



Connect

Empowering Mankind

Open Source and the Final Frontier of AI

Anneli Sara Banderby
Senior Solution Architect

32 years old

Bachelor in Biomedicine

Master in Toxicology

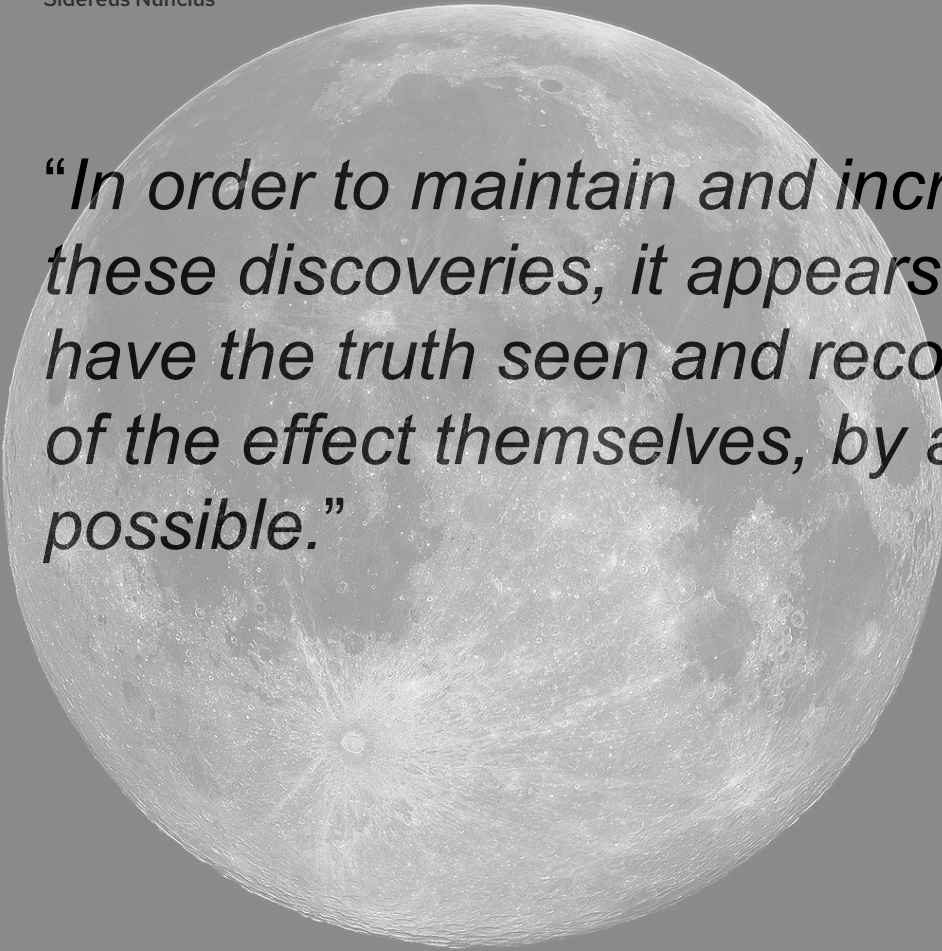
Statistics, machine learning
and bioinformatics



© Patrik Grönqvist

Agenda

- Why is Open Source relevant in the field of AI?
- How can we collaborate with the InstructLab methodology?
- What does it solve and not solve?
- Call to action



“In order to maintain and increase the renown of these discoveries, it appears to me necessary... to have the truth seen and recognized, by the means of the effect themselves, by as many people as possible.”

