Free Software Foundation's criteria for free machine learning applications

Zoë Kooyman, FSF executive director
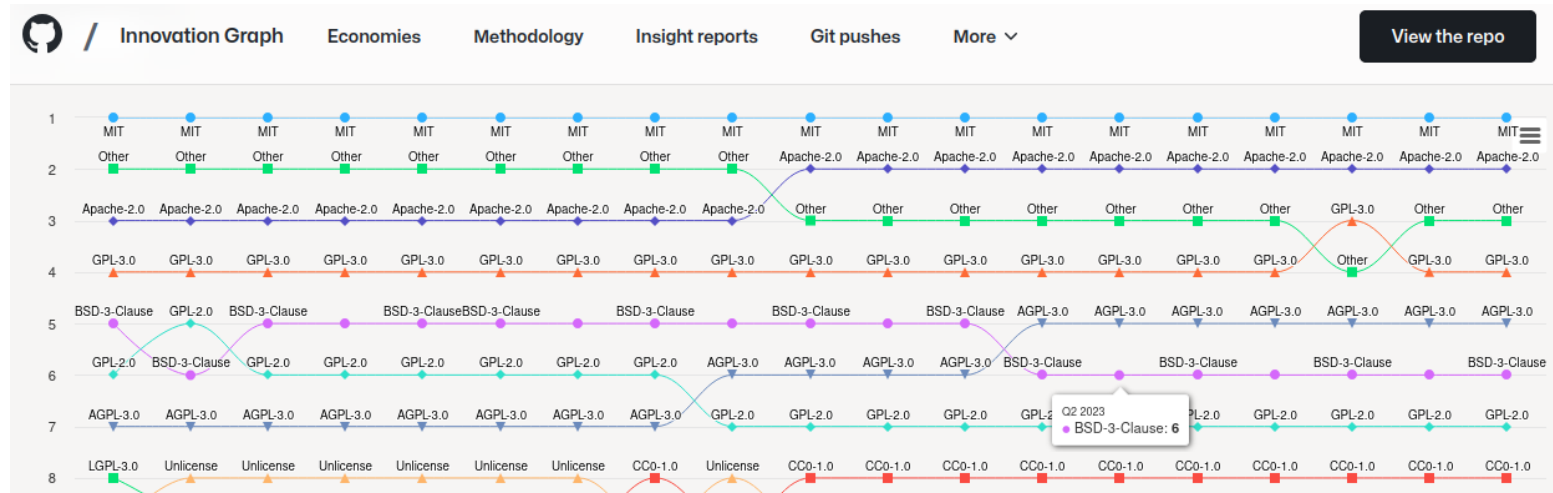Krzysztof Siewicz, FSF licensing and compliance manager

Disclaimer

- A work in progress

- Not machine learning experts

- We don't have all answers (yet?)

- We are here to listen and collect your questions and thoughts

Free Software Foundation

- Nonprofit with worldwide mission to promote computer user freedom.

