KIs, Open-Source, Openwashing

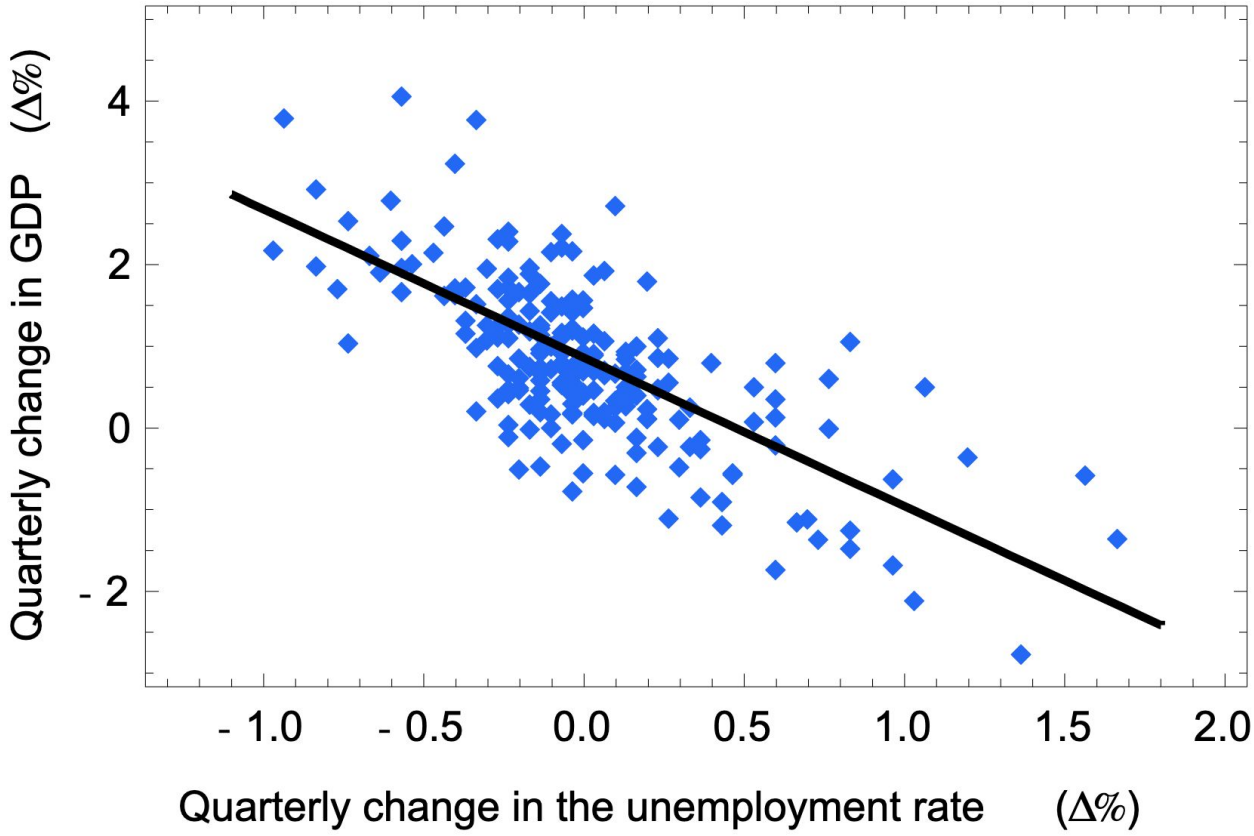
Yannic Kilcher

AG

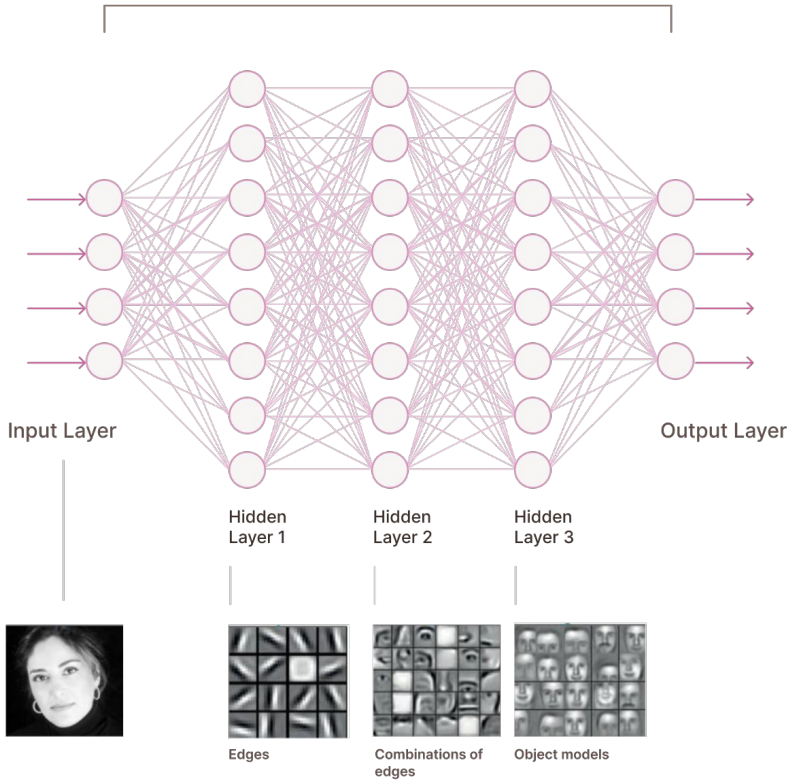
explain software development in the style of justin beiber

