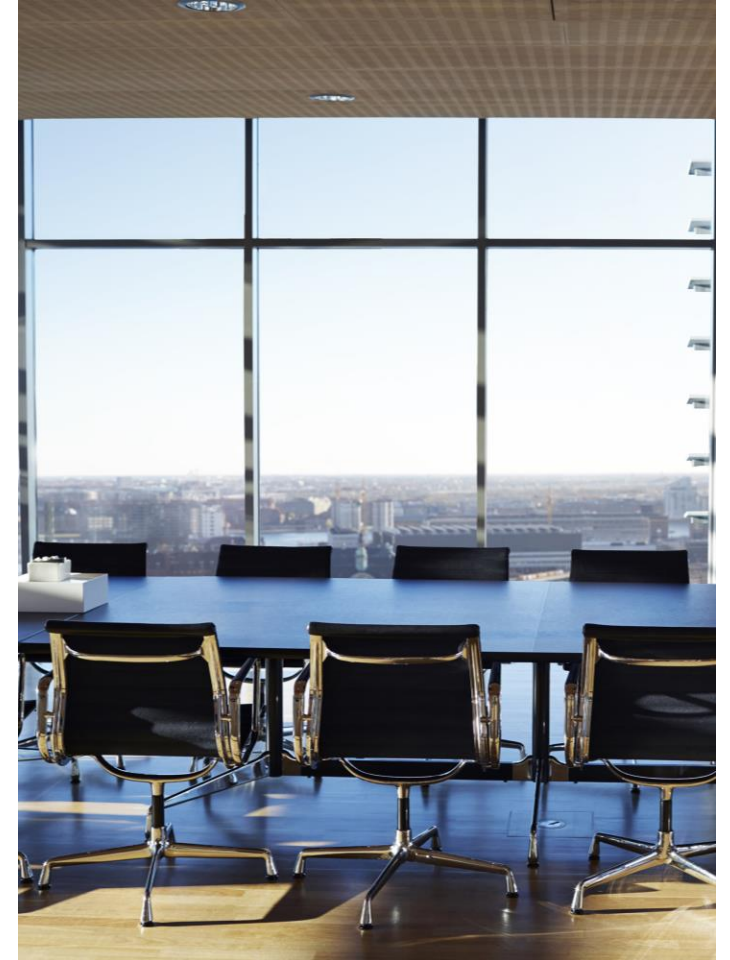
Why is OS AI Surging?

- AI is the most impactful technology in the history of the world – it will have a greater impact than the industrial revolution, computers or the internet
- The leading LLM companies are a duopoly spending 100's of billions of dollars on training “closed” models to produce the most powerful AI tools
- The worldwide AI “race” has geopolitical significance



Why is OS AI Surging?

- Few companies can afford to build their own LLMs
- Many companies pick one LLM provider – creates many risks:
 - Companies pay a lot of money to the LLM provider
 - Vendor lock-in: because the models are “closed,” companies are beholden to how the LLMs are trained
 - Lack of transparency in training data impacts bias and potential infringement
 - Companies often access LLMs hosted by LLM provider via API – this exposes data regarding inputs to and outputs from the models



Why is OS AI Surging?

- Microsoft CEO Satya Nadella has warned that enterprises using proprietary LLM APIs are effectively "paying twice": once in token fees, and again by revealing proprietary knowledge — strategy, code, internal processes — through the prompts, corrections, and feedback needed to make the models useful
- This is one reason that Microsoft is considering greater use of OS models

Microsoft considers replacing ChatGPT and Claude with China's cheaper Kimi K3

Story by Joseph Zeballos-Roig • 9h •  4 min read



 Microsoft mulls cheaper Chinese AI to
© Feature China/Getty Images

The OS “Shot” Heard Around the World

- In Jan. 2025, DeepSeek released a free, open-source LLM allegedly built in just two months for under \$6 million, using less powerful Nvidia H800 chips instead of the most advanced H100s.
- Third-party benchmarks showed it outperformed Meta’s LLaMA 3.1, OpenAI’s GPT-4o, and Anthropic’s Claude Sonnet 3.5 in accuracy across complex problem-solving, math, and coding.
- The follow-up DeepSeek-R1 reasoning model also beat OpenAI’s then latest o1 in several tests.

How China’s new AI model DeepSeek is threatening U.S. dominance

PUBLISHED FRI, JAN 24 2025 8:00 AM EST | UPDATED MON, JAN 27 2025 12:03 PM EST



Jasmine Wu
@JASWU_

Deirdre Bosa

SHARE



What is China's Kimi K3 and why is the US so rattled by it?

JUL 23, 2026

By  [John Liu](#)



KIMI K3

The Open Model That Just Cornered the AI Frontier

2.8T total parameters. Open weights.
Top 3 on the Intelligence Index. #1 on Frontend Code Arena.



INTELLIGENCE INDEX v4.1

1	Claude Fable 5	59.9
2	GPT-5.6 Sol	58.9
3	Kimi K3	57.1
4	Opus 4.8	55.0

FRONTEND CODE ARENA (BLIND TEST)

2	1	3
Claude Fable 5	Kimi K3	GPT-5.6 Sol
1,631	1,679	1,604





 2.8T
TOTAL PARAMETERS

 ~50B
ACTIVE PARAMETERS

 1M
TOKEN CONTEXT

 NATIVE
IMAGE UNDERSTANDING

 6.3x
FASTER at 1M CONTEXT
(Kimi Delta Attention)

 JULY 27
WEIGHTS GO PUBLIC

Microsoft considers replacing ChatGPT and Claude with China's cheaper Kimi K3

Story by Joseph Zeballos-Roig • 9h •  4 min read



 Microsoft mulls cheaper Chinese AI to
 © Feature China/Getty Images

LLM Counter-punches

- Biggest LLM providers have called for a ban on OS models
- Allege the decentralized control prevents efficient remedy if problems arise
- Have also sought to have the government regulate LLMs
- Many allege this is to effect Regulatory Capture

See, [David Sacks Says Dario Amodei's AI Safety Push Is Regulatory Capture — and an Open-Source Ban Is Next](#)

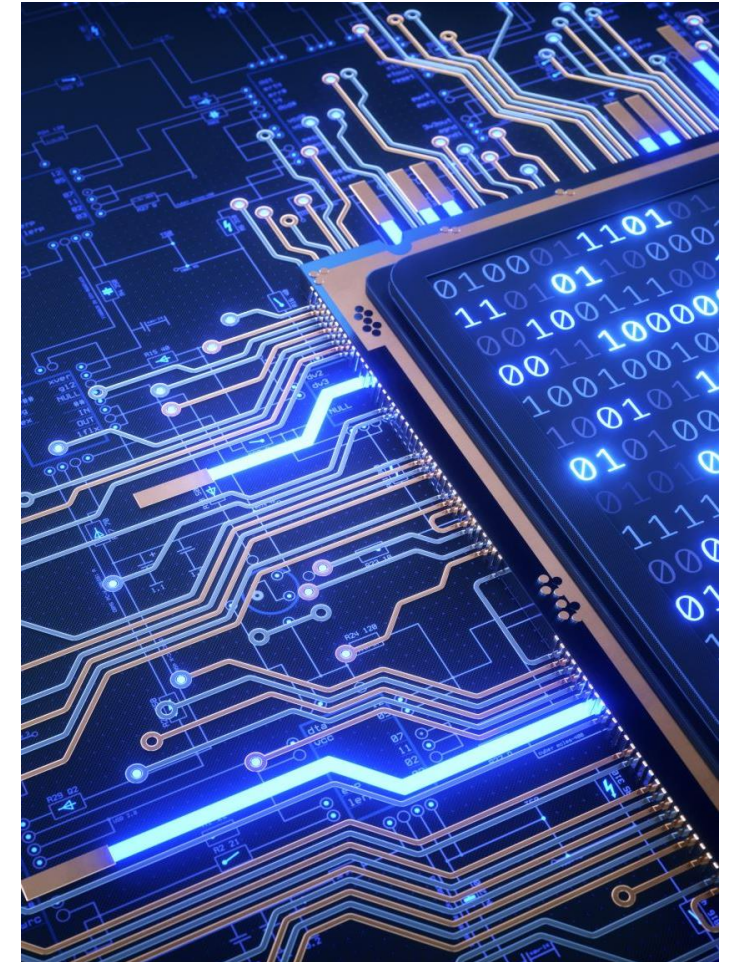
OS Overview

- **OS Licenses**

- Primarily a copyright license to copy, modify and redistribute **source code**; many different licenses with many different terms
- Enables recipients to **use, study, modify, and share** the source code
- The freedom to run the program as you wish, for any purpose
- Most licenses include various compliance obligations
 - e.g., must make the **source code** available to enable exercise of these rights

