Machine Learning practices and infrastructures

Glen Berman
glen.berman@anu.edu.au

PhD candidate, Australian National University

In proceedings: AIES '23 2023 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society

Outline

Part 1 Introduction to interactive notebooks and clarification of terms: Machine Learning (ML) practitioner, practice, infrastructure.

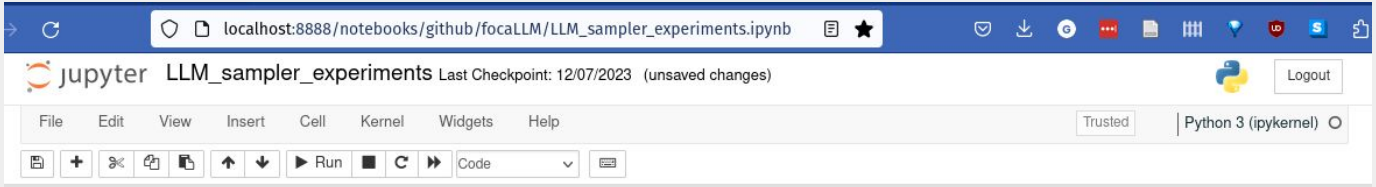
Part 2 Review of empirical findings, showing that notebooks are widely enrolled in ML practices, such that the relationship between practitioners and notebooks is infrastructural.

Part 3 Argument for the importance of considering the role of ML infrastructures in shaping: what it means to be an ML practitioner; the normative commitments manifested in ML practices; and, the space in which AI ethics discourse operates.

Part 1 Introduction to interactive notebooks and clarification of terms: Machine Learning (ML) practitioner, practice, infrastructure.

Part 2 Review of empirical findings, showing that notebooks are widely enrolled in ML practices, such that the relationship between practitioners and notebooks is infrastructural.

Part 3 Argument for the importance of considering the role of ML infrastructures in shaping: what it means to be an ML practitioner; the normative commitments manifested in ML practices; and, the space in which AI ethics discourse operates.



Experiments with using OpenAI's API to recreate classic economics experimnents, inspired by John Horton's working paper: https://john-joseph-horton.com/papers/llm_ask.pdf

In [2]:
```
# assumes
# pip install openai

# requirements
import os
import openai
import sys
import pandas as pd
import random
import collections

openai.api_key = os.environ["OPENAI_API_KEY"]

print("requirements loaded")
```

```
requirements loaded
```

In [3]:
```
# check which models are avaliable
models = openai.Model.list()
models_df = pd.DataFrame(models["data"])
models_df.head(20)
```

Out[3]:

	id	object	created	owned_by	permission	root	parent
0	whisper-1	model	1677532384	openai-internal	[{'id': 'modelperm-KlsZlfft3Gma8pl6A8rTnyjs', ...	whisper-1	None
1	babbage	model	1649358449	openai	[{'id': 'modelperm-49FUp5v084tBB49tC4z8LPH5', ...	babbage	None
2	text-davinci-003	model	1669599635	openai-internal	[{'id': 'modelperm-jepinXYt59ncUQrjQEIUEDyC', ...	text-davinci-003	None
3	davinci	model	1649359874	openai	[{'id': 'modelperm-U6ZwlyAd0LyMk4rcMdz33Yc3', ...	davinci	None
4	text-davinci-edit-001	model	1649809179	openai	[{'id': 'modelperm-otmQSS0hmabtVGHI9QB3bct3', ...	text-davinci-edit-001	None
5	babbage-code-search-code	model	1651172509	openai-dev	[{'id': 'modelperm-4qRnA3Hj8HlJbgo0cGbcmErn', ...	babbage-code-search-code	None
6	text-similarity-babbage-001	model	1651172505	openai-dev	[{'id': 'modelperm-48kcCHhfzvnfY84OtJf5m8Cz', ...	text-similarity-babbage-001	None

Clarification of terms

Machine Learning practitioners	People who are doing Machine Learning, i.e. people at the applied end of the Machine Learning community.
Practices	Conventions or patterns of doing that exist across a social group.
Infrastructures	Sociotechnical systems upon which practices rely; a relational concept.