- Galileo Galilei



Tasks

Libraries

Datasets

Languages

Licenses

Other

Multimodal

Image-Text-to-Text

Visual Question Answering

Document Question Answering

Video-Text-to-Text

Any-to-Any

Computer Vision

Depth Estimation

Image Classification

Object Detection

Image Segmentation

Text-to-Image

Image-to-Text

Image-to-Image

Image-to-Video

Unconditional Image Generation

Video Classification

Text-to-Video

Zero-Shot Image Classification

Mask Generation

Zero-Shot Object Detection

Text-to-3D

Image-to-3D

Image Feature Extraction

★ Keypoint Detection

Models 1,063,266

Full-text search

↑↓ Sort: Trending

nvidia/Llama-3.1-Nemotron-70B-Instruct-HF

Text Generation • Updated 5 days ago • ↓ 60.3k • ♥ 847

nvidia/Llama-3.1-Nemotron-70B-Instruct

Updated 4 days ago • ↓ 4k • ♥ 334

SWivid/F5-TTS

Text-to-Speech • Updated 7 days ago • ↓ 71.5k • ♥ 533

deepseek-ai/Janus-1.3B

Any-to-Any • Updated about 22 hours ago • ↓ 2.37k • ♥ 270

black-forest-labs/FLUX.1-dev

Text-to-Image • Updated Aug 16 • ↓ 1.15M • ⚡ • ♥ 5.67k

mistralai/Ministral-8B-Instruct-2410

Updated 3 days ago • ↓ 36 • ♥ 224

rain1011/pyramid-flow-sd3

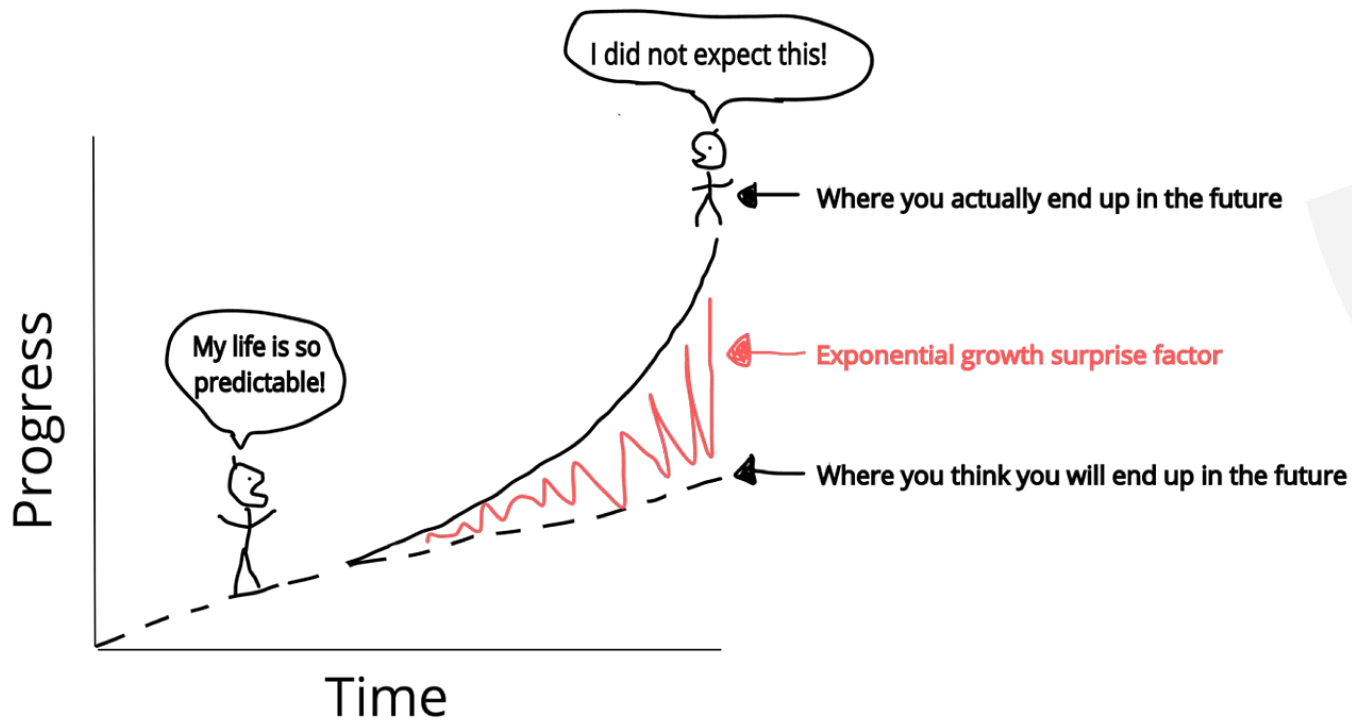
Text-to-Video • Updated 10 days ago • ♥ 684

rhymes-ai/Aria

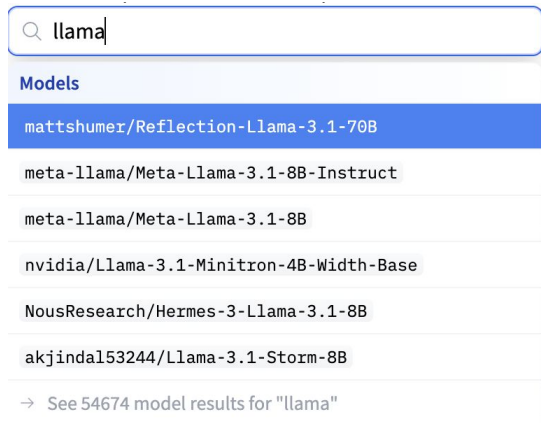
Image-Text-to-Text • Updated about 2 hours ago • ↓ 14.4k • ♥ 510