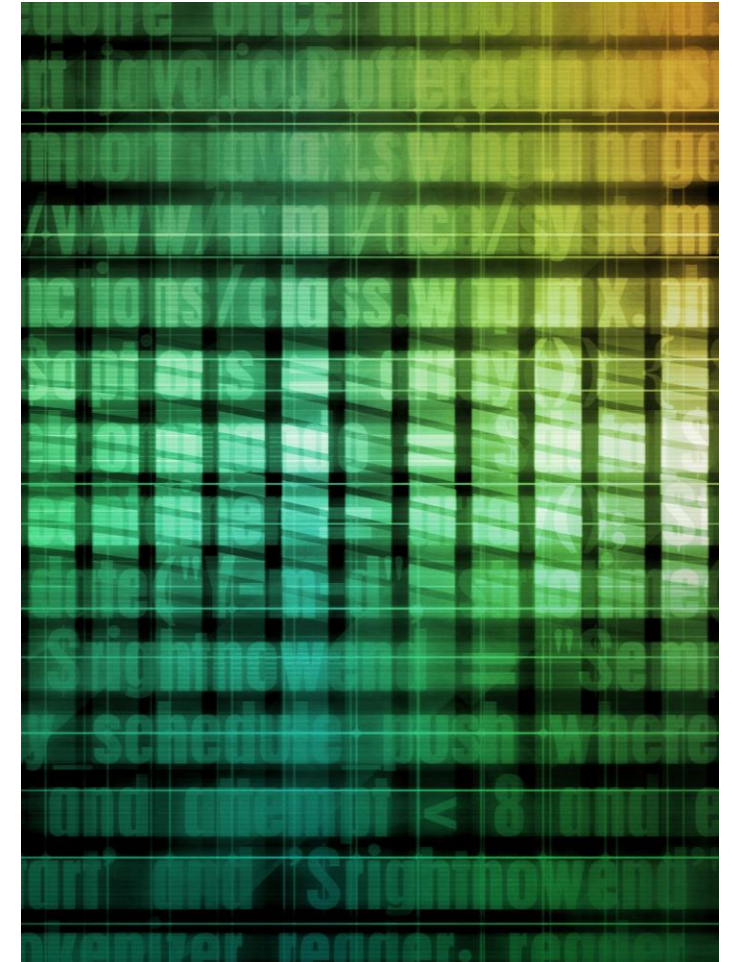
- 2 “standard” definitions – [FSF](#) and [OSI](#)

- **GPL License** - The “Corresponding Source” for a work in object code form means all the **source code** needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities.



Open Source-ish Licenses

- Many products touted as open source are not “fully” open source
 - See, [Open Source-ish! What Defines Open Source and Why it Really Matters in Investments and Acquisitions](#)
- Open Source was originally used with software – the “source” referred to the source code
- Over time, “open” was used in other contexts
- “Open” _____ [fill in the blank]
 - Data
 - Research
 - Hardware
 - Many others
- The problem is that some of these adaptations of open-ness rely on the same licenses originally applied to software – yet the licenses do not perfectly fit some other uses
- “OS AI” is no exception!

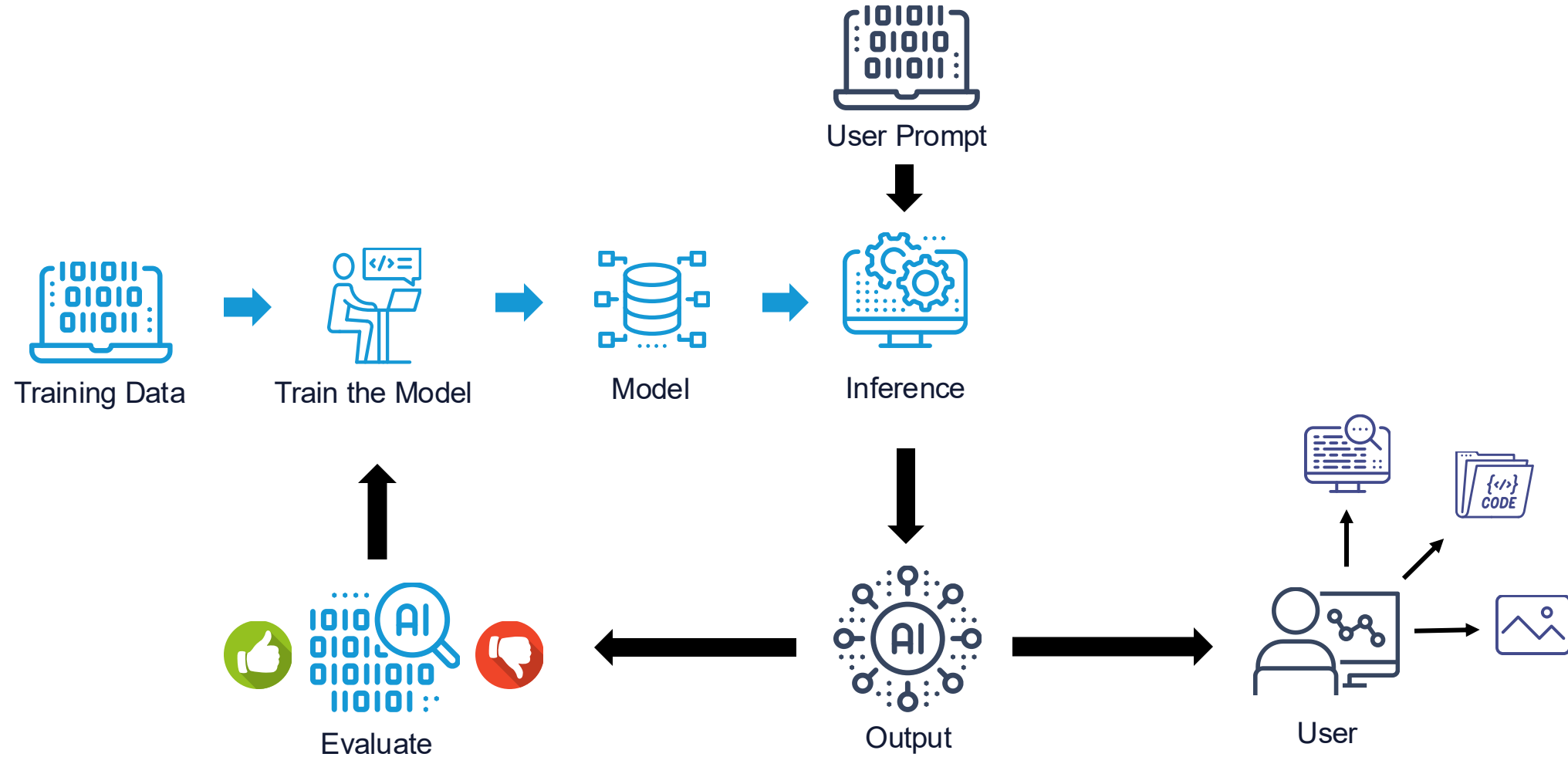


What is AI?