Watson, Matt, and Elizabeth Shove. 2022. 'How Infrastructures and Practices Shape Each Other: Aggregation, Integration and the Introduction of Gas Central Heating'. Sociological Research Online.
Star, Susan Leigh. 1999. 'The Ethnography of Infrastructure'. *American Behavioral Scientist* 43 (3): 377–91.

Part 1 Introduction to interactive notebooks and clarification of terms: Machine Learning (ML) practitioner, practice, infrastructure.

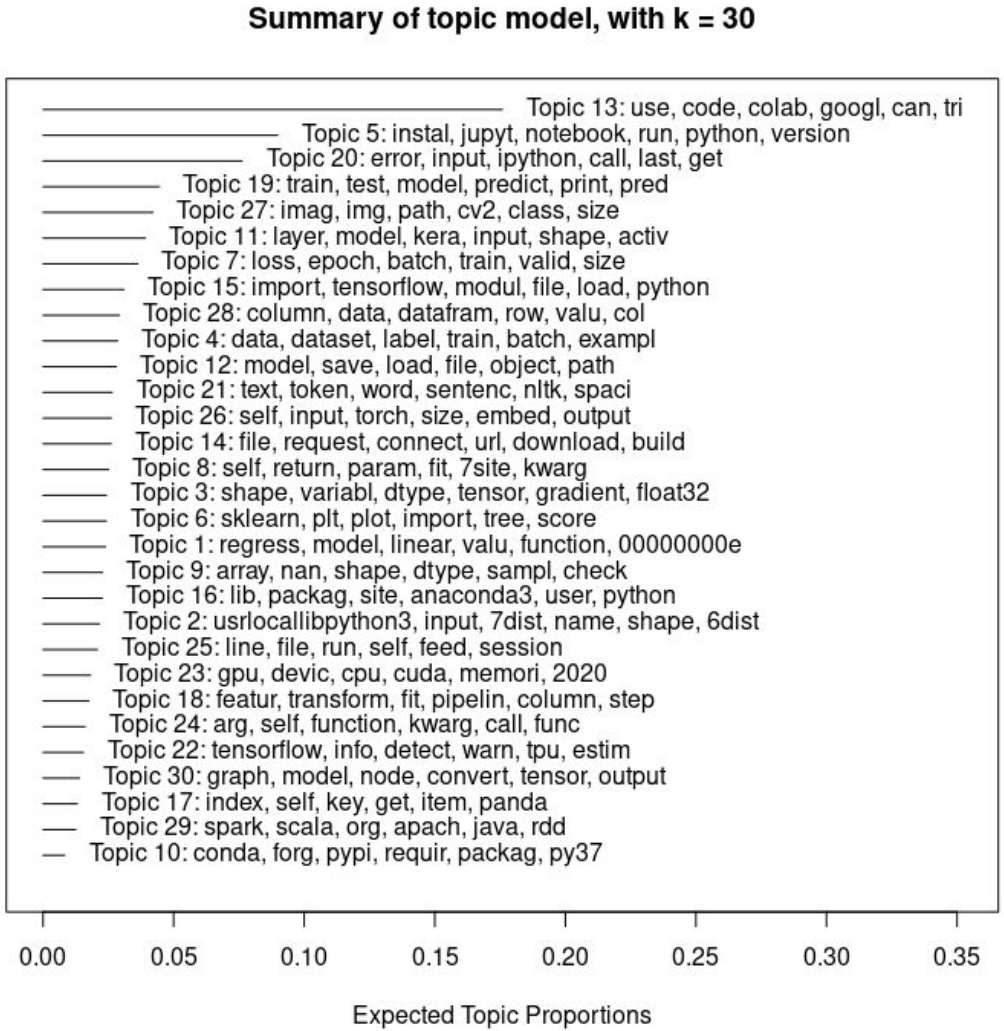
Part 2 Review of empirical findings, showing that notebooks are widely enrolled in ML practices, such that the relationship between practitioners and notebooks is infrastructural.