The future is exponential



Open source development

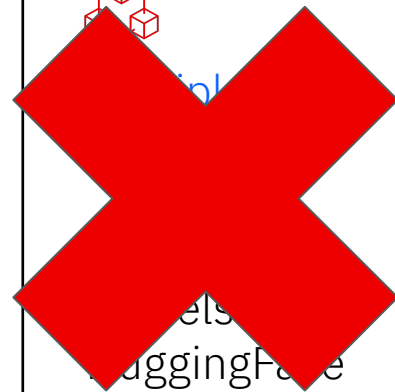


No way to
contribute to
one model

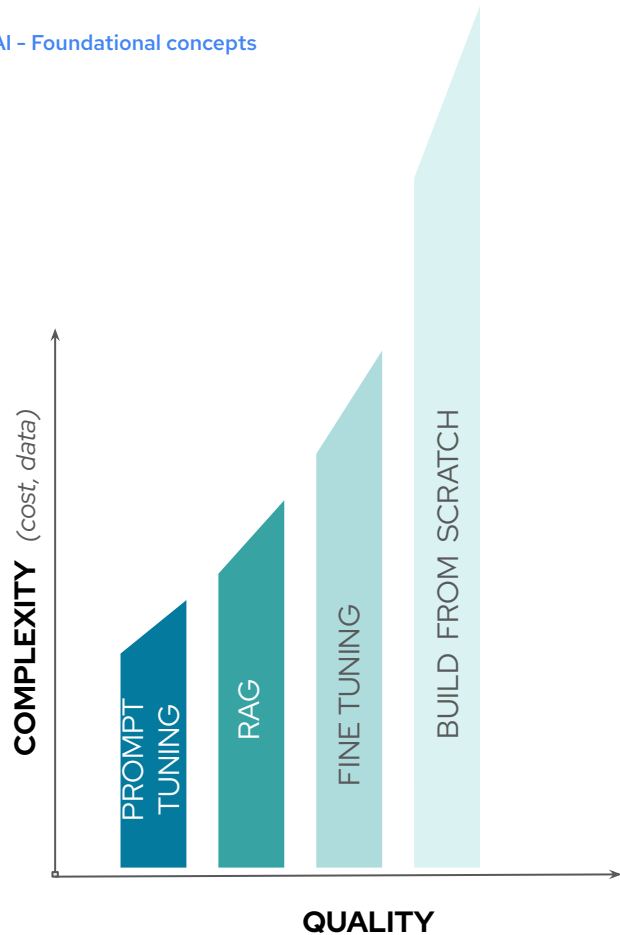


Not community
driven

No way to share
LLMs
enhancements



There is a clear need for AI
developers to collaborate
and update existing
Language Models without
creating multiple clones



The balancing act of using foundation models

Foundation models will still need more work to be useful

- ▶ **Prompt tuning**
- ▶ **Retrieval augmented generation (RAG)**
- ▶ **Fine tuning foundation models**
- ▶ **Training a Foundation Model from scratch**

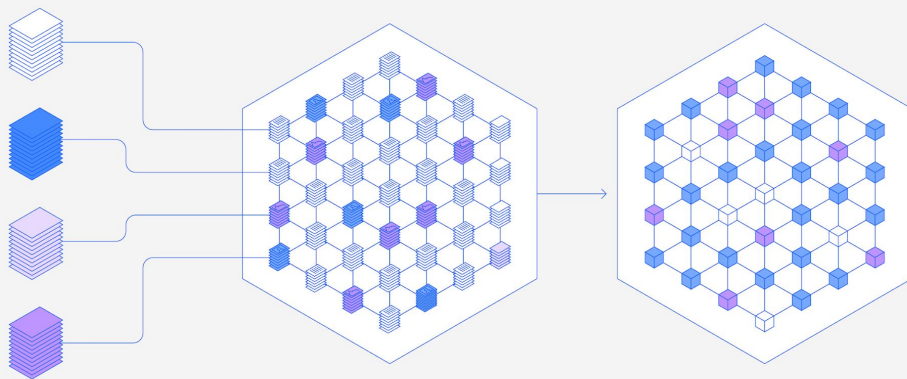
InstructLab: A new methodology!