Yeah, yeah, yeah, software development, oh yeah
Gotta write the code, gotta make it clean, oh yeah
Gonna build it right, gonna make it sleek, oh yeah
We'll test it out, gonna make it bug-free, oh yeah

Software development, yeah it's like a melody
Got the flow and rhythm, gonna make it harmony
Gonna bring it all together, like a symphony
Make it beautiful, gonna make it just like me, oh yeah


Quarterly change in GDP (Δ%)
4
2
0
- 2
- 1.0
- 0.5
0.0
0.5
1.0
1.5
2.0
Quarterly change in the unemployment rate (Δ%)

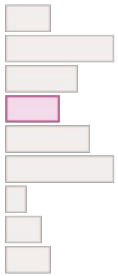
Deep Neural Network



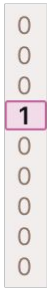
we want the model
to predict this

Training example: **I saw a cat** on a mat <eos>

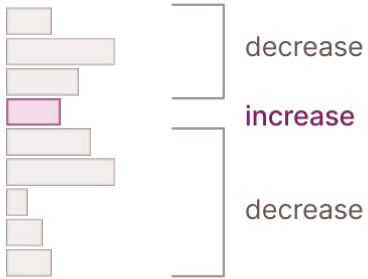
Model prediction: p (* | **I saw a**)

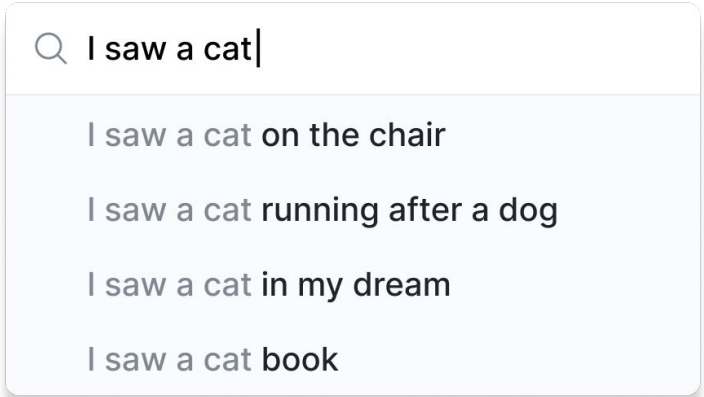


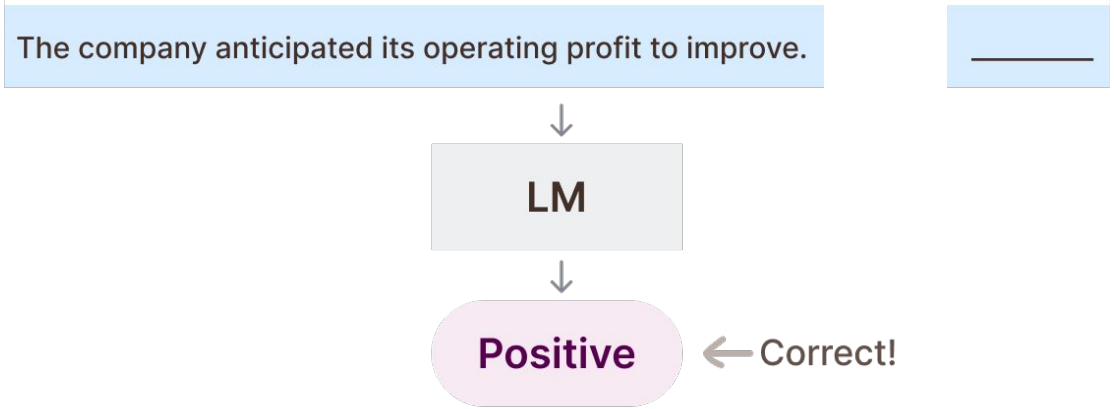
Target



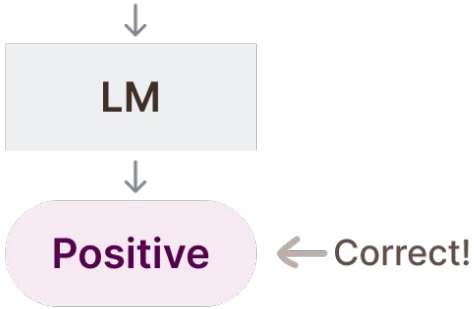
Loss = -log(**cat**)) → min




I saw a cat
I saw a cat on the chair
I saw a cat running after a dog
I saw a cat in my dream
I saw a cat book


The company anticipated its operating profit to improve.
_____
LM
Positive
Correct!

Circulation revenue has increased by 5% in Finland.	Positive
Panostaja did not disclose the purchase price.	Neutral
Paying off the national debt will be extremely painful.	Negative
The company anticipated its operating profit to improve.	_____