Part 3 Argument for the importance of considering the role of ML infrastructures in shaping: what it means to be an ML practitioner; the normative commitments manifested in ML practices; and, the space in which AI ethics discourse operates.

How are interactive notebooks used in ML practices?

Data	21,555 Stack Overflow questions on interactive notebooks and Machine Learning.
Analysis	Structural topic model (STM) of questions to surface thematic patterns (topics).
Interpretation	Review most probable words associated with each topic; analyse relationship between topics by calculating their correlation; review highly representative questions.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, and Edoardo M. Airoldi. 2013. 'The Structural Topic Model and Applied Social Science'. In Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation, 4:1–20.


Summary of topic model, with k = 30
Topic 13: use, code, colab, googl, can, tri
Topic 5: instal, jupyt, notebook, run, python, version
Topic 20: error, input, ipython, call, last, get
Topic 19: train, test, model, predict, print, pred
Topic 27: imag, img, path, cv2, class, size
Topic 11: layer, model, kera, input, shape, activ
Topic 7: loss, epoch, batch, train, valid, size
Topic 15: import, tensorflow, modul, file, load, python
Topic 28: column, data, datafram, row, valu, col
Topic 4: data, dataset, label, train, batch, exampl
Topic 12: model, save, load, file, object, path
Topic 21: text, token, word, sentenc, nltk, spaci
Topic 26: self, input, torch, size, embed, output
Topic 14: file, request, connect, url, download, build
Topic 8: self, return, param, fit, 7site, kwarg
Topic 3: shape, variabl, dtype, tensor, gradient, float32
Topic 6: sklearn, plt, plot, import, tree, score
Topic 1: regress, model, linear, valu, function, 00000000e
Topic 9: array, nan, shape, dtype, sampl, check
Topic 16: lib, packag, site, anaconda3, user, python
Topic 2: usrlocallibpython3, input, 7dist, name, shape, 6dist
Topic 25: line, file, run, self, feed, session
Topic 23: gpu, devic, cpu, cuda, memori, 2020
Topic 18: featur, transform, fit, pipelin, column, step
Topic 24: arg, self, function, kwarg, call, func
Topic 22: tensorflow, info, detect, warn, tpu, estim
Topic 30: graph, model, node, convert, tensor, output
Topic 17: index, self, key, get, item, panda
Topic 29: spark, scala, org, apach, java, rdd
Topic 10: conda, forg, pypi, requir, packag, py37
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Expected Topic Proportions

How are interactive notebooks used in ML practices?

Learning laboratory	Interactive notebooks are used as environments in which practitioners experiment with Machine Learning techniques and tools.
Coordination hub	Interactive notebooks are used as spaces for the assembly of many layers of Machine Learning infrastructure into coherent systems.

Part 1 Introduction to interactive notebooks and clarification of terms: Machine Learning (ML) practitioner, practice, infrastructure.

Part 2 Review of empirical findings, showing that notebooks are widely enrolled in ML practices, such that the relationship between practitioners and notebooks is infrastructural.

Part 3 Argument for the importance of considering the role of ML infrastructures in shaping: what it means to be an ML practitioner; the normative commitments manifested in ML practices; and, the space in which AI ethics discourse operates.

The significance of Machine Learning infrastructures for AI ethics

The emergence of infrastructures and formation of social groups are intertwined.

Star, Susan Leigh. 1999. 'The Ethnography of Infrastructure'. *American Behavioral Scientist* 43 (3): 377–91.

The significance of Machine Learning infrastructures for AI ethics

The emergence of infrastructures and formation of social groups are intertwined.

Infrastructures can constrain the practices of the social groups they are related to.

Star, Susan Leigh. 1999. 'The Ethnography of Infrastructure'. *American Behavioral Scientist* 43 (3): 377–91.
Watson, Matt, and Elizabeth Shove. 2022. 'How Infrastructures and Practices Shape Each Other: Aggregation, Integration and the Introduction of Gas Central Heating'. Sociological Research Online.

The significance of Machine Learning infrastructures for AI ethics

The emergence of infrastructures and formation of social groups are intertwined.