- **AI is a System** of technologies that mimic human intelligence
 - **AI Models** are trained on vast amounts of **data**
 - An **AI model** includes the model architecture/model parameters (including weights)
 - **AI weights** are the set of learned parameters that overlay the model architecture to produce an output from a given input.
 - **Inference code** for running the model
 - May include other elements



Overview of AI Tech



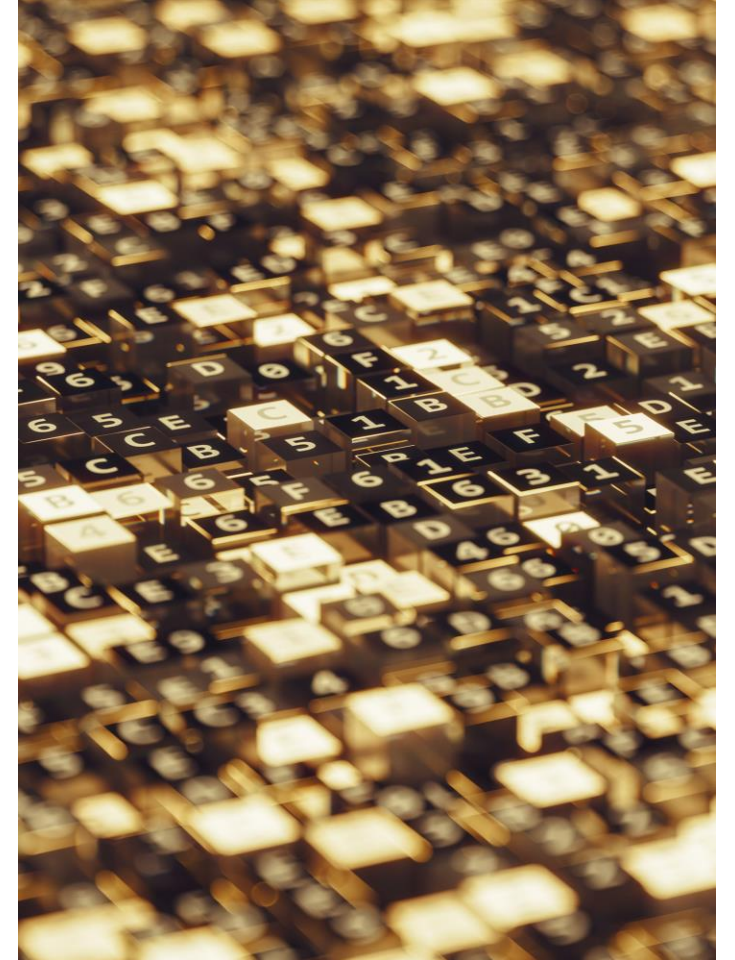


OS License Issues

**Traditional OS licenses were
not designed with AI in mind**

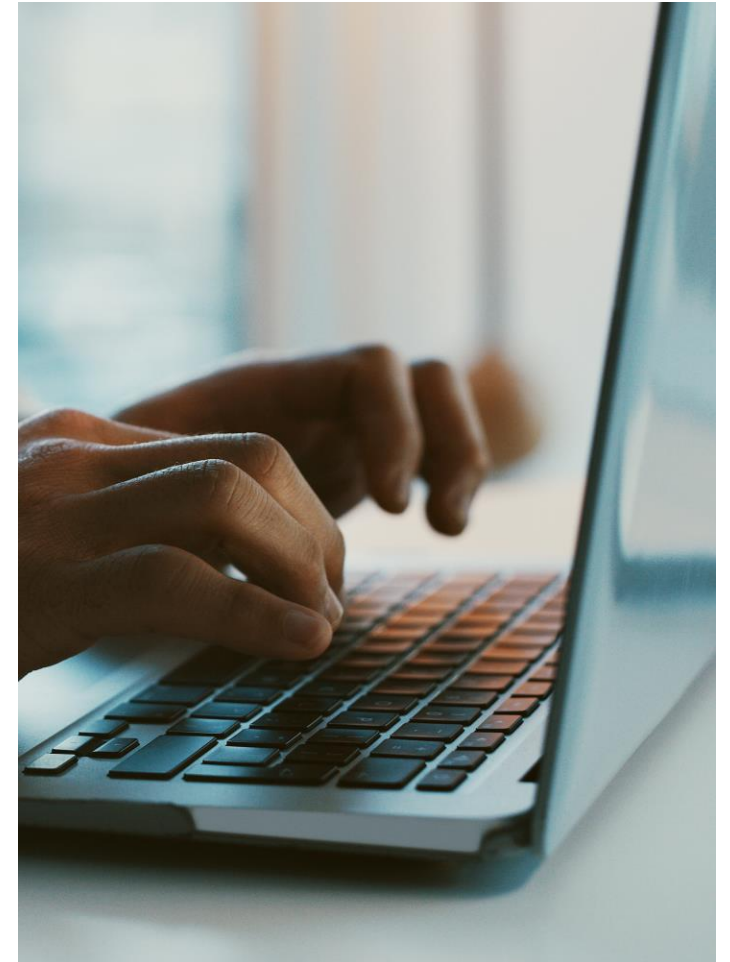
What is OS AI?

- It is like OS software, but more complicated
- Just getting source code is not enough
- Should cover what is needed to **use, study, modify,** and **share** an AI system or element
- No single, standard definition
- [OSI](#) - OS AI definition



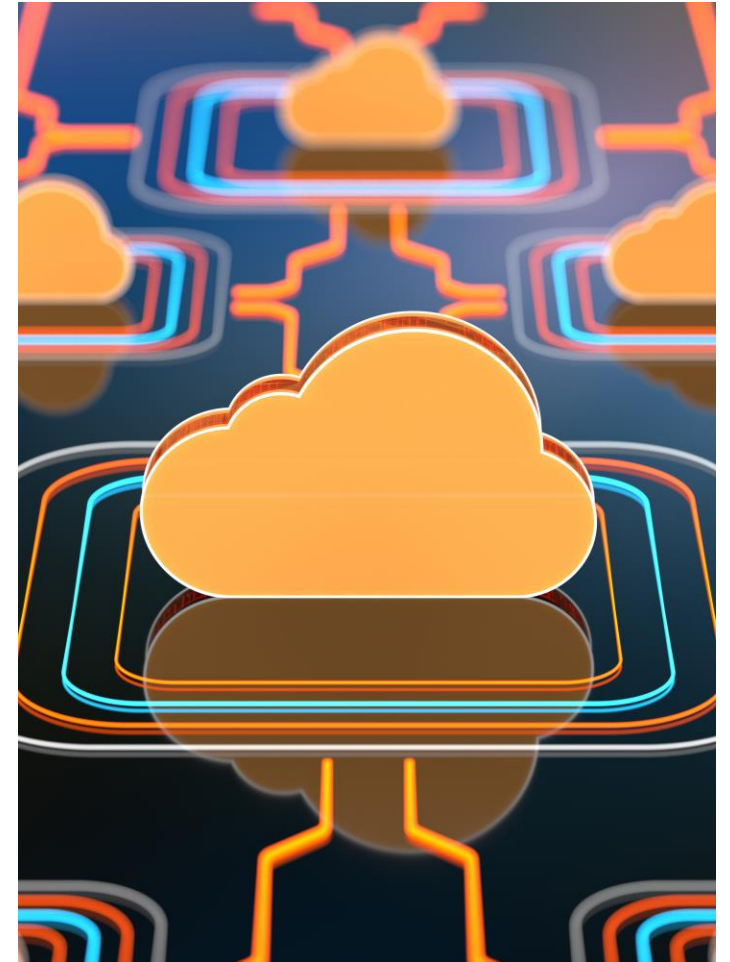
OSI Definition of OS AI

- To be considered Open Source, the requirements are the same, whether applied to a **system, a model, weights and parameters**, or other structural elements.
- OS AI is 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.
- These freedoms apply both to a fully functional system and to discrete elements of a system. A precondition to exercising these freedoms is to have **access to the preferred form to make modifications to the system**.



