Domain-Specific Pre-Training for Improved Abstract Summarization in Neuroscience

Dylan Daniels

Problem Overview

- Too many publications, not enough time
- Information overload
- Relying on shortcuts, heuristics

Image:https://img.freepik.com/premium-photo/stress-woman-scientist-with-headache-laboratory-suffering-from-burnout-migraine-pain-overworked-exhausted-frustrated-tired-worker-working-science-research-with-fatigue-tension_590464-150113.jpg

What would a scientist want from an ‘LLM assistant’ ?

- Given a paper...
  - Summarize key points of an article
  - Find all the bits related to their own research

What does our model need to have?

- High level of domain specificity
- Recognize and extract pertinent information
- Return a fluent, reduced representation

Why can’t we do that with Chat GPT, Gemini, etc.?

- We can, but ... there are issues with doing so
- These models tend to ‘fill in the gaps’
- We want a model that is a domain expert

Could we build a model from scratch?

- We could, but ...

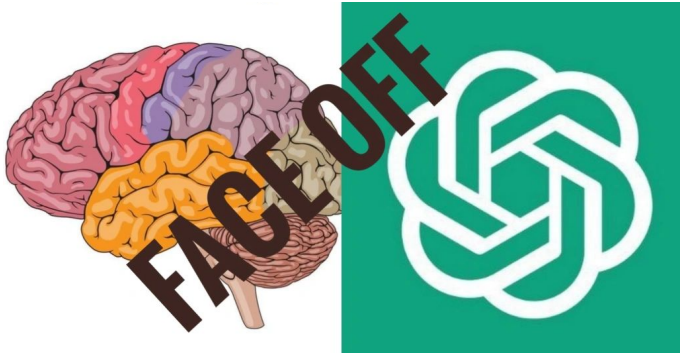

Image:https://hikmaahblog.wordpress.com/wp-content/uploads/2023/04/img_20230404_092315-01.jpeg?w=1024

BrainGPT

This is the homepage for BrainGPT

A Large Language Model tool to assist **neuroscientific research**.

BrainGPT functions as a generative model of the scientific literature, allowing researchers to propose study designs as prompts for which BrainGPT would generate likely data patterns reflecting its current synthesis of the scientific literature. Modellers can use BrainGPT to assess their models against the field's general understanding of a domain (e.g., instant meta-analysis). BrainGPT could help identify anomalous findings, whether because they point to a breakthrough or contain an error.

Importantly, BrainGPT does not summarize papers nor retrieve articles. In such cases, large-language models often confabulate, which is potentially harmful. Instead, BrainGPT stitches together existing knowledge too vast for human comprehension to assist humans in expanding scientific frontiers.

https://braingpt.org/

If BrainGPT isn’t willing to try, do we give up on this highly-detailed, in-domain summarization?

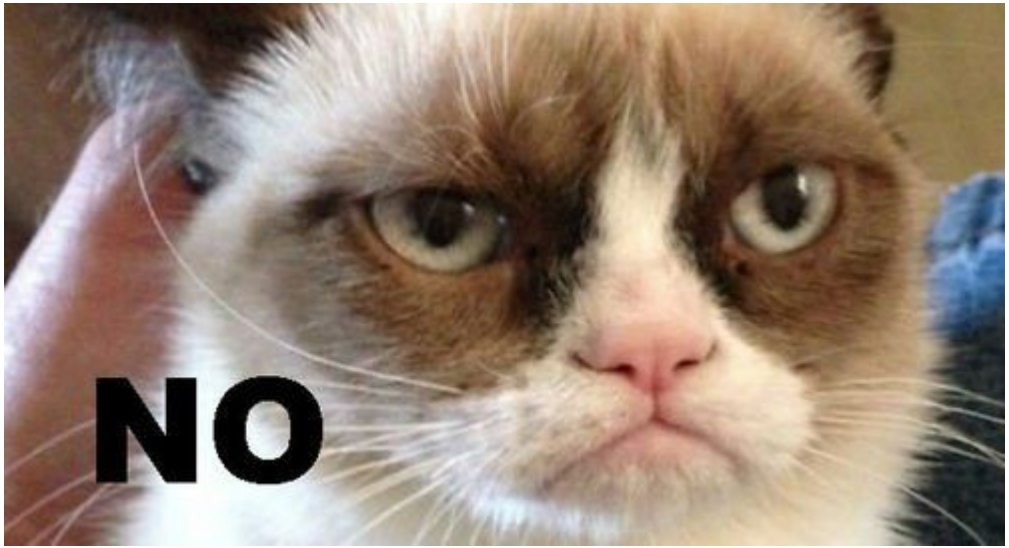

Image:https://nocamels.com/wp-content/uploads/2015/01/grumpycatmeme.jpg

The question: can we adapt a publicly-available model to do in-domain summarization ‘cheaply’

The data

Cold Spring Harbor Laboratory

bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

HOM

18 Results for author "Stephanie R Jones"

Items/Page 10 **Order by** Best Match

Biophysical modeling of frontocentral ERP generation links circuit-level mechanisms of action-stopping to a behavioral race model