Infrastructures can constrain the practices of the social groups they are related to.

Infrastructures enact particular politics, in part by rendering certain labour, relations, and materials invisible to users.

Larkin, Brian. 2013. 'The Politics and Poetics of Infrastructure'. Annual Review of Anthropology 42: 327–43.
Perrigo, Billy. 2023. 'The Workers Behind AI Rarely See Its Rewards.' Time Magazine online.
Star, Susan Leigh. 1999. 'The Ethnography of Infrastructure'. *American Behavioral Scientist* 43 (3): 377–91.

The significance of Machine Learning infrastructures for AI ethics

The emergence of infrastructures and formation of social groups are intertwined.

Infrastructures can constrain the practices of the social groups they are related to.

Infrastructures enact particular politics, in part by rendering certain labour, relations, and materials invisible to users.

What is the responsibility of infrastructure operators?

Gillespie, Tarleton. 2020. Content Moderation, AI, and the Question of Scale. Big Data and Society. Vol. 7. 2.
Gorwa, Robert. 2020. 'Towards Fairness, Accountability, and Transparency in Platform Governance'. AoIR Selected Papers of Internet Research, February.

Machine Learning practices and infrastructures

Highlighted findings

Interactive notebooks are learning laboratories for ML practices, through which practitioners gain an experiential understanding of ML techniques.

Interactive notebooks are coordination hubs, central points through which a web of infrastructures are marshalled during ML practices.

An infrastructural relationship exists between interactive notebooks and ML practitioners (who use them).

Significance

The emergence of infrastructures and formation of social groups are intertwined.

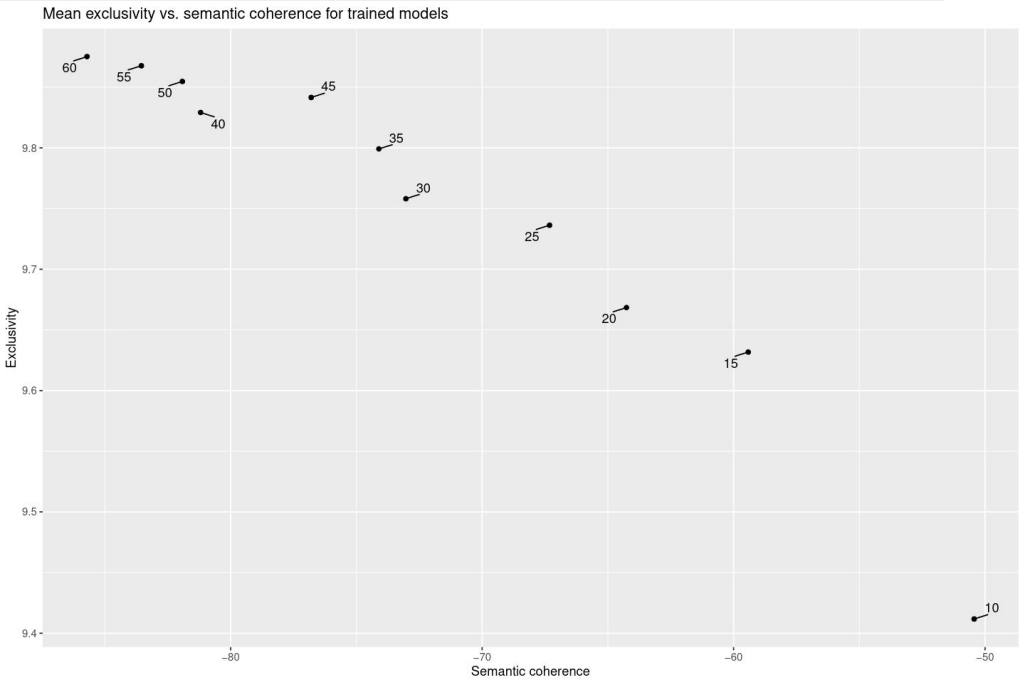
Infrastructures can constrain the practices of the social groups they are related to.