There is a clear need
for AI professionals to
collaborate and
update existing
Language Models
without creating
multiple clones

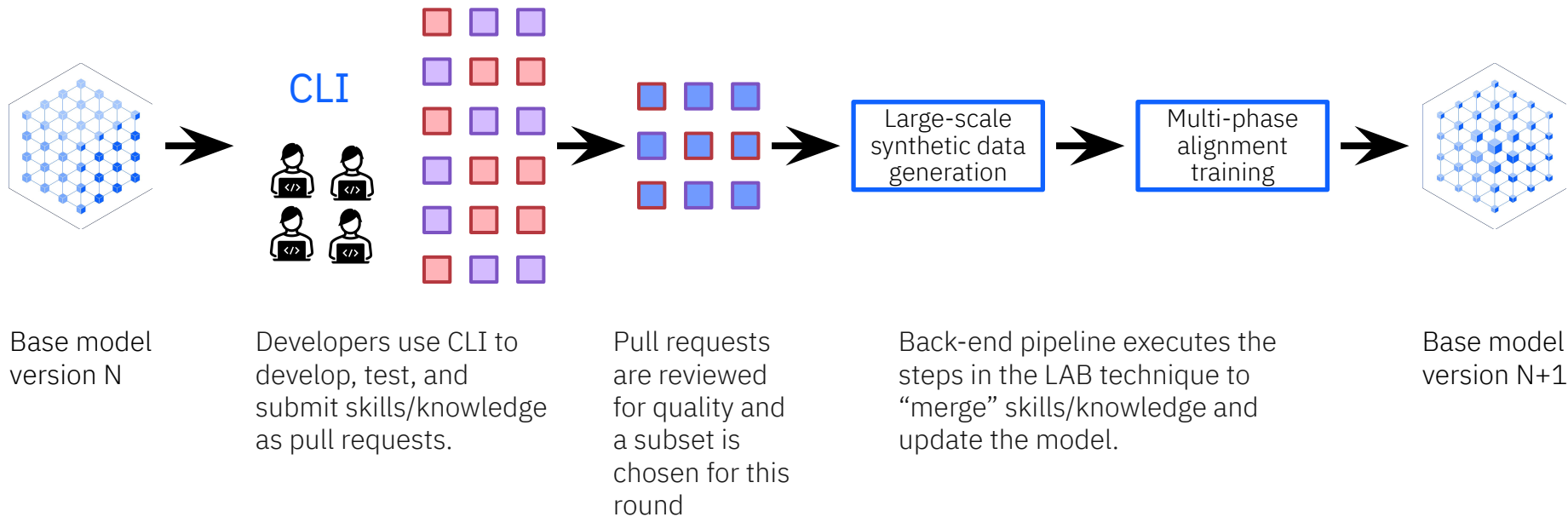
→ Makes LLMs truly open-source with collaborative model development

