Background

Trends in Model Reuse on Hugging Face 🤗

- **Pretraining is costly**
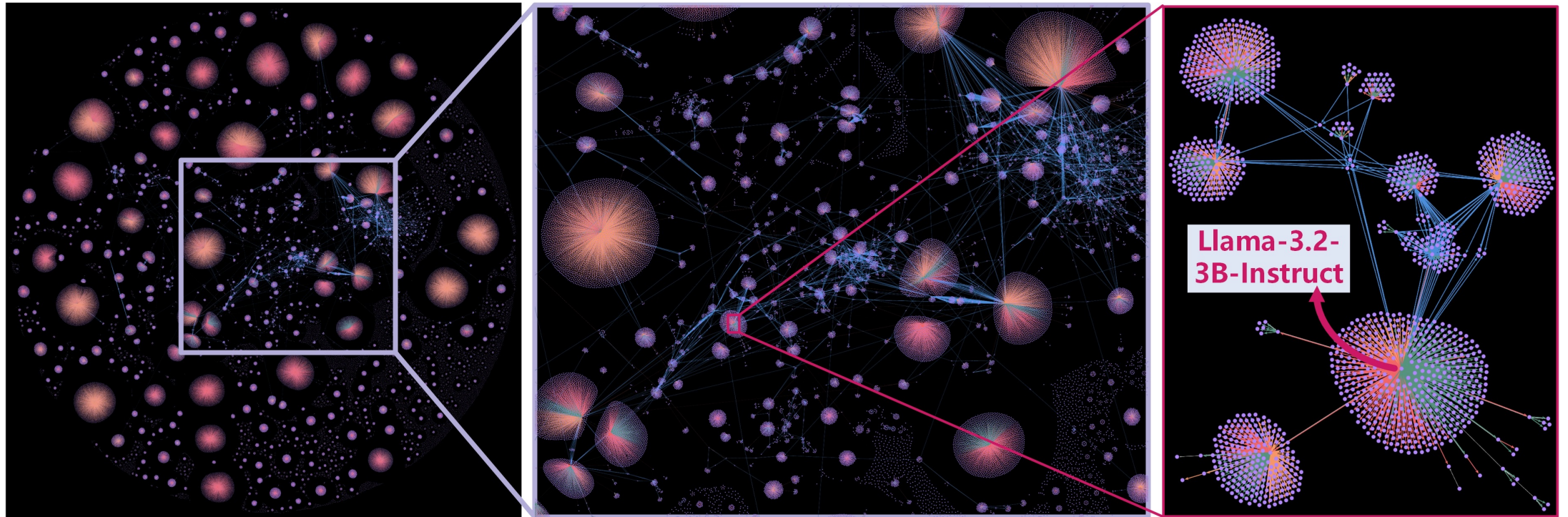 - 10^25 FLOPs
  - Billions of tokens
  - Several months with thousands of GPUs

- **Model reuse is cheap**
  - PEFT: LoRA (0.1% – 1%)
  - 1 GPU can run
  - 1.4 million models on HF