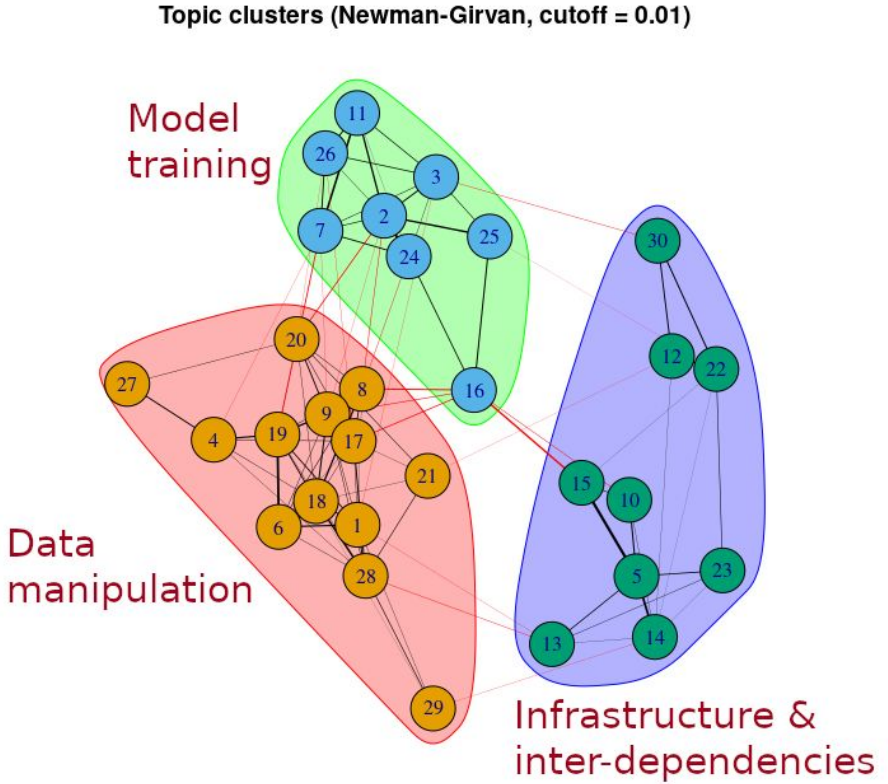
Infrastructures enact particular politics, in part by rendering certain labour, relations, and materials invisible to users.

What is the responsibility of infrastructure operators?

Thank you!

Glen Berman glen.berman@anu.edu.au


Mean exclusivity vs. semantic coherence for trained models
Exclusivity
Semantic coherence
9.8
9.7
9.6
9.5
9.4
−80
−70
−60
−50
60
55
50
45
40
35
30
25
20
15
10


Topic clusters (Newman-Girvan, cutoff = 0.01)
Model training
Data manipulation
Infrastructure & inter-dependencies

How to load CSV file in Jupyter Notebook?

Asked 3 years, 11 months ago Modified 2 days ago Viewed 182k times

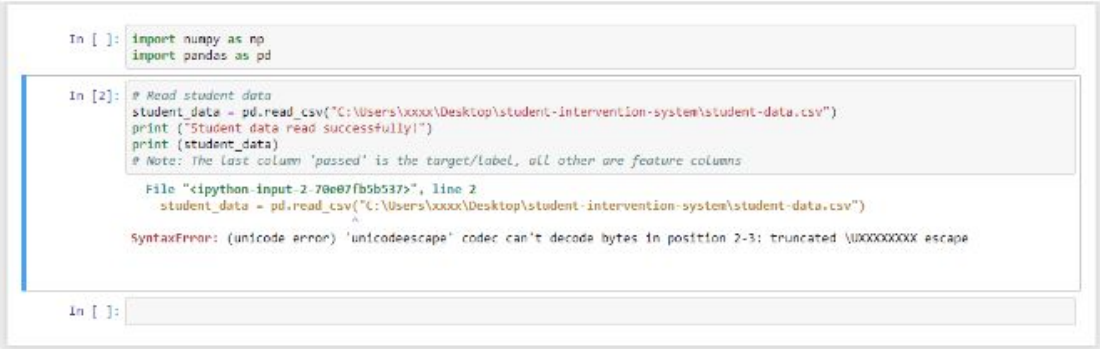
13

I'm new and studying machine learning. I stumble upon a tutorial I found online and I'd like to make the program work so I'll get a better understanding. However, I'm getting problems about loading the CSV File into the Jupyter Notebook.