What's needed for AI model to be OS

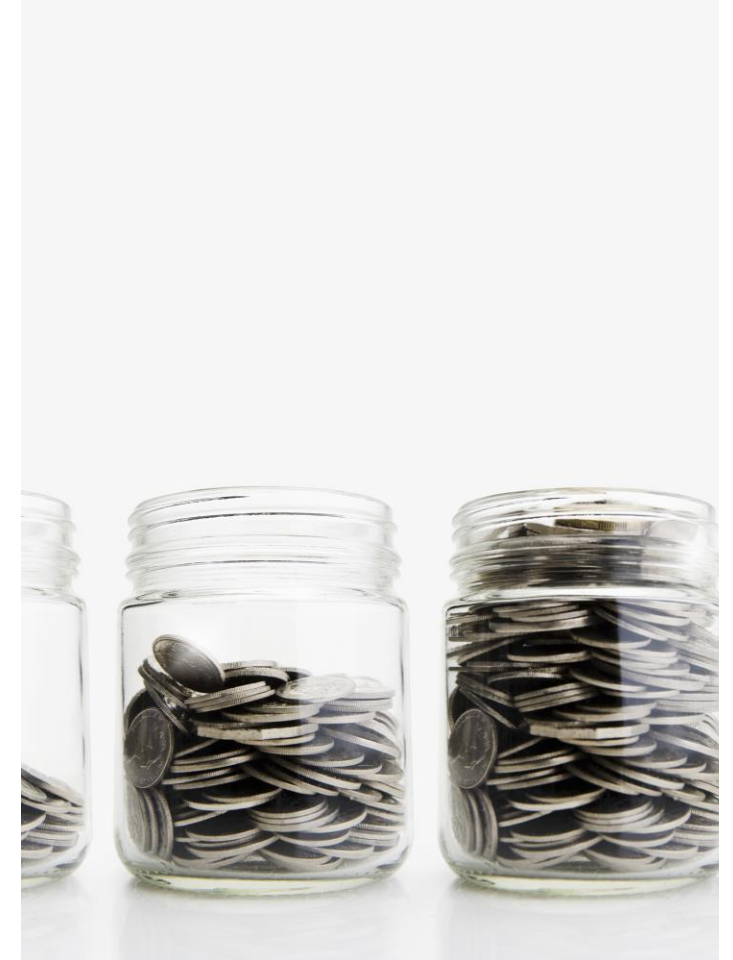
- The preferred form of making modifications to a machine-learning system must include all the elements below:
 - **Data Information:** Sufficiently detailed information about the data used to train the system so that a skilled person can build a substantially equivalent system. Data Information shall be made available under OSI-approved terms.
 - **Code:** 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.
 - **Parameters:** The model parameters, such as weights or other configuration settings.
 - *See appendix slides for more detail*



Many AI Model Licenses are Open Source-ish

KIMI K3 License Restrictions

- If the Licensee or any of its affiliates operates a Model as a Service business, and the aggregate revenue of the Licensee and its affiliates exceeds 20 million US dollars (or the equivalent in other currencies) in total over any consecutive 12 months, the Licensee must enter into a separate agreement with Moonshot AI before using the Software or its derivative works for any commercial purpose.
- If the Software (or any derivative works thereof) is used for any of the Licensee's commercial products or services that have more than 100 million monthly active users, or more than 20 million US dollars (or equivalent in other currencies) in monthly revenue, "Kimi K3" must be prominently displayed on the user interface of such product or service.

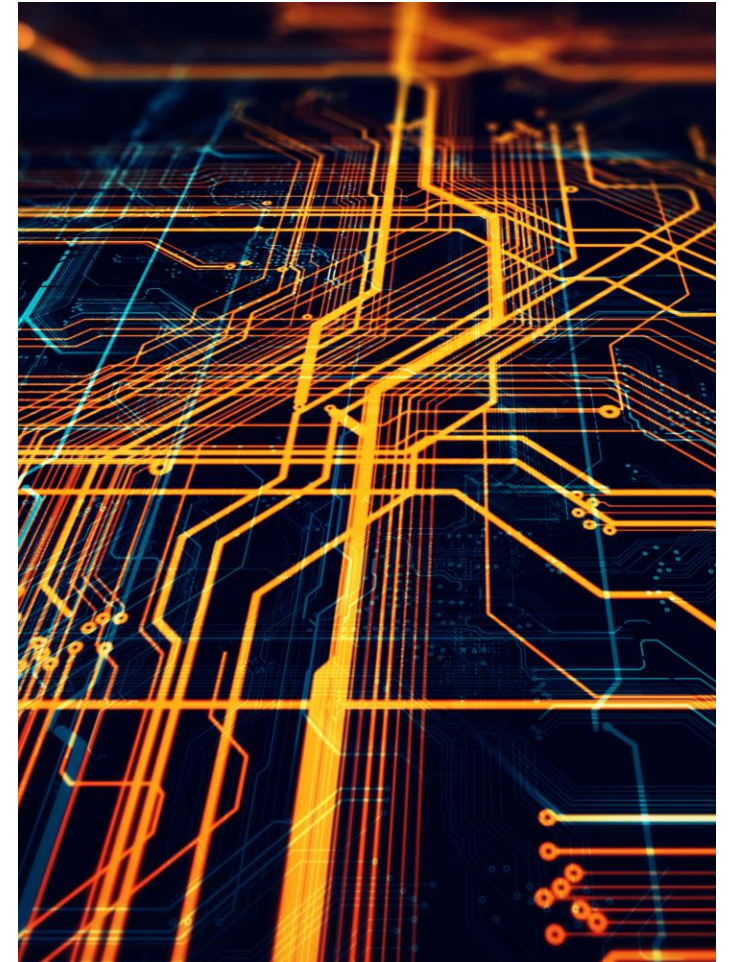


Some Advantages and Disadvantages of Open AI

Advantages	Disadvantages
“Free”	Must operate and maintain the model and supporting systems
Can fine-tune (train on your own data) without expenses of training a model from scratch	Potential liability/infringement risks based on how data was used
Can run on your own platform without using third party API - avoids exposing their prompts/data	May need data licenses
Understand the training data (if fully open)	Potentially increased regulatory issues as a model “provider” under EU AI Act

Open Weight Models

- Make available the weights (and parameters) needed to **run** the model, but typically does not release the training data, detailed information about that data, or the training algorithms
- Enables user to fine-tune the model on their own data to modify or add weights and parameters
- Does not require provider to disclose all its trade secrets and know-how ... But some laws/proposed laws require transparency on how and on what a model is trained
- Does not let the user fully understand, reproduce, or adapt the underlying model - including any inherent biases - because there's no access to the training data or training algorithms
- The lack of visibility into training data can make it impossible for a developer to know whether the training process used copyrighted works without permission
- False claims of openness because only some components are released as fully open (“AI Washing”)



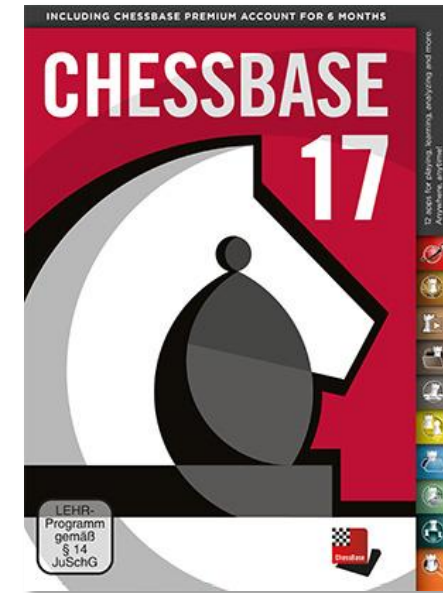


Sample Lawsuits involving Open Source and AI

Stockfish v. ChessBase

Background

- Stockfish licensed a chess engine under GPL-3.0
- ChessBase used the chess engine in its products and licensed under proprietary licenses
- Stockfish repeatedly accused ChessBase of violating the GPL-3.0 license
- Stockfish **terminated** the license for ChessBase but ChessBase continued to sell the products
- Stockfish sued ChessBase in a German court in July 2021
- The parties reached a [settlement](#) in November 2022



GPLv3 License Obligations

Conveying Non-Source Forms

- You may convey a **covered work** in object code form provided that you also convey the machine-readable **Corresponding Source** under the terms of this License
- **Corresponding Source** means “all the **source code** needed to generate, install, and (for an executable work) run the object code and to modify the work...”