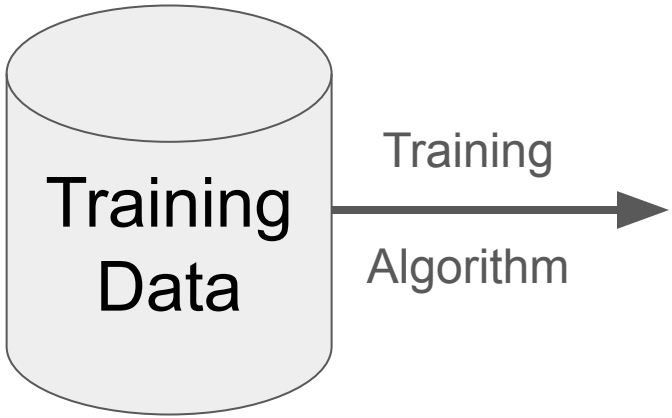


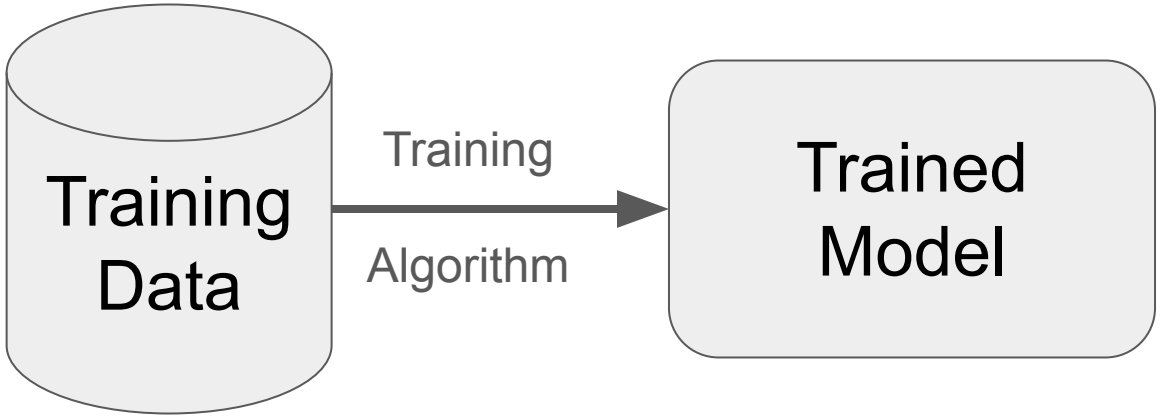
hcebn → bench