I get this error:

```
File "<ipython-input-2-70e07fb5b537>", line 2
    student_data = pd.read_csv("C:\Users\xxxx\Desktop\student-intervention-
system\student-data.csv")
                                ^
SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in
position 2-3: truncated \UXXXXXXXX escape
```

and here is the code:



I followed tutorials online regarding this error but none worked. Does anyone know how to fix it?

FailedPreconditionError: Attempting to use uninitialized in Tensorflow

Ask Question

Asked 7 years, 3 months ago Modified 2 years, 9 months ago Viewed 116k times

75

I am working through the TensorFlow tutorial, which uses a "weird" format to upload the data. I would like to use the NumPy or pandas format for the data, so that I can compare it with scikit-learn results.

I get the digit recognition data from Kaggle: https://www.kaggle.com/c/digit-recognizer/data.

Here the code from the TensorFlow tutorial (which works fine):

```
# Stuff from tensorflow tutorial
import tensorflow as tf

sess = tf.InteractiveSession()

x = tf.placeholder("float", shape=[None, 784])
y_ = tf.placeholder("float", shape=[None, 10])

W = tf.Variable(tf.zeros([784, 10]))
b = tf.Variable(tf.zeros([10]))

y = tf.nn.softmax(tf.matmul(x, W) + b)

cross_entropy = -tf.reduce_sum(y_ * tf.log(y))

train_step = tf.train.GradientDescentOptimizer(0.01).minimize(cross_entropy)
```

Here I read the data, strip out the target variables and split the data into testing and training datasets (this all works fine):

Keras, how do I predict after I trained a model?

Asked 6 years, 8 months ago Modified 2 years, 3 months ago Viewed 217k times

108

I'm playing with the reuters-example dataset and it runs fine (my model is trained). I read about how to save a model, so I could load it later to use again. But how do I use this saved model to predict a new text? Do I use `models.predict()`?

Do I have to prepare this text in a special way?

I tried it with

```
import keras.preprocessing.text

text = np.array(['this is just some random, stupid text'])
print(text.shape)

tk = keras.preprocessing.text.Tokenizer(
        nb_words=2000,
        filters=keras.preprocessing.text.base_filter(),
        lower=True,
        split=" ")

tk.fit_on_texts(text)
pred = tk.texts_to_sequences(text)
print(pred)

model.predict(pred)
```

Can I run Keras model on gpu?

Asked 5 years, 7 months ago Modified 1 year, 7 months ago Viewed 326k times

166

I'm running a Keras model, with a submission deadline of 36 hours, if I train my model on the cpu it will take approx 50 hours, is there a way to run Keras on gpu?

I'm using Tensorflow backend and running it on my Jupyter notebook, without anaconda installed.

python tensorflow keras jupyter

Share Improve this question Follow

edited Aug 14, 2017 at 18:48

halfer
19.8k ● 17 ● 97 ● 185

asked Aug 13, 2017 at 15:58

Ryan
7,689 ● 13 ● 36 ● 64

3 I found this: medium.com/@kegui/... It feels like one could peruse highly rated questions in a narrow field here, and then make a full "answer" on Medium, and make actual money from views. – EngrStudent Dec 9, 2019 at 19:34