Termination Provision

- You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License



Legal Issues

Issues

- Did ChessBase fail to comply with the GPL-3.0 license?
- Whether the licensor can unilaterally terminate the GPL-3.0 license?
- How should neural networks and weights be addressed in the context of GPL-3.0 licensed software?



Are Neural Network Weights “Corresponding Source”

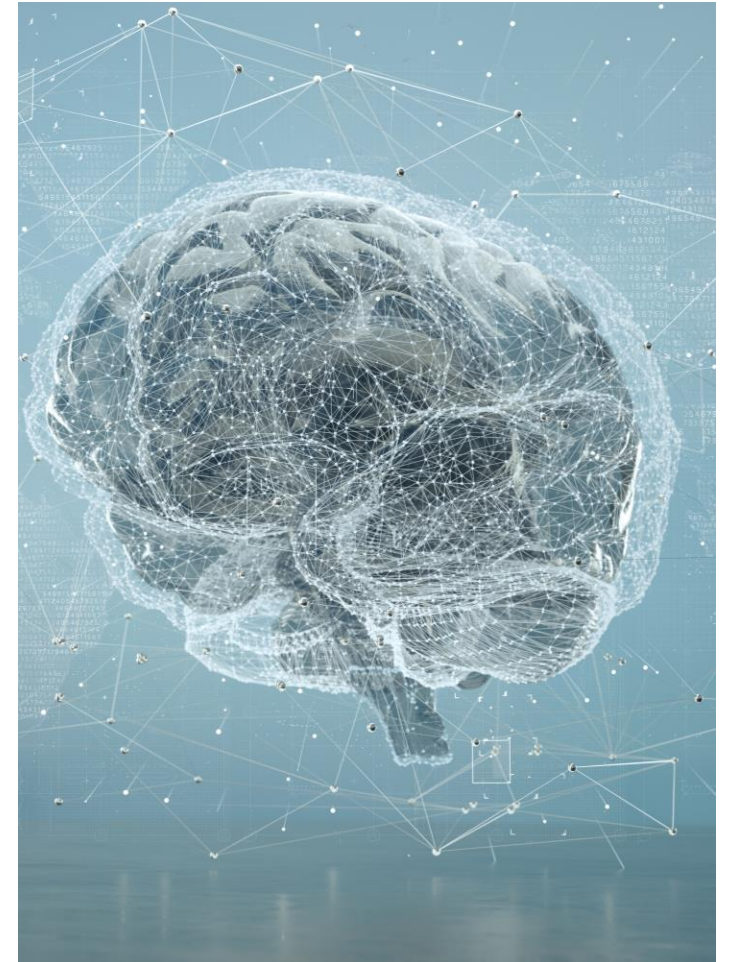
Are neural network weights source code? Or data?

Typically, under GPL 3.0 - Corresponding Source primarily includes code but not data

But no case has decided this specific point

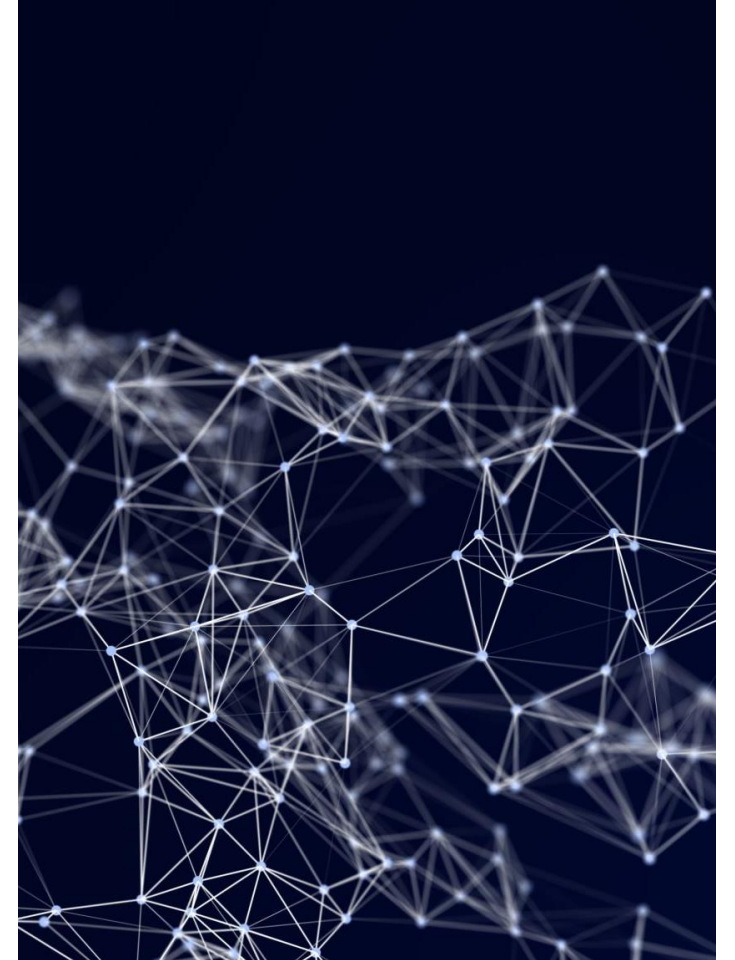
Settlement

- Any **neural networks** offered by ChessBase for use with Stockfish that are **included** in the compilation or **dynamically loaded** at runtime **to initialize the data structures** and logic of the Software must be subject to GPL-3.0 or a compatible license



Neural Network Weights

- One key issue not answered or expressly addressed in the settlement:
- Whether **proprietary neural network weights** must be released in order to comply with the “Corresponding Source” code requirement of GPL 3.0?



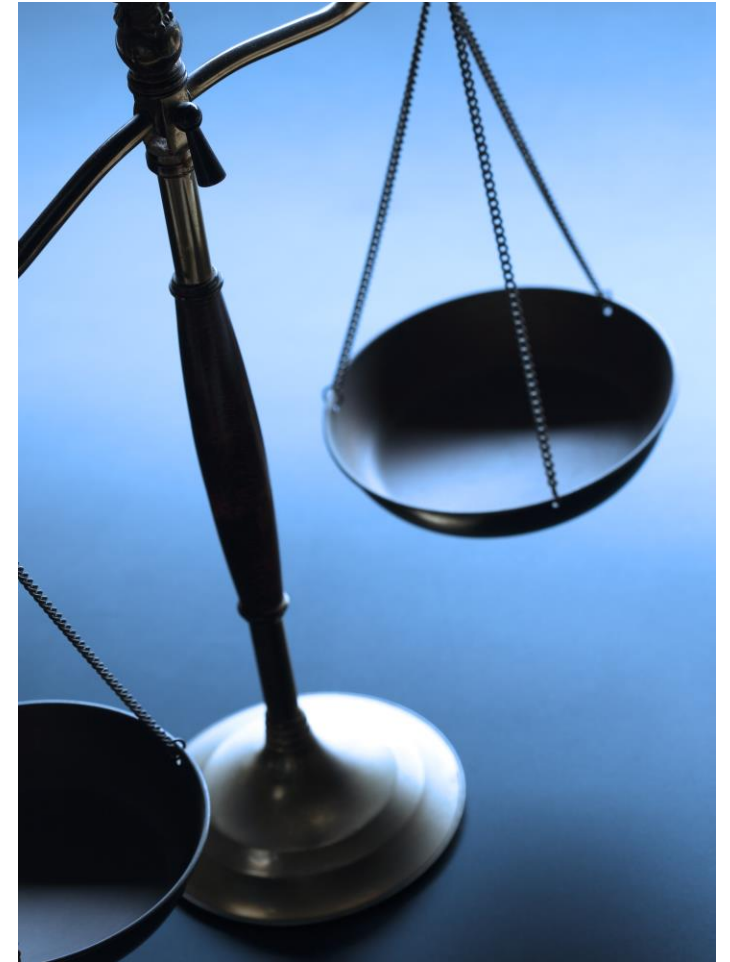
AI Code Generators

Overview of AI Code Generators

- AI models trained on software (typically OS software)
- Generates code for developers
- One of the most high-valued AI tools