→ Allows LLMs to learn as humans do, using knowledge and skills

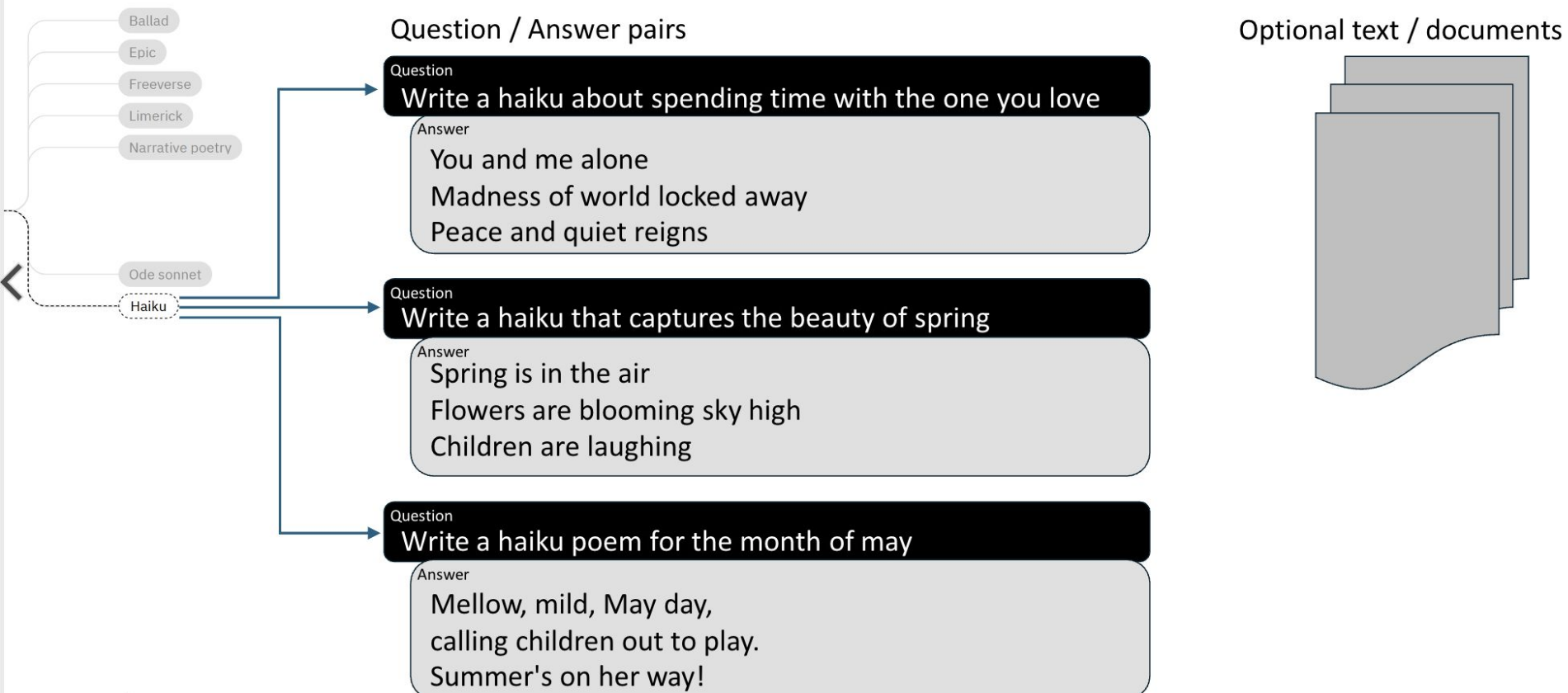
→ Enables incremental skill teaching



InstructLab Workflow

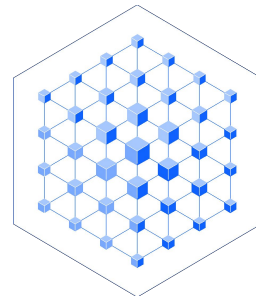
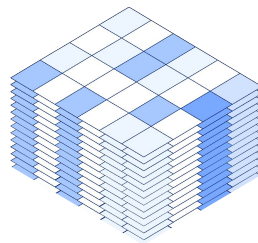
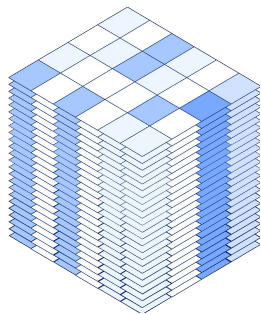
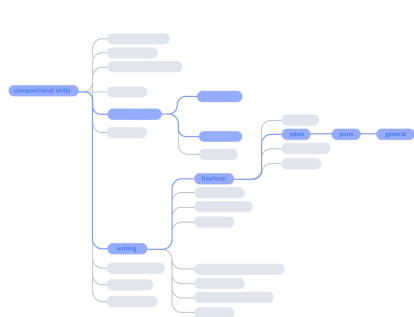


Knowledge example



A taxonomy of skills and knowledge helps to identify gaps in desired capabilities, and then generate enough diversity in synthetic data to tune base models efficiently.

LAB (Large-scale Alignment for ChatBots) Method



Taxonomy-based skill & knowledge representation

<https://arxiv.org/abs/2403.01081>

Represent any missing model knowledge or skills in a hierarchical taxonomy, providing 5+ exemplifying data points of the missing behavior per missing skill.

Synthetic data generation with teacher model

A teacher model generates a “curriculum” of millions of questions and answers across the taxonomy.