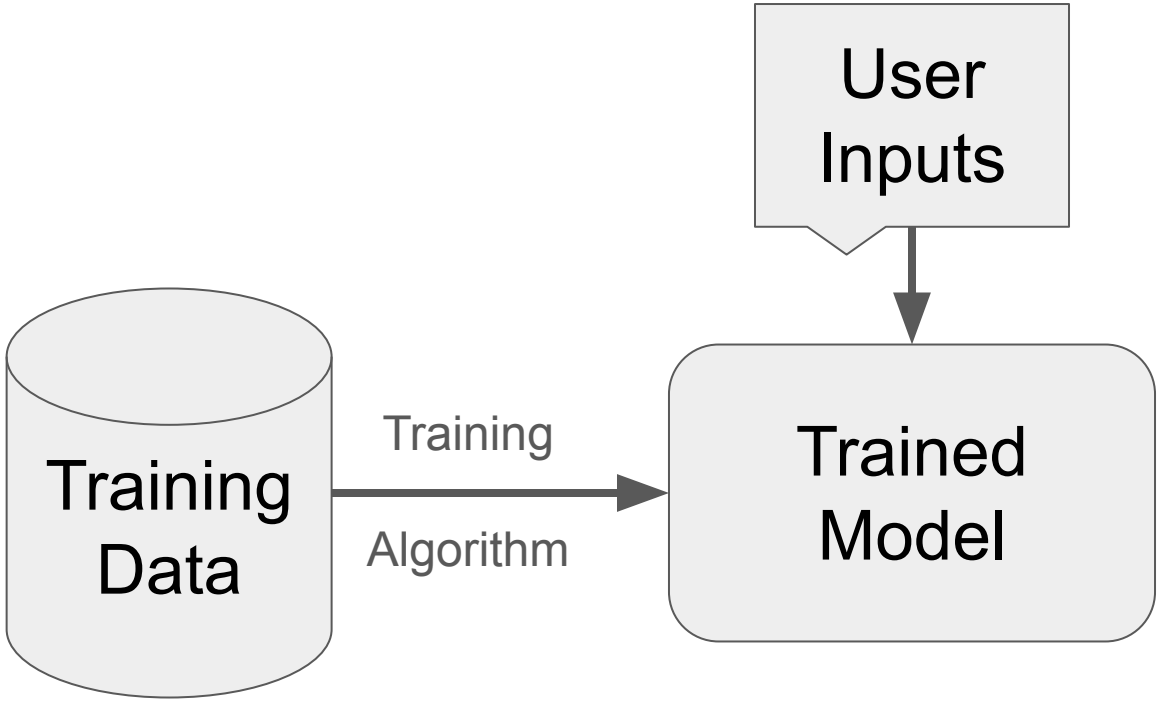
lsain → snail

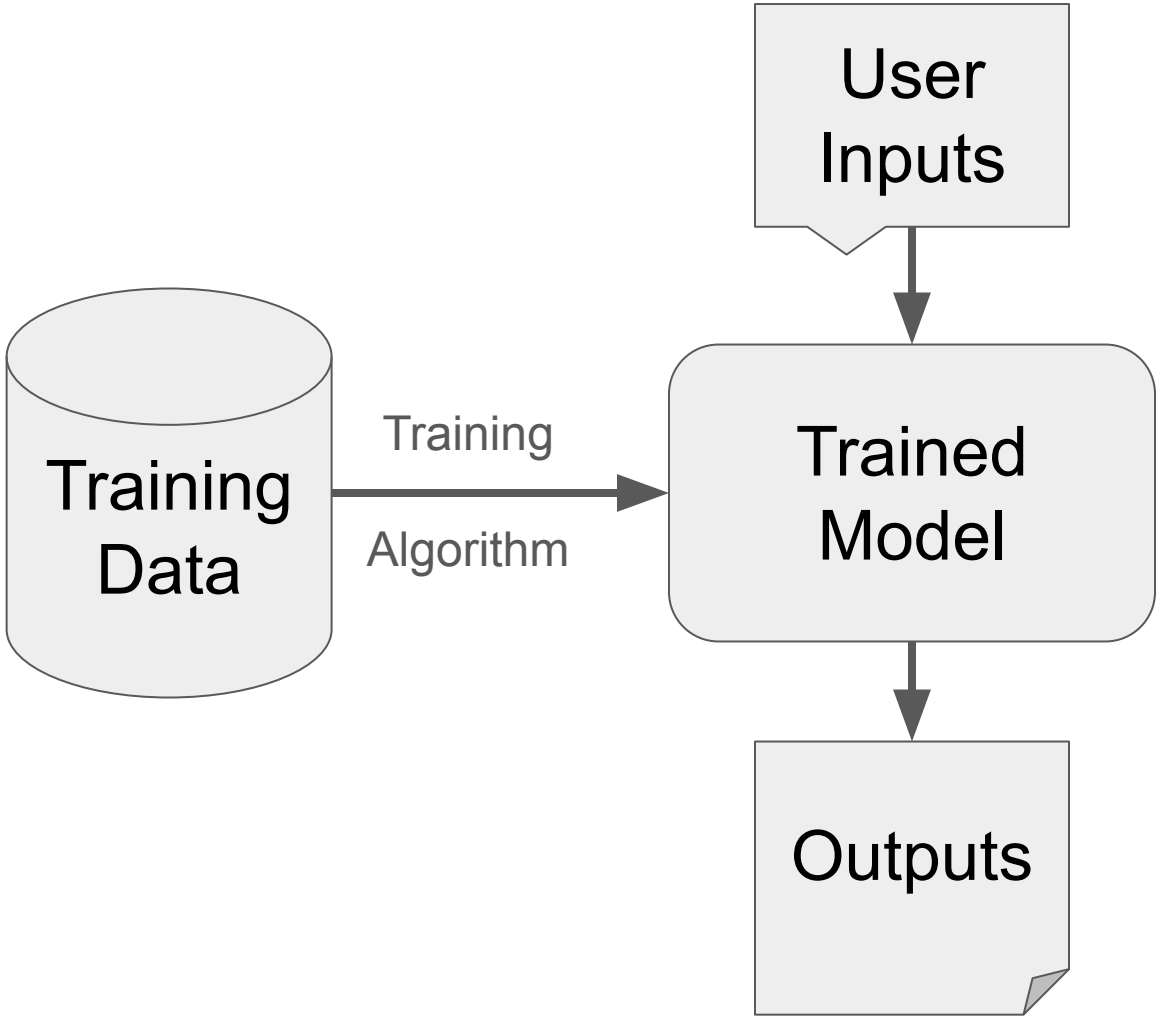
fefoce → ??????