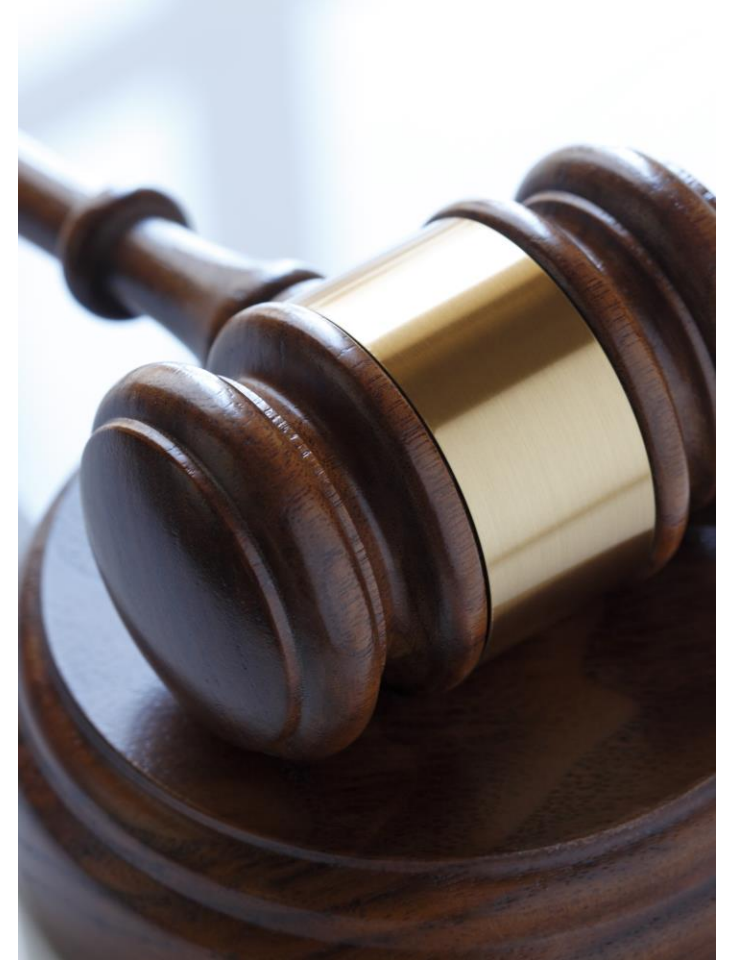
AI-Based Code Generators Raise Various Legal Issues

1. Does training AI models using OS code constitute infringement?
2. Does use of AI-generated code by developers creating a new software application require that application to be licensed under an OS license and its source code to be made available (“Tainting”)?
3. Do OS license compliance conditions or restrictions apply?



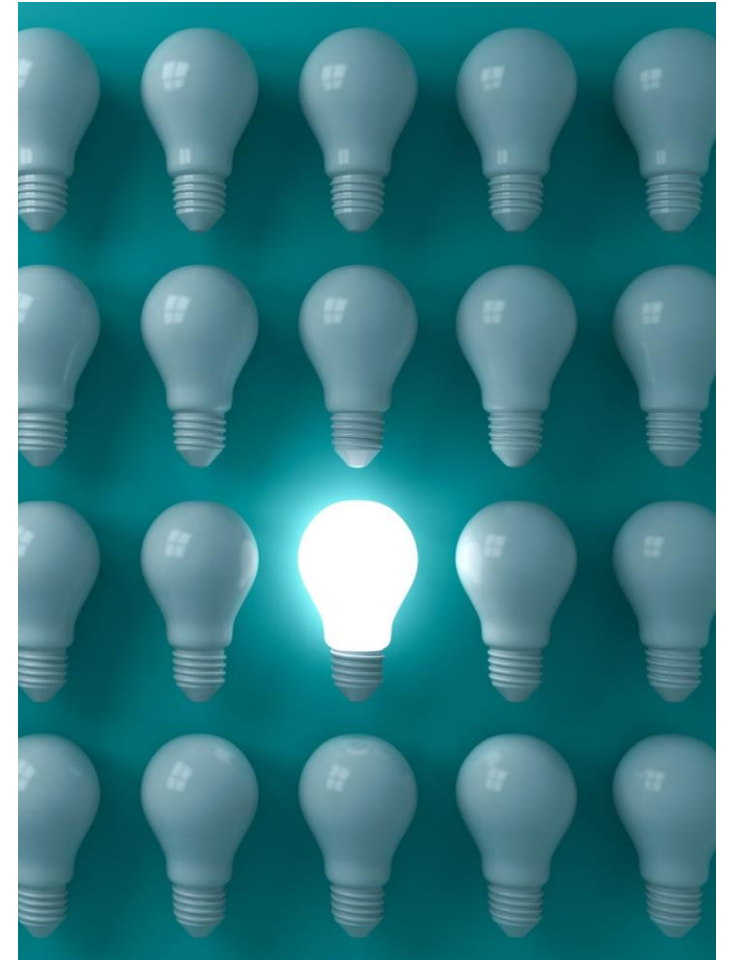
Doe v. Github

- Section 1202 DMCA violations (for stripping copyright-related information from input data)
- Relates to AI code generators trained largely on OS code
- 9th circuit just affirmed dismissal of DMCA claim finding that the **output** was **not a copy** of the work bearing the CMI, but rather a **new work**.
- Plaintiff waived arguments regarding the input
- This will not be the final case on this issue!



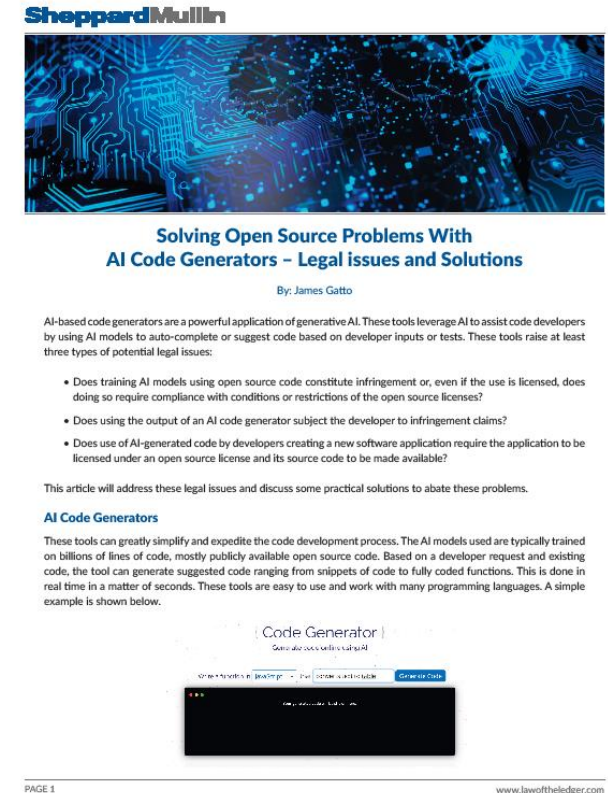
Potential Technical Solutions to Mitigate OS Legal Risks with AI Code Generators