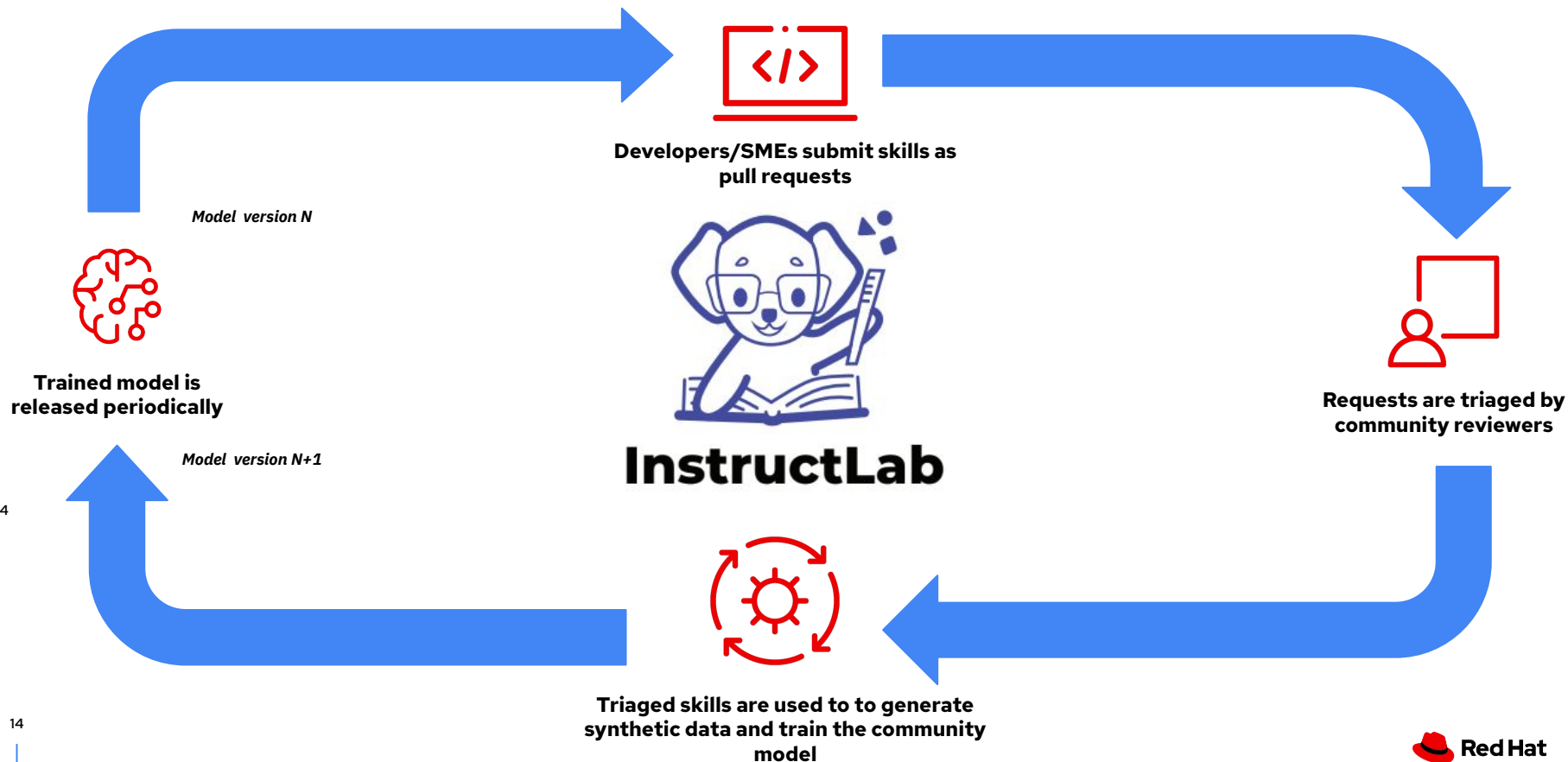
Synthetic data validation with critic model

Critic models filter the questions for correctness and quality. Synthetic data is scanned for prohibited material.

Skill and knowledge training on top of student model(s)

The student model is trained with the curriculum using a novel training approach.

InstructLab



What does InstructLab solve?

Reduces expensive human annotation and proprietary models

Transparency of data

Subject matter knowledge without mathematical skills

Uncertainty of models

Incremental knowledge addition

Community collaboration of one model

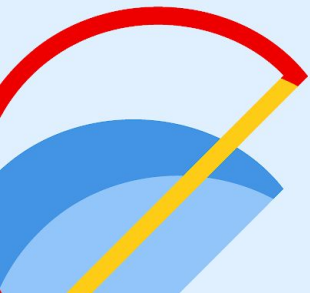
What does InstructLab not solve?