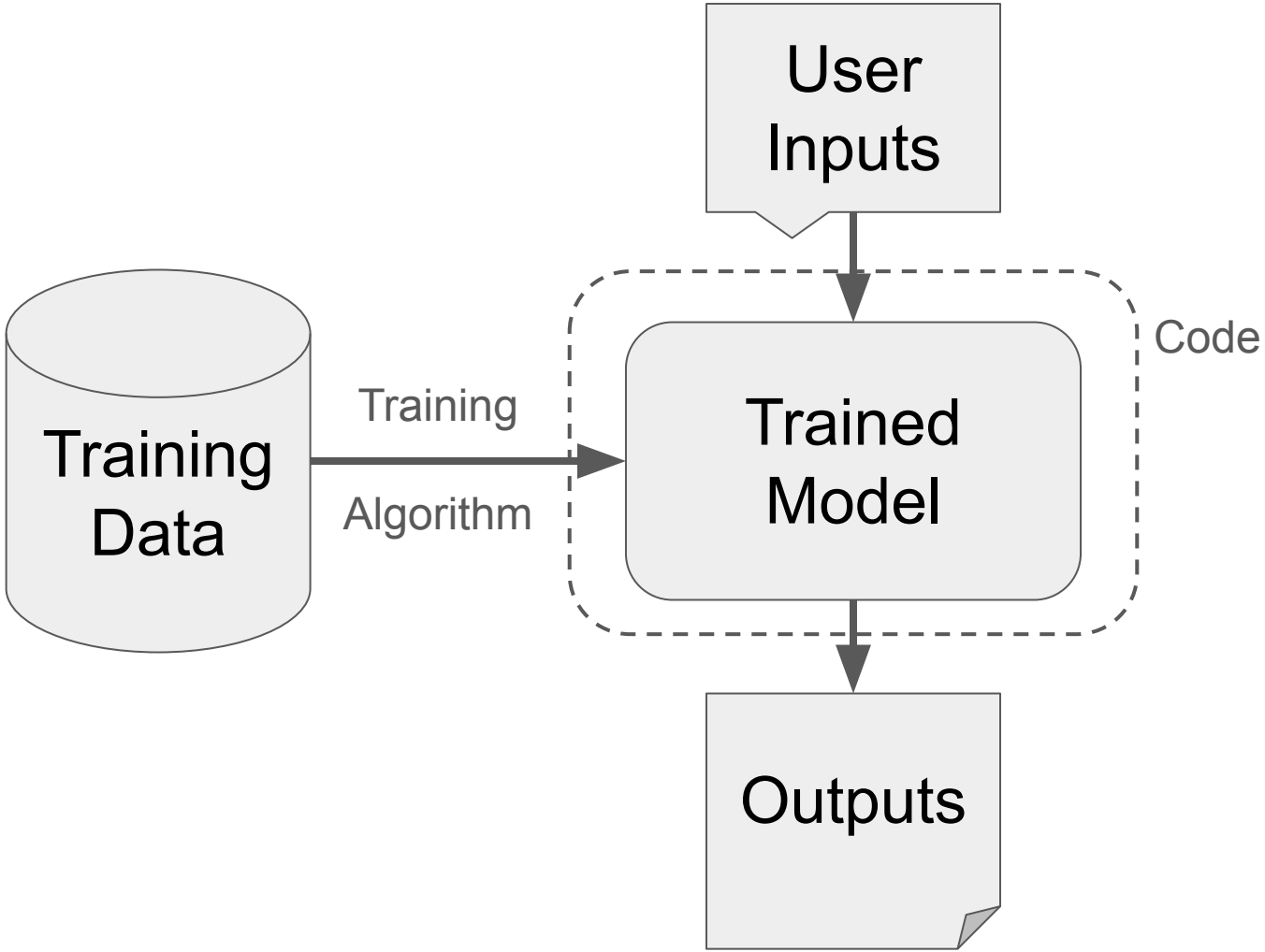

Training
Data

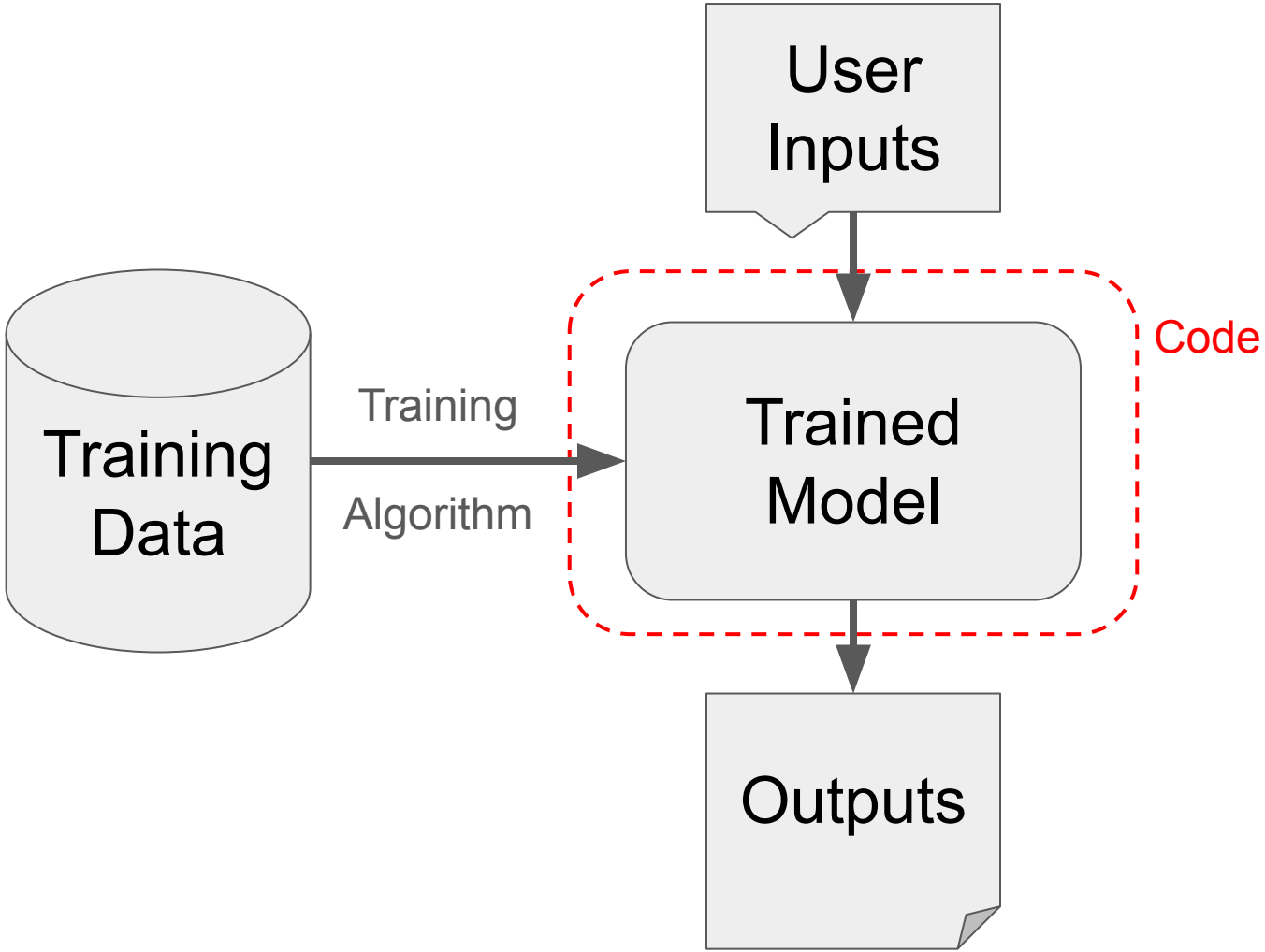

Training Data
Training
Algorithm


Training Data
Training
Algorithm
Trained Model


User Inputs
Training Data
Training
Algorithm
Trained Model

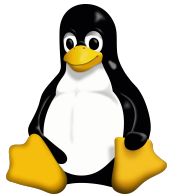

User Inputs
Training Data
Training
Algorithm
Trained Model
Outputs

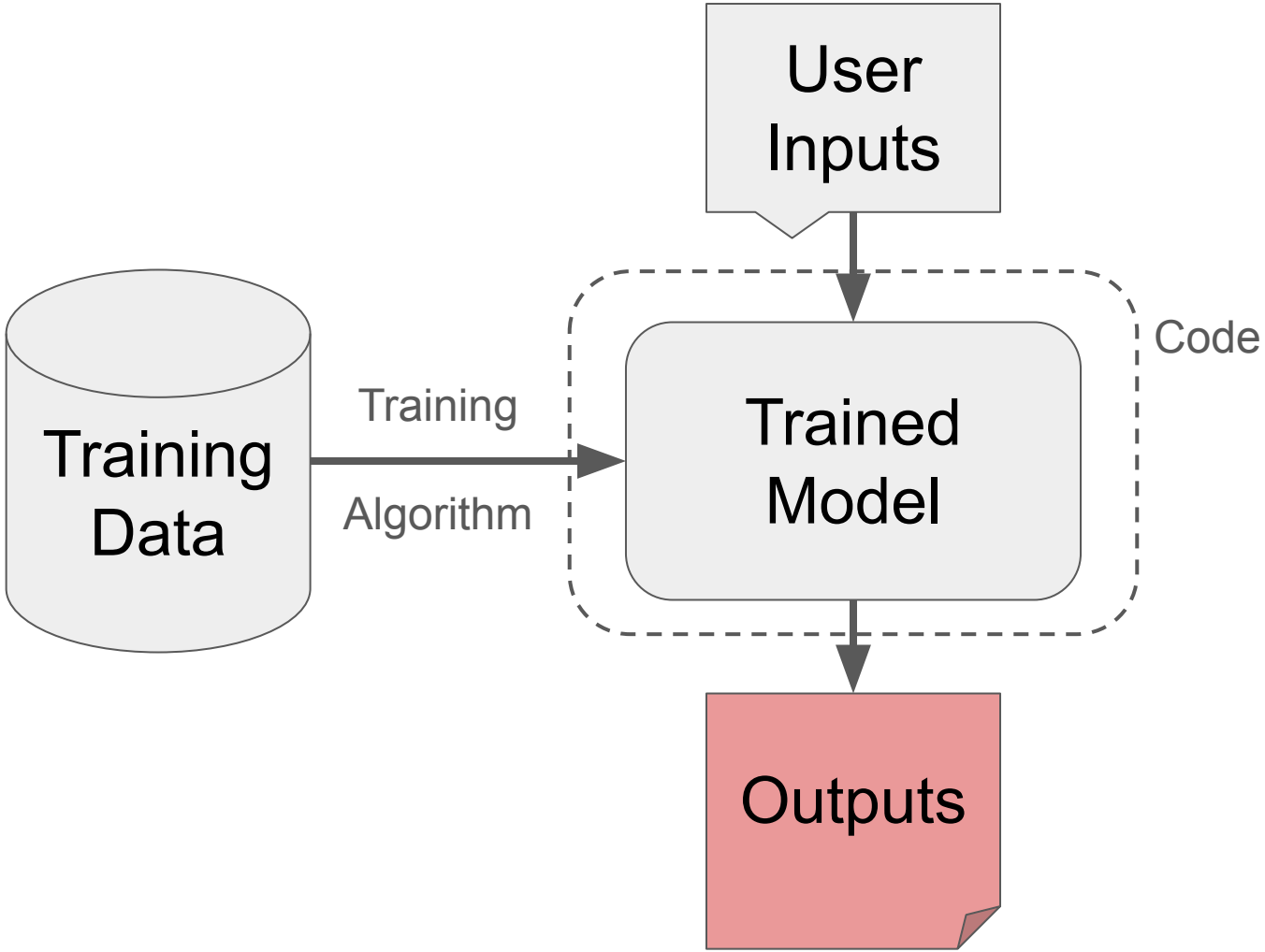

User Inputs
Training Data
Training
Algorithm
Trained Model
Code
Outputs


User Inputs
Training Data
Training
Algorithm
Trained Model
Code
Outputs

Source Code

- Der un-wichtigste Teil der modernen KI-Entwicklung
- Geschützt durch Urheberrecht
  - Automatisch garantiert für den Autor eines kreativen Werkes
  - Regelt u.a. Publikation, Replikation, Verteilung
  - Geschützt ist das Werk, nicht die Idee
- Open-Source Lizenzen
  - Mach-was-du-willst (MIT / Apache / Creative Commons)
  - Copyleft (GPL)