1 For AMD GPU. See this post. stackoverflow.com/a/60016869/6117565 – bikram Feb 2, 2020 at 7:50

Add a comment

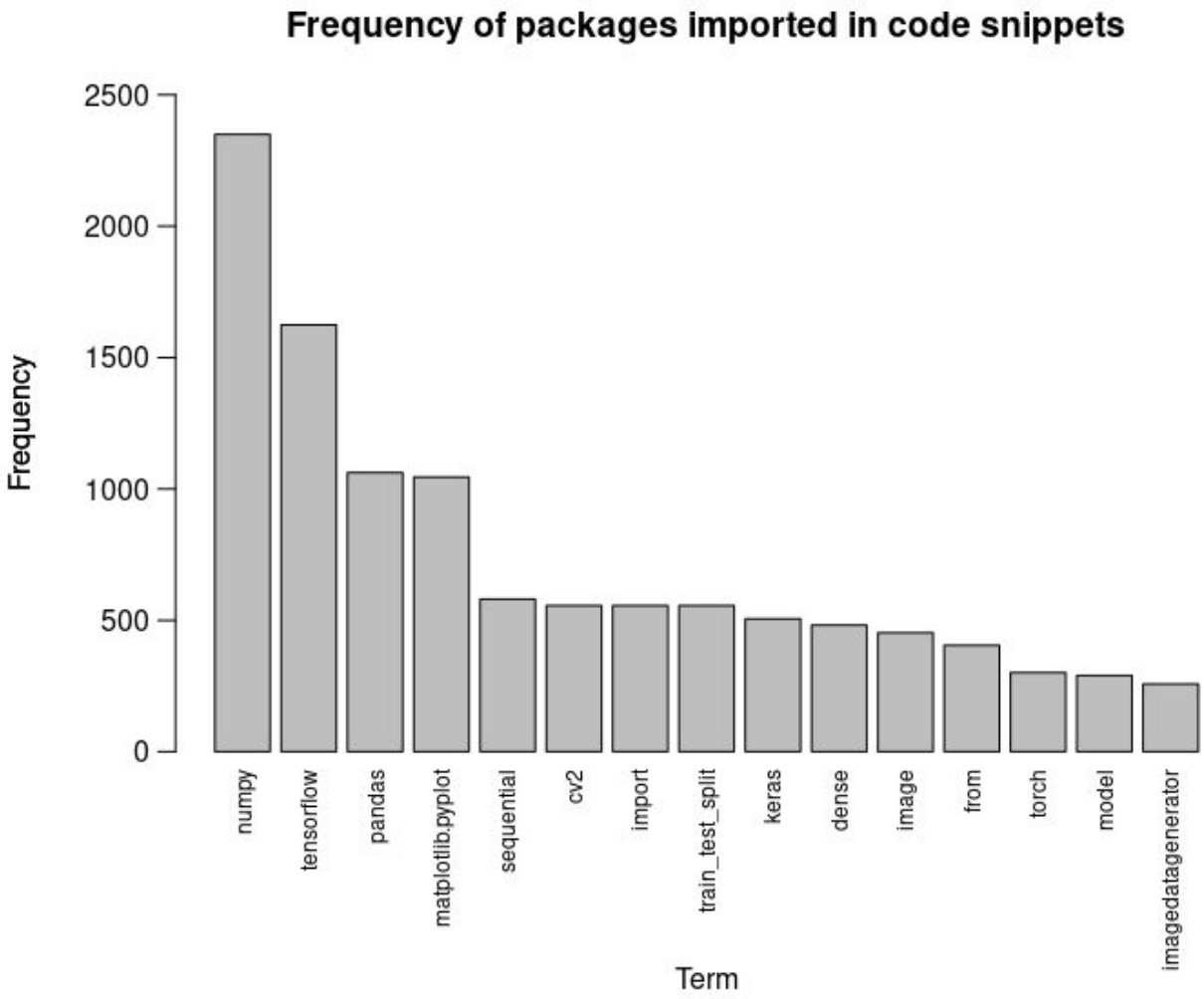
7 Answers

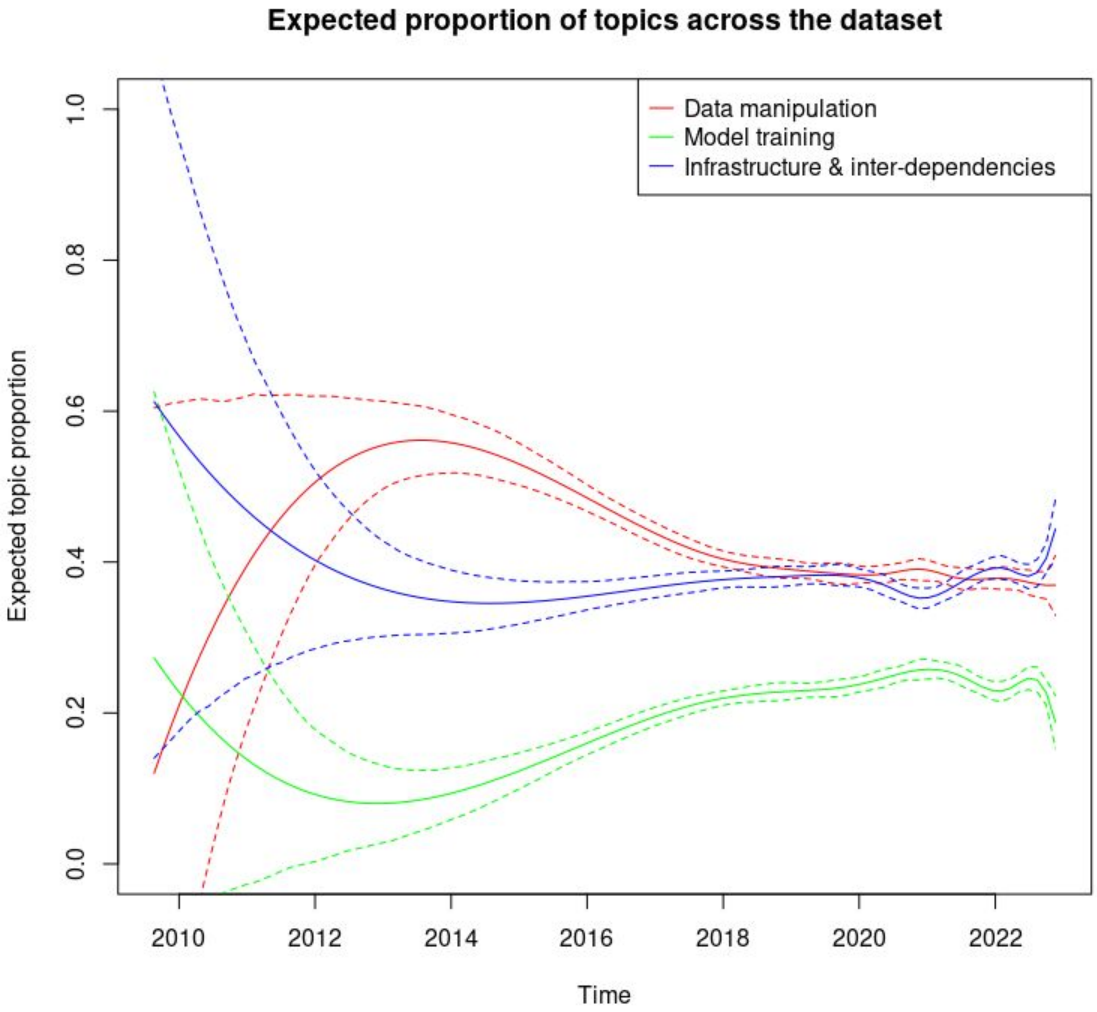
Sorted by: Highest score (default)

220

Yes you can run keras models on GPU. Few things you will have to check first.

1. your system has GPU (Nvidia. As AMD doesn't work yet)
2. You have installed the GPU version of tensorflow


Frequency of packages imported in code snippets
Frequency
2500
2000
1500
1000
500
0
numpy
tensorflow
pandas
matplotlib.pyplot
sequential
cv2
import
train_test_split
keras
dense
image
from
torch
model
imagedatagenerator
Term


Expected proportion of topics across the dataset
Data manipulation
Model training
Infrastructure & inter-dependencies
Expected topic proportion
0.0
0.2
0.4
0.6
0.8
1.0
2010
2012
2014
2016
2018
2020
2022
Time