Infrastructure

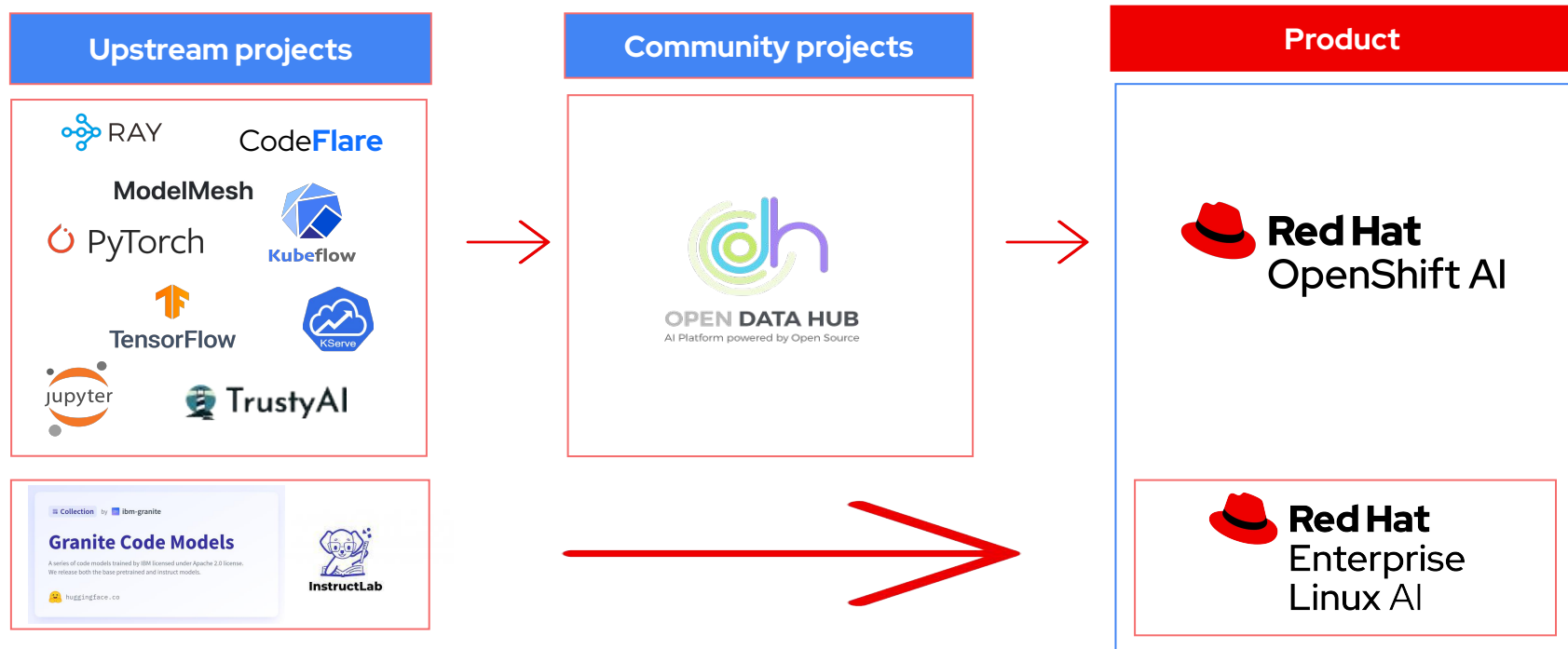
User access and management

RAG-additions and production settings

Monitoring, automatic pipelines



Red Hat's AI/ML engineering is 100% open source



Red Hat AI Portfolio



InstructLab

Open Source

Learn & experiment via limited desktop-scale training method (qlora) on small datasets. *Future potential Podman Desktop integration.*



Laptop / desktop



Red Hat Enterprise Linux AI

Small Scale

Production-grade model training using full synthetic data generation, teacher and critic models. CLI tooling with building blocks.



Server / VM



Red Hat OpenShift AI

Large Scale

Production-grade model training, serving, and monitoring using full power of hybrid cloud app platform for scaling, automation, and MLOps services.



Cluster

Call to action!

<https://github.com/instructlab>



Connect

Thank you



linkedin.com/company/red-hat



facebook.com/redhatinc



youtube.com/user/RedHatVideos



twitter.com/RedHat