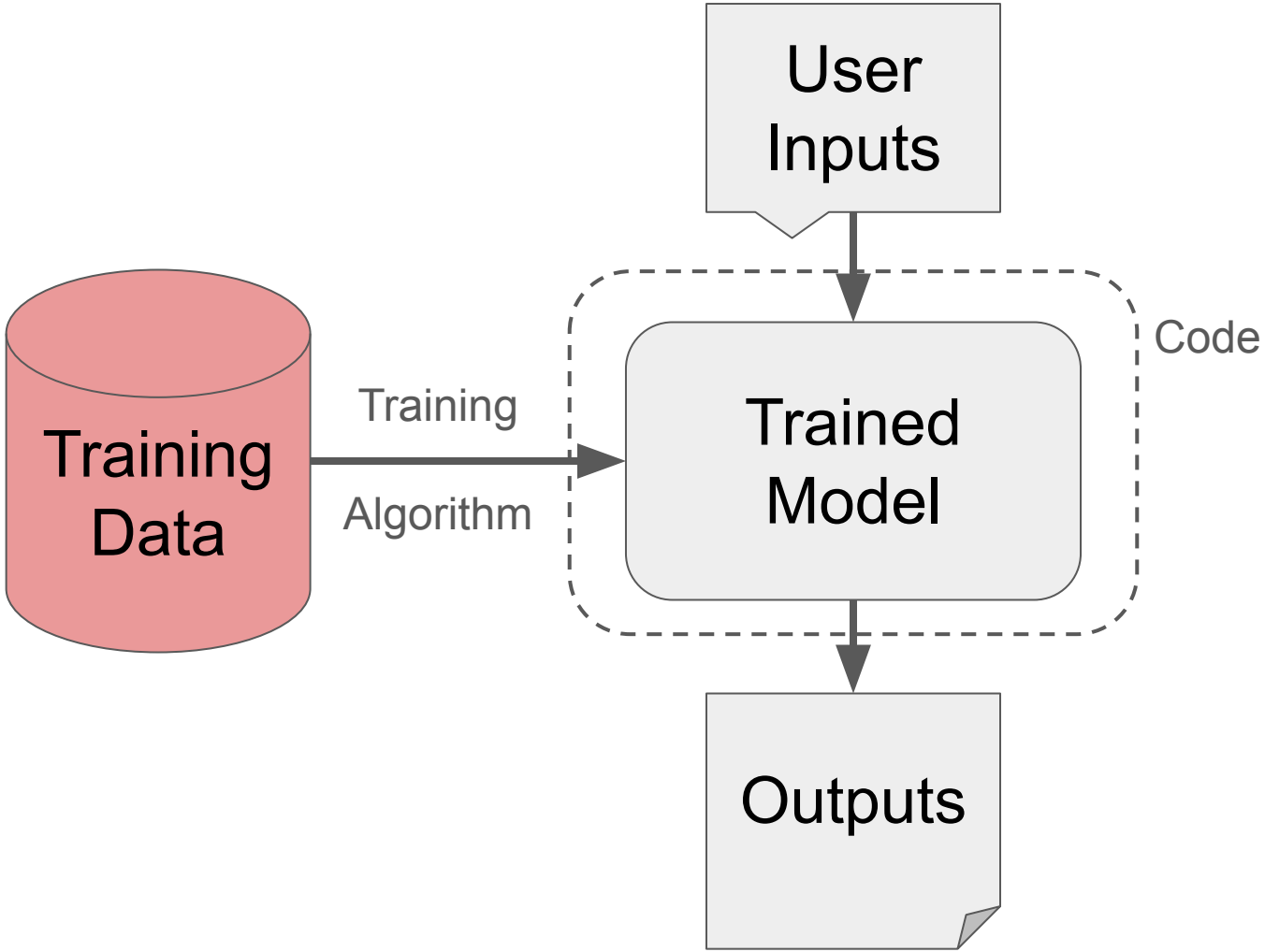

User Inputs
Training Data
Training
Algorithm
Trained Model
Code
Outputs

Model Outputs

- Besteht ein Urheberrecht?
  - Menschlicher Autor
  - Kreativität
- Wer hätte Urheberrecht?
  - Der Benutzer?
  - Der Betreiber des Modells?
  - Der Erschaffer des Modells?
  - Das Modell selbst?
- AGBs
  - Klassische Vertragsbindung

Model Outputs

- Können Outputs das Urheberrecht anderer schädigen?
  - Wie nah ist zu nah?
- Was, wenn die Outputs…
  - schlecht sind?
  - falsch sind?
  - illegal sind?


User Inputs
Training Data
Training
Algorithm
Trained Model
Code
Outputs

Training Data

- Wurden die Daten legal gesammelt?
  - Web scraping
  - Bezahlte Arbeiter
  - Freiwillige
- Haben diese Daten Urheberrecht?
- Was ist mit Datenschutz?
- Menschen lernen auch von geschützten Daten