Dataset card | Viewer | Files | Community 9

Dataset Viewer (First 5GB) Auto-converted to Parquet API View in Dataset Viewer

Subset (2): arxiv · 156k rows

Split (3): train · 143k rows

article string · *lengths* 0 – 699k	abstract string · *lengths* 0 – 93.1k	section_names string · *lengths* 0 – 25.2k
additive models @xcite provide an important family of models for…	additive models play an important role in…	introduction main resu learning rates compari
the leptonic decays of a charged pseudoscalar meson @xmath7 are…	we have studied the leptonic decay @xmath0 …	[sec:introduction]intr [sec:detector]data and
the transport properties of nonlinear non - equilibrium dynamical systems…	in 84 , 258 (2000) , mateos conjectured that…	introduction regularit chaos in single-partic
studies of laser beams propagating through turbulent atmospheres are…	the effect of a random phase diffuser on…	introduction the metho photon distribution fu
the so - called `` nucleon spin crisis '' raised by the european muo…	with a special intention of clarifying the…	introduction model lag with pion mass term…
let @xmath1 . let	we improve the currently	introduction linear

‹ Previous 1 2 3 ... 1,430 Next ›

The models

1

BART Base

2

BART Fine-Tuned on In-Domain Summarization

3

BART With MLM Training and Fine Tuning In Domain

ROUGE / BERT Score Results

A. Best Run Averaged Across Abstracts

Model	rouge1	rouge2	rougeL	rougeLsum
base BART	0.2400	0.0641	0.1647	0.1647
FTO	0.4192	0.1778	0.3579	0.3579
MLM-FT (3)	0.4140	0.1755	0.3337	0.3337
MLM-FT (6)	0.4212	0.1635	0.3396	0.3396

A. Best Run Averaged Across Abstracts

Model	Precision	Recall	F1
base_bart_model	0.847459	0.876872	0.859904
finetuned_model	0.884751	0.898202	0.890938
mlm_ft_3e_model	0.889835	0.897664	0.892372
mlm_ft_6e_model	0.891861	0.897321	0.892154

B. Average Across All Runs and Abstracts

Model	rouge1	rouge2	rougeL	rougeLsum
base BART	0.2024	0.0438	0.1441	0.1441
FTO	0.3533	0.1346	0.2939	0.2939
MLM-FT (3)	0.3157	0.1099	0.2567	0.2567
MLM-FT (6)	0.3195	0.1103	0.2539	0.2539

B. Average Across All Runs and Abstracts

Model	Precision	Recall	F1
base_bart_model	0.839080	0.863863	0.851181
finetuned_model	0.872938	0.886233	0.879343
mlm_ft_3e_model	0.873287	0.880210	0.876595
mlm_ft_6e_model	0.874708	0.883119	0.878798

ROUGE metrics for this task

Image:https://lh5.googleusercontent.com/proxy/_6_O3nSUGrEsA_giqDBs-qGFKhtHNTvdiFZLa2IOt6SdVm9IM8_sQiYyY34x7gicPXIUTV5LjgxGQf_MzQFxo16q8Id4vP96fSIq6fU27Rh64_0v5hEpaP3toKHbBLEmDx55JkPZarpCzz-1K8cGOG8Sd6t3

A better approach? (maybe...maybe not)

MISTRAL AI_

Image:https://docs.mistral.ai/img/logo.svg

"""**You are a scientist.** You will be given an abstract and three summaries of the abstract, and your job is to determine which summary does the best job of describing the abstract. **Keep in mind, you should select the summary that will be most useful to you as a scientist** in understanding what the paper will be about based on the abstract. **More specific summaries are preferred, as you want to understand the fine details of the science.**

Here is the abstract: {abs}

Here are the summaries you can choose from:

Summary 1: {sum_finetuned}

Summary 2: {sum_mlm_ft_3e}

Summary 3: {sum_mlm_ft_6e}

Which summary best describes the abstract? You can only choose one answer.

Start your response with 'Summary', followed by the summary number and then a colon.

For example, if you choose summary 3, start with 'Summary 3:'

"""

So what did Mistral think?

Metadata Set	Runs (n)	Proportion MLM-FT
A	6	0.5417
B	3	0.5000
C	3	0.4583
D	3	0.4583

Proportion of the time Mistral chose MLM-FT for 'Best' Runs

Metadata Set	Runs (n)	Proportion MLM-FT
A	10	0.5375
B	5	0.5250
C	5	0.4875
D	5	0.5000

Proportion of the time Mistral chose MLM-FT for All Runs

Image:https://i.kinja-img.com/image/upload/c_fill,h_675,pg_1,q_80,w_1200/f13b27b54cb5e11c7505487ddda2a1b9.png

Limitations, Ideas for Improvement

Image:https://cloudarchitectmusings.com/wp-content/uploads/2016/01/625955-whats-next-meme.jpg