1. **Filters** to prevent the output of problematic code
2. Code **Reference** tools to flag problematic output
3. Code **Scanning** tools to assist developers with OS compliance



OS Policies

- Companies Need OS Policies
 - [Open-Source Policies — Why You Need Them And What They Should Include](#)
- OS Policies Need Updating to Address AI
 - Use of OS AI
 - Release of OS AI
 - Use of AI Code Generators - Link to article shown here
- Compliance with all applicable requirements
 - Must understand what compliance obligations per applicable license
 - Ensure procedures to comply



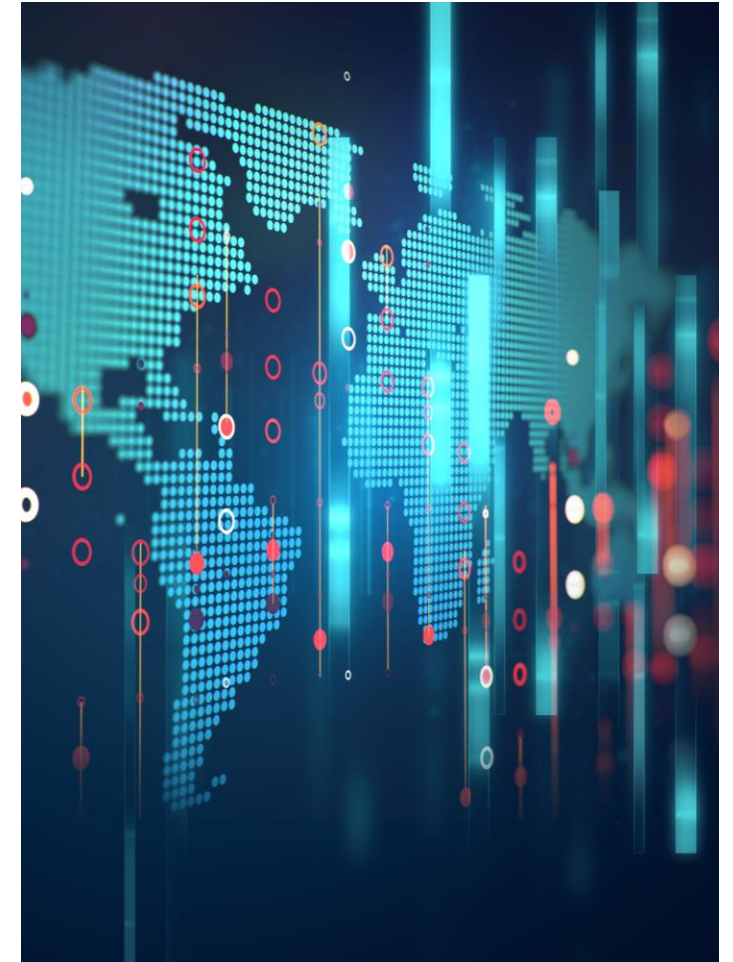
Regulatory Issues with OS AI

Sample Regulatory Issues

- Some laws require disclosure of training data – use of open weight model may be problematic to comply
- [NIST - AI RISK MANAGEMENT FRAMEWORK](#) (not law) but referenced as standard requires: Transparency, Fairness, Validity, etc. – may not be able to demonstrate compliance with Open Weight models
- May be liable for use of a model that was trained on data for which there was no right to use (specifically with PII, HIPPA, BIPA)

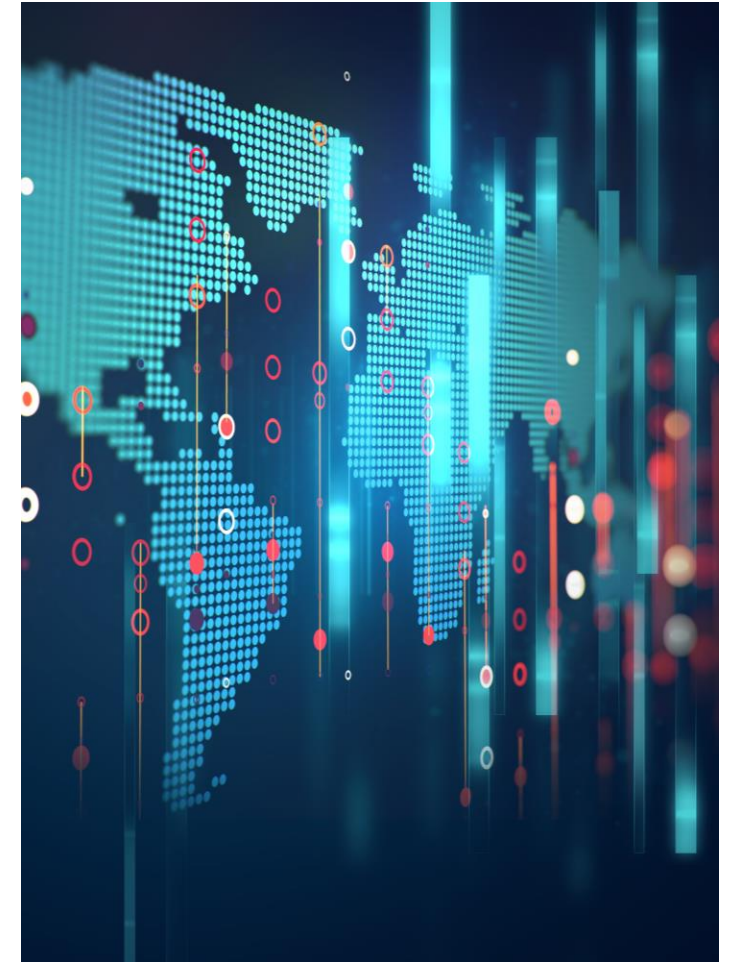
Sample Regulatory Issues

- Open-sourcing AI models can implicate export controls if models incorporate controlled technology, were developed with export-controlled information, or have military or intelligence applications.
- SBOM – often need to disclose AI/software, including OS
- May be country of origin issues
- Section 1532 of the Fiscal Year 2026 National Defense Authorization Act (NDAA) bars DoD contractors from using certain “covered artificial intelligence” — including AI systems developed by China-based DeepSeek — when performing on DoD contracts
- US government even banned Anthropic
- technology that may be regulated under ITAR or the EAR



Sample Regulatory Issues

- GSA proposed AI-specific safeguarding clause in the federal acquisition system — GSAR 552.239-7001, most recently revised in June 2026 as *Basic Safeguarding of Data within Large Language Model Artificial Intelligence Systems* — which would flow down data-handling, disclosure, and use-rights obligations for GSA Multiple Award Schedule contracts whenever a large language model processes government data.
- Technology that may be regulated under ITAR or the EAR
- [Commerce Takes on AI: Recent Developments from BIS on AI](#)



Thank You!

