AI Advancements

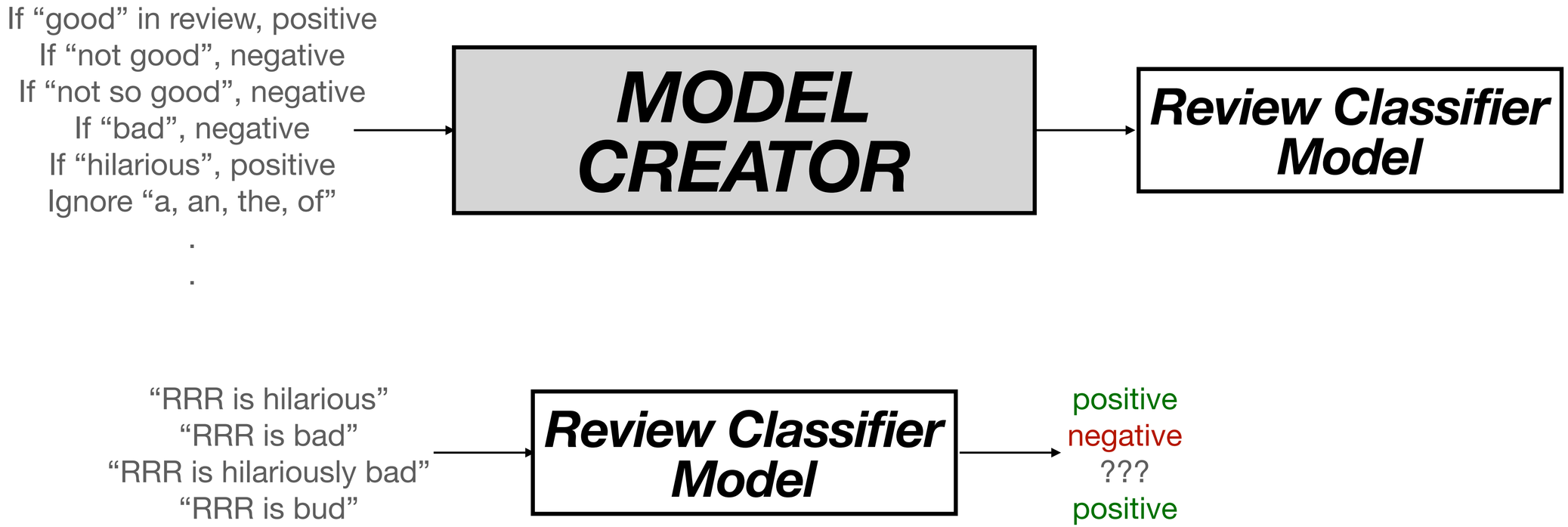
1) Rule-based Models (1990 - 2012)
2) Deep Learning Models (2012 - 2021)
3) Prompt-based Models (2021 - present) [Most relevant to us]

7th July 2022 @ Tech4Dev Sprint, Garudmachi
suryateja@avantifellows.org

Two decades ago…

Rule-based models

- Example Goal: Read a movie review and tell if it is positive or negative



Two decades ago…

Rule-based models

- Example Goal: tell if the image input is plant or animal

If “lots of green”, plant
If “black white stripes alternating”, animal
…

→ ***MODEL CREATOR*** → ***Plant/Animal Classifier***

→ ***Plant/Animal Classifier*** → Plant

→ ***Plant/Animal Classifier*** → Plant

Two decades ago…

Rule-based models —> *no meaning-making skills*

"RRR is hilarious"
"RRR is bad"
"RRR is hilariously bad"
"RRR is bud"

→ ***Review Classifier Model*** →