- Founded in 1985 by Richard Stallman to support the GNU Project

- Maintains the Free Software Definition, wrote and steward of the GNU General Public License(s)
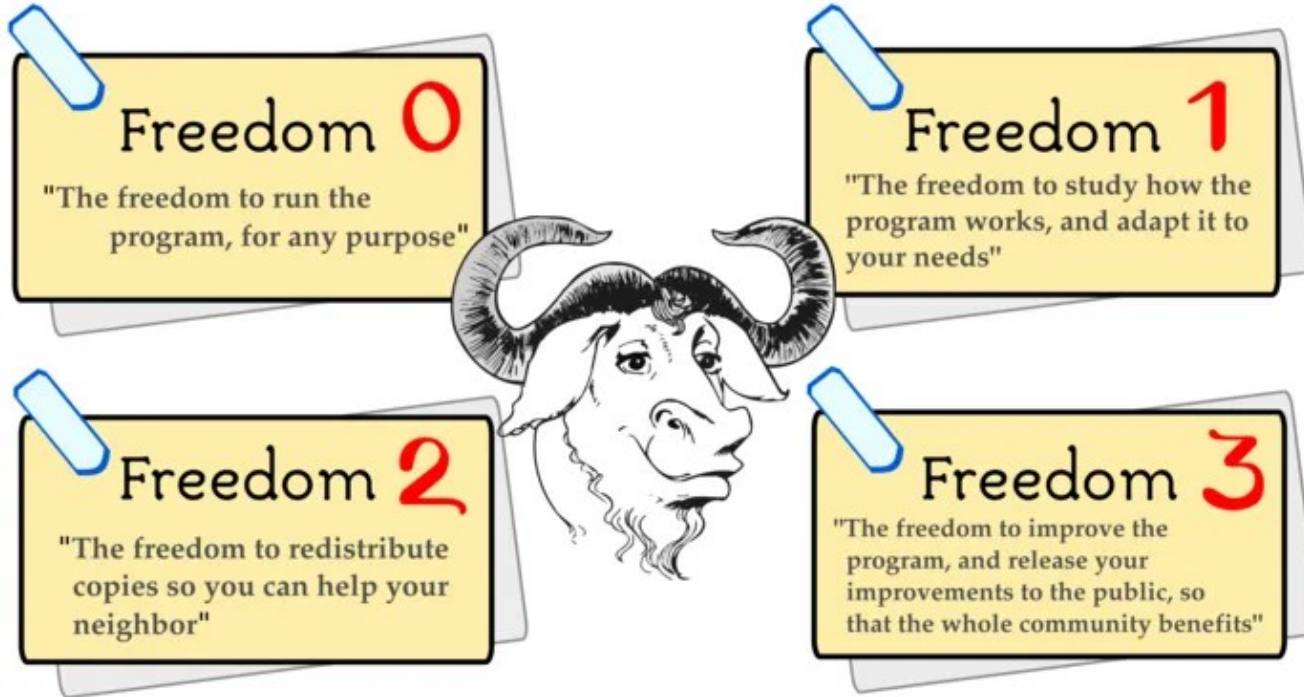
Why free software is important

- Lets you control your computing

- Heart of software development, runs almost everywhere now

- People choose freedom

- GNU GPL and GNU AGPL among the 5 most popular licenses on GitHub

https://innovationgraph.github.com/global-metrics/licenses#license-rankings

4

The Four Freedoms

Freedom **0** — "The freedom to run the program, for any purpose"

Freedom **1** — "The freedom to study how the program works, and adapt it to your needs"

Freedom **2** — "The freedom to redistribute copies so you can help your neighbor"

Freedom **3** — "The freedom to improve the program, and release your improvements to the public, so that the whole community benefits"

Access to source code is precondition to 4 freedoms

40 years of free software!

FREE SOFTWARE
FOUNDATION

- FSD Published by Richard M. Stallman in February 1986

- Changed around 1990 to 3 freedoms and freedom 0 was added (the freedom to run)

- 3 versions of the GNU GPL: GPL, AGPL, LGPL

- Still the most used copyleft license, AGPL on the rise

- "Despite significant challenges during the pandemic, the market has surged by over 60% since 2020. Still, that is nothing compared to growth projections for the following years."

- "Machine learning is set to skyrocket by a remarkable 535% and hit over half a trillion dollar value by 2030."

  https://altindex.com/news/machine-learning-drives-ai-growth

Why bother with machine learning

- computing done with ML is on the rise

  - ML is an agent between users and software

  - ML is used to develop software

  - ML is used to do software tasks

- Freewashing attempts: ML as giving users a degree of freedom, while it is actually designed not to

- A lot of people are looking at FSF for guidance

Machine learning and its freedom-related challenges

FSF FREE SOFTWARE FOUNDATION

- MLs do not have "intelligence" (no knowledge or understanding)
- Generative MLs (including LLMs) are "stochastic parrots"
- Some MLs can reach conclusions about specific subjects, e.g. recognize patterns
- ML include software and non-software elements
- No equivalent of "source code" for non-software elements
- Various non-software elements important for studying or modifying ML, e.g., training and testing data, parameters for training, like the model structure, initialization, learning rate functions, batch sizes, regularization

Towards criteria for freedom-respecting machine learning applications

- All software and non-software elements play a role but training data seems the most debated topic

- We should be aspirational

- We should not lower standards

- Four freedoms

- Focus on studying and modifying.

What we know

- Users should be given all they need to control their computing done with ML

- Training data should be free

- Not just incremental training, but also retraining from scratch

Questions to answer

- Is it morally justifiable to use an ML application that has no available training data?

- Give us your questions! Now … or email to info@fsf.org to make contact

12

What's next?

- Discussions on technical and philosophical details with experts

- We'll listen, discuss, and work towards a statement of criteria

- We don't have deadlines, but we do have ambition to set a standard we can live towards

Thank you for your attention