Keep Up To Date

See articles on
Artificial Intelligence



Subscribe
to blog for updates





Questions & Answers



James G. Gatto

AI Team Leader | **Partner**

jgatto@sheppard.com

[Connect on LinkedIn](#)

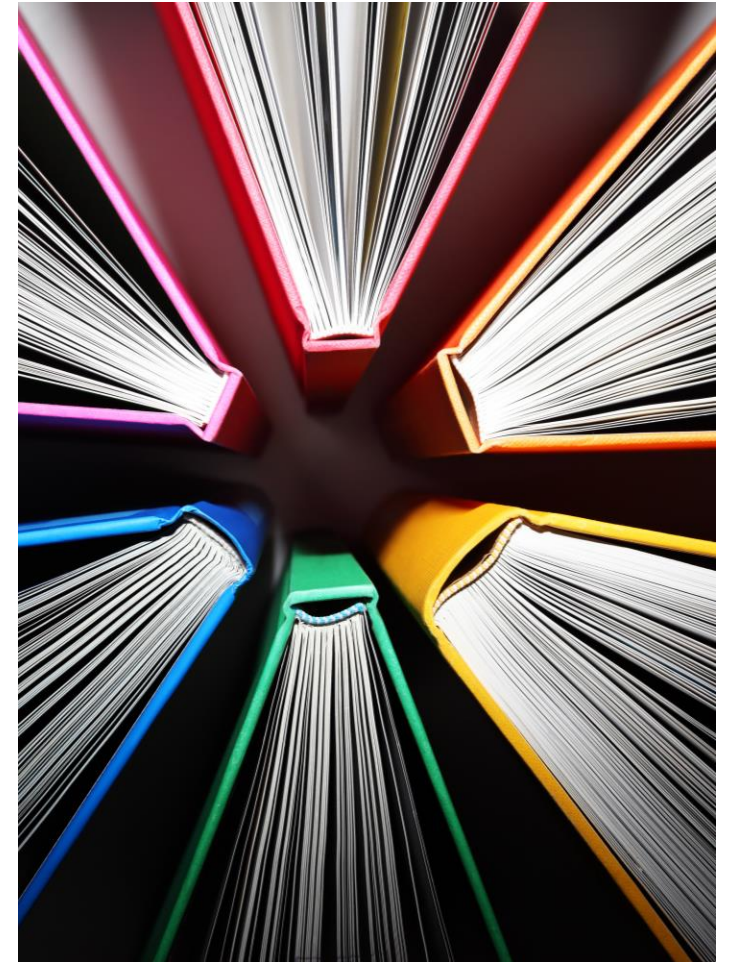
+1.202.747.1945

Appendix Materials

What's needed for AI model to be OS

Data Information:

- The complete description of all data used for training, including (if used) of unshareable data,
- Disclosing the provenance of the data, its scope and characteristics,
- How the data was obtained and selected
- The labeling procedures, and
- Data processing and filtering methodologies;
- A listing of all publicly available training data and where to obtain it; and
- A listing of all training data obtainable from third parties and where to obtain it, including for fee.

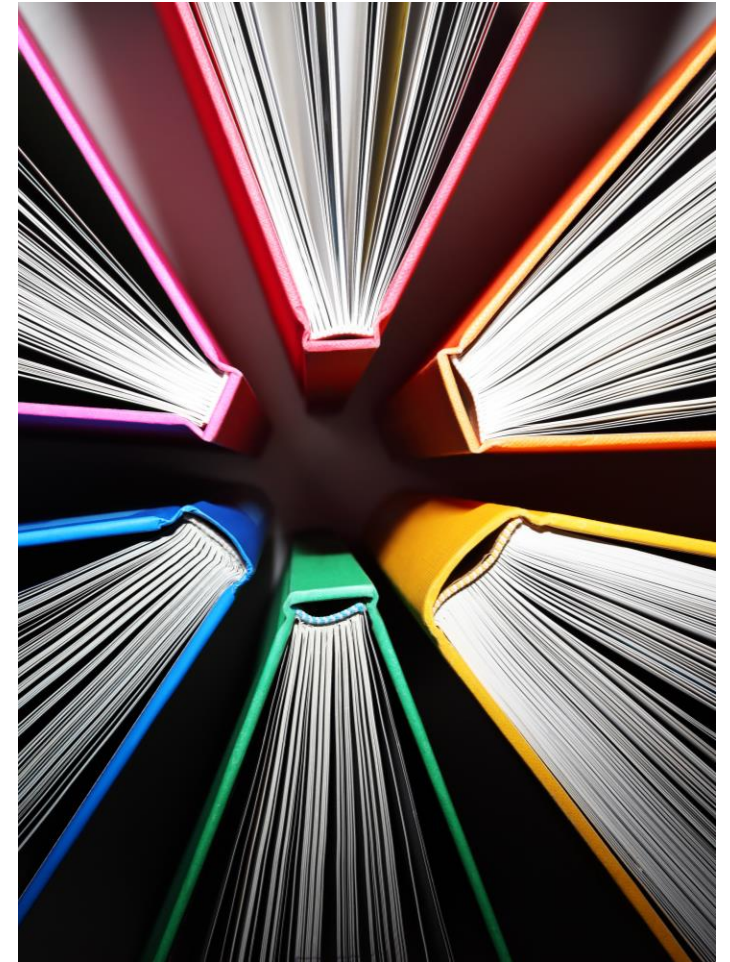


What's needed for AI model to be OS

Code must include:

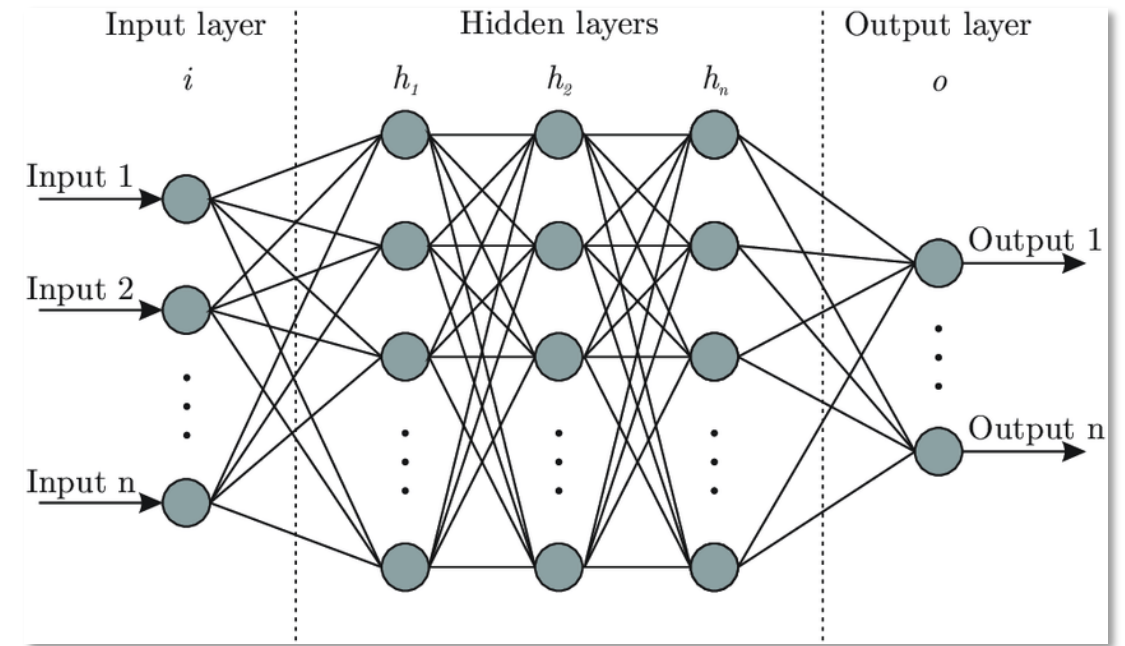
- Code used for processing and filtering data
- Code used for training including arguments and settings used
- Validation and testing, supporting libraries like tokenizers and hyperparameters search code
- inference code
- model architecture

Parameters includes checkpoints from key intermediate stages of training as well as the final optimizer state.



Neural Networks

- A neural network is a series of nodes (or [neurons](#)) each node having a set of inputs, **weights**, and a bias value
- The neural network contains a series of hidden layers which apply transformations to the input data
- Input layer takes the input signals and passes them to the next layer
- It is within the nodes of the hidden layers that the weights are applied



Node

- The input to a node gets multiplied by a weight value and the output is observed or passed to the next layer in the neural network
- **Weights** refer to the strength of the connection - affects the amount of influence a change in the input will have upon the output
- **Bias** represents by how much the predictions differ from their intended value; makes up the difference between the function's output and its intended output
- Weights can change over time