Background

- The Collaboration Chain of ML Models

Visualization of Model Dependencies on 🤗
Finetune (50%), Adapter (33%), Quantization (10%), and Merge (4%)
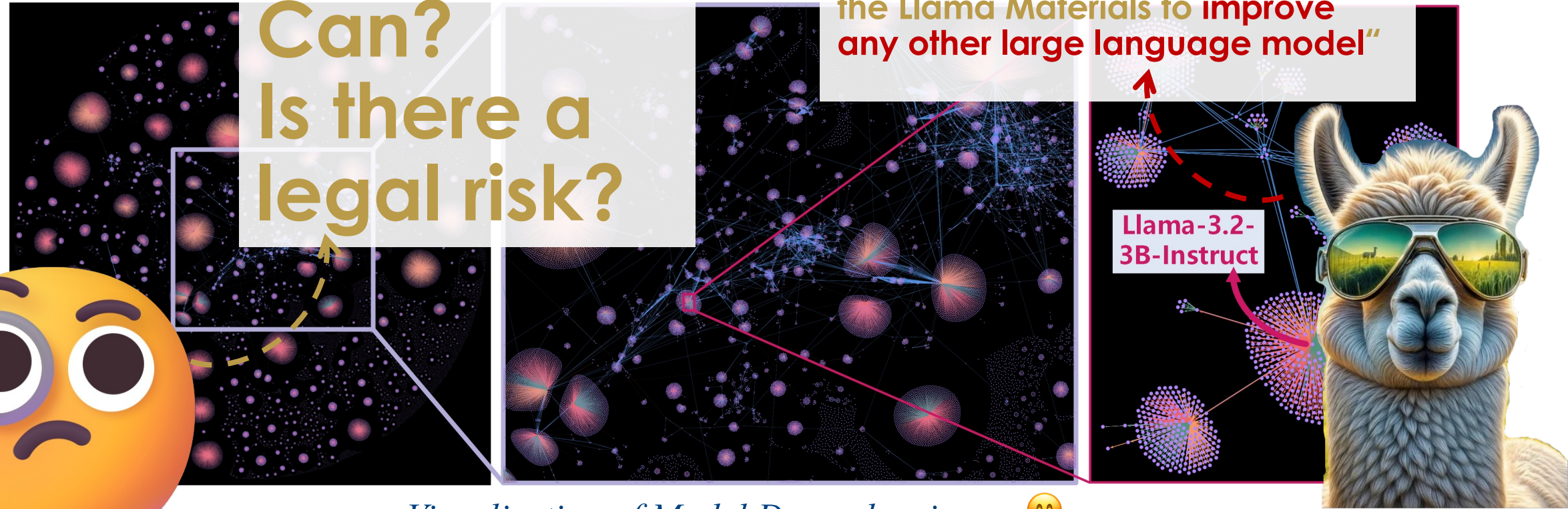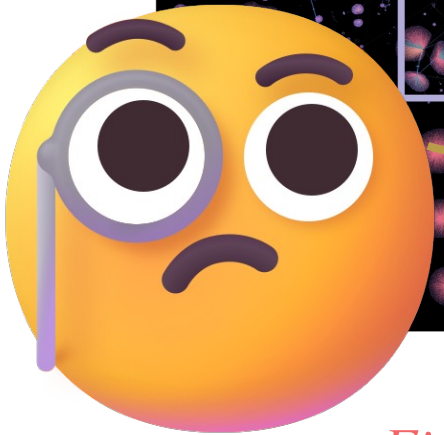
Background

- The Collaboration Chain of ML Models

Can?
Is there a legal risk?

Llama-3.2-3B-Instruct

Visualization of Model Dependencies on 🤗
Finetune (50%), Adapter (33%), Quantization (10%), and Merge (4%)

Type of Noncompliance

- **License Mismatch:** Model published under OSS/free-content licenses

- **License Proliferation:** Copyleft violations and inconsistencies in copyleft terms

- **License Conflict:**

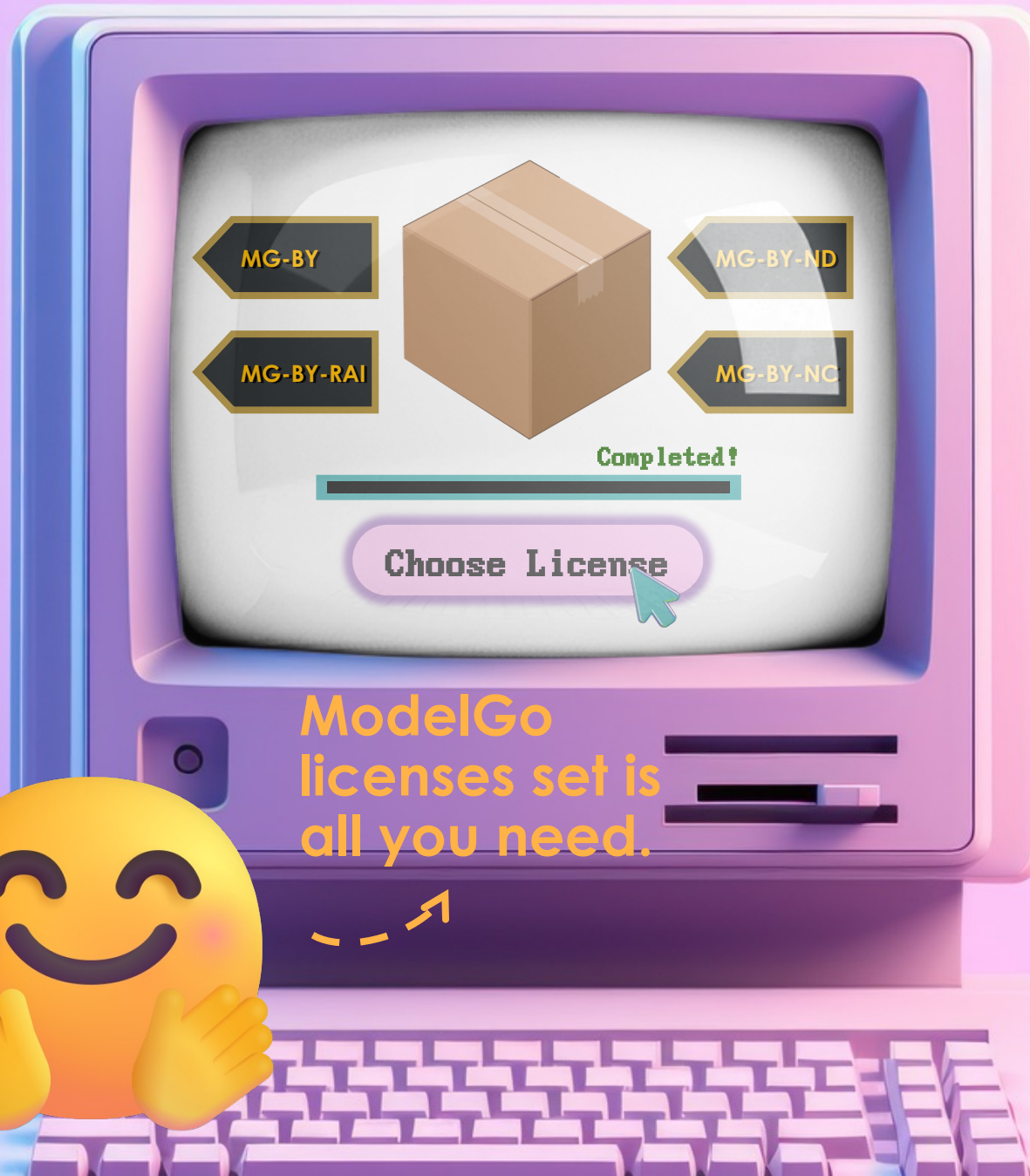| Conflict Type (Abbrv.)	License	Conflict-Causing Clause
No Derivate (ND)	CC-BY(-NC)-ND-4.0	Section 3 Clause a: *You do not have permission under this Public License to Share Adapted Material.*
FSF-freedom (FSFf)	GPL-3.0, AGPL-3.0	Clause 10: *You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License.*
CC-freedom (CCf)	CCs (excl. CC0-1.0)	Section 2 Clause 5B: *You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.*
Llama2(/3) Exclusive (La2E/La3E)	Llama2(/3) License	Clause 1.v: *You will not use the Llama Materials or any output or results of the Llama Materials to improve any other large language model (excluding Llama 2(/3) or derivative works thereof).*