Open Assistant

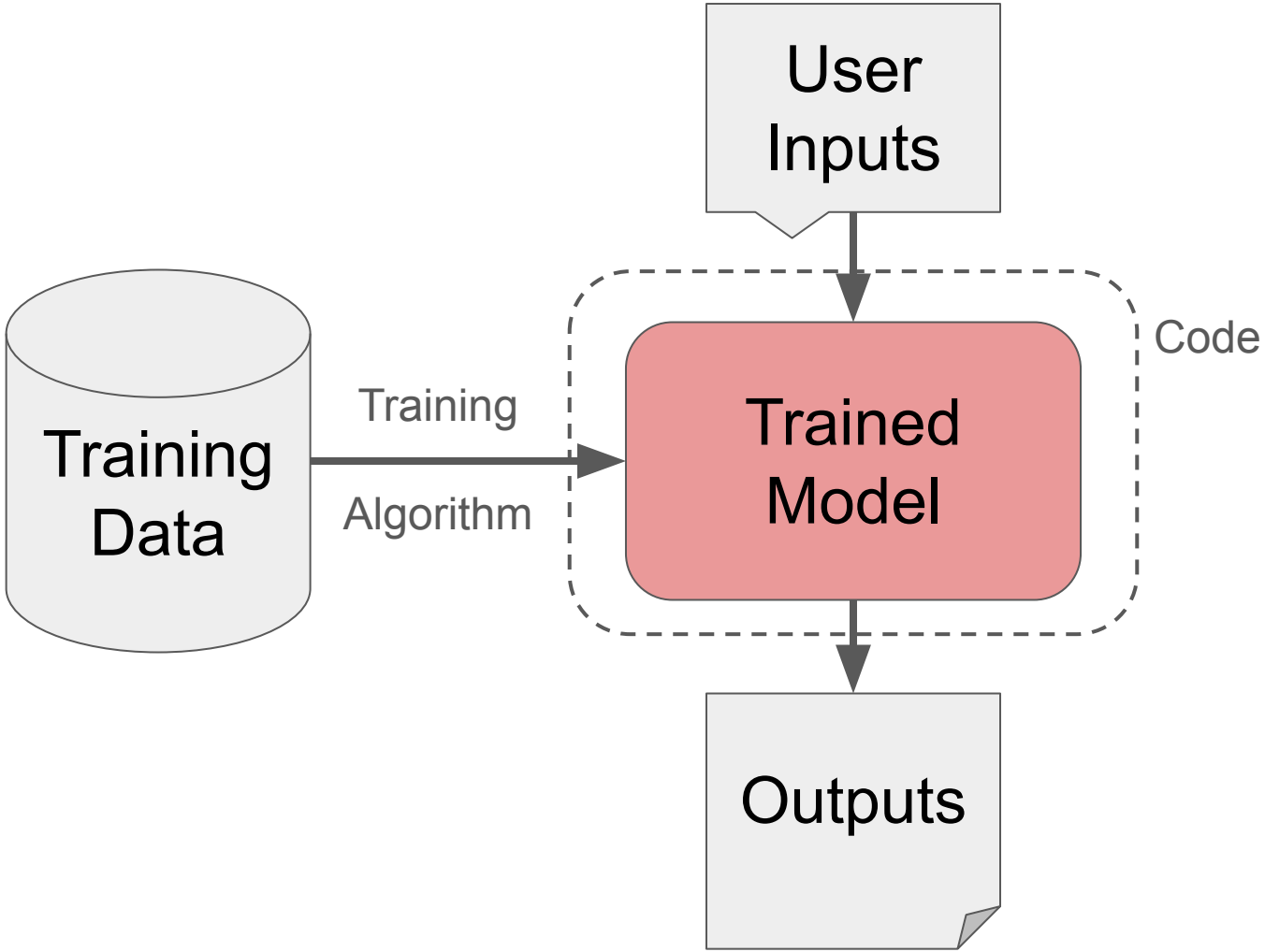
Conversational AI for everyone.

We believe we can create a revolution.

In the same way that Stable Diffusion helped the world make art and images in new ways, we want to improve the world by providing amazing conversational AI.

Try our assistant

Help us improve


User Inputs
Training Data
Training
Algorithm
Trained Model
Code
Outputs

Trained Model

- Besteht Urheberrecht auf das Modell selbst?
- Vermehrt kommen hier AGBs zum Einsatz
- Unknoventionelle "Lizenzen" (Nicht Open-Source)
  - BLOOM → OpenRAIL
  - The dangers of DIY → OpenRAIL++
  - Research-only licenses (Llama v1)
  - Not-my-competitor licenses (Llama v2)

OpenRAIL

- **Responsible:** OpenRAIL licenses embed a specific set of restrictions for the use of the licensed AI artifact in identified critical scenarios. Use-based restrictions are informed by an evidence-based approach to ML development and use limitations which forces to draw a line between promoting wide access and use of ML against potential social costs stemming from harmful uses of the openly licensed AI artifact. Therefore, while benefiting from an open access to the ML model, the user will not be able to use the model for the specified restricted scenarios.

OpenRAIL

7. Updates and Runtime Restrictions. To the maximum extent permitted by law, Licensor reserves the right to restrict (remotely or otherwise) usage of the Model in violation of this License, update the Model through electronic means, or modify the Output of the Model based on updates. You shall undertake reasonable efforts to use the latest version of the Model.

OpenRAIL++-M License Template

Copyright (c) [year] [authors]

[project/org name] Open RAIL++-M
dated August 22, 2022

(This license is based on the CreativeML Open RAIL-M license for stable diffusion.)

Section I: PREAMBLE

Llama 2

2. Additional Commercial Terms. If, on the Llama 2 version release date, the monthly active users of the products or services made available by or for Licensee, or Licensee's affiliates, is greater than 700 million monthly active users in the preceding calendar month, you must request a license from Meta, which Meta may grant to you in its sole discretion, and you are not authorized to exercise any of the rights under this Agreement unless or until Meta otherwise expressly grants you such rights.

Openwashing Example 1: OpenAI

- OpenAI AGB: Outputs dürfen nicht dazu verwendet werden, ein konkurrenzierendes Modell zu trainieren.
- Person 1
  - Akzeptiert OpenAI AGBs
  - Speichert und publiziert die Outputs
  - Trainiert selbst nichts
- Person 2
  - Akzeptiert OpenAI AGBs nicht
  - Nimmt die Daten von Person 1
  - Trainiert ein konkurrenzierendes Modell

Openwashing Example 2: Kunst

- Kunstwerke sind urheberrechtlich geschützt
- Urheberrecht verbietet (bisher) nicht das trainieren auf Daten
- Outputs davon sind keine exakten Kopien, aber nahe
- Outputs haben kein Urheberrecht

Letzte Bemerkungen

- Gesetze aus vergangener Zeit
- Viele offene Fragen
- Viele Grauzonen
- Die Politik wird wahrscheinlich reaktiv handeln
- Was ist Kreativität? Was bedeutet es, zu lernen?