Results

Noncompliance	# of NCM	# of CTM	Finetune	Merge	Quantization	Adapter
● License Mismatch	123,707 (81.8%)	151,302	n/a	n/a	n/a	n/a
○ Copyleft-style Terms	26,679 (63.5%)	41,019	11,011 (35.8%)	1,453 (4.7%)	4,775 (15.5%)	13,546 (**44.0%**)
● Copyleft Violation	365 (38.5%)	949	296 (**81.1%**)	32 (8.8%)	37 (10.1%)	n/a
● ND	36 (100%)	36	1 (2.8%)	12 (33.3%)	23 (**63.9%**)	n/a
● CCf	76 (2.1%)	3,675	17 (20.0%)	67 (**78.8%**)	1 (1.2%)	n/a
● FSFf	98 (45.8%)	214	84 (85.7%)	0	14 (14.3%)	n/a
● La2E-license	1,000 (38.4%)	2,602	182 (18.1%)	29 (2.9%)	331 (32.9%)	465 (46.2%)
● La2E-merge	11 (55.0%)	20	n/a			
● La3E-license	966 (24.7%)	3,904	518 (47.2%)	(7.6%)	(27.6%)	195 (17.6%)
● La3E-merge	81 (56.6%)	143	n/a	81 (100%)	n/a	n/a

A Quagmire That's Hard to Remedy

LICENSING OPTIONS

BY - Attribution

NC - NonCommerical

ND - NoDerivatives

RAI - Responsible AI

SA - ShareAlike
       (Copyleft)

MG-BY

MG-BY-RAI

MG-BY-ND

MG-BY-NC

Completed!

Choose License

ModelGo licenses set is all you need.

KEY FEATURES

Specifically Designed for Model Licensing

Feasible Licensing Options Easy to Learn & Use

Clearly Defined and Safer Model Reuse

Our Views

1. **Model Reuse Can Lead to License Violations.**

2. **Violations May Proliferate to Downstream Derivatives.**

3. **Remediation is Difficult due to Unknown Dependencies.**

Alternative Views

1. **No need to comply with model licenses because the copyrightability of models remains legally uncertain.**

2. **No need to comply with a license/term of use if you have not explicitly entered to it or signed.**

3. **Distillation does not create a derivative work, since the output is not considered part of the original model.**

For more interesting topics in ML model governance, welcome to our related tutorial at VLDB 2025!

ML-Asset Management: Curation, Discovery, and Utilization

https://ml-assets-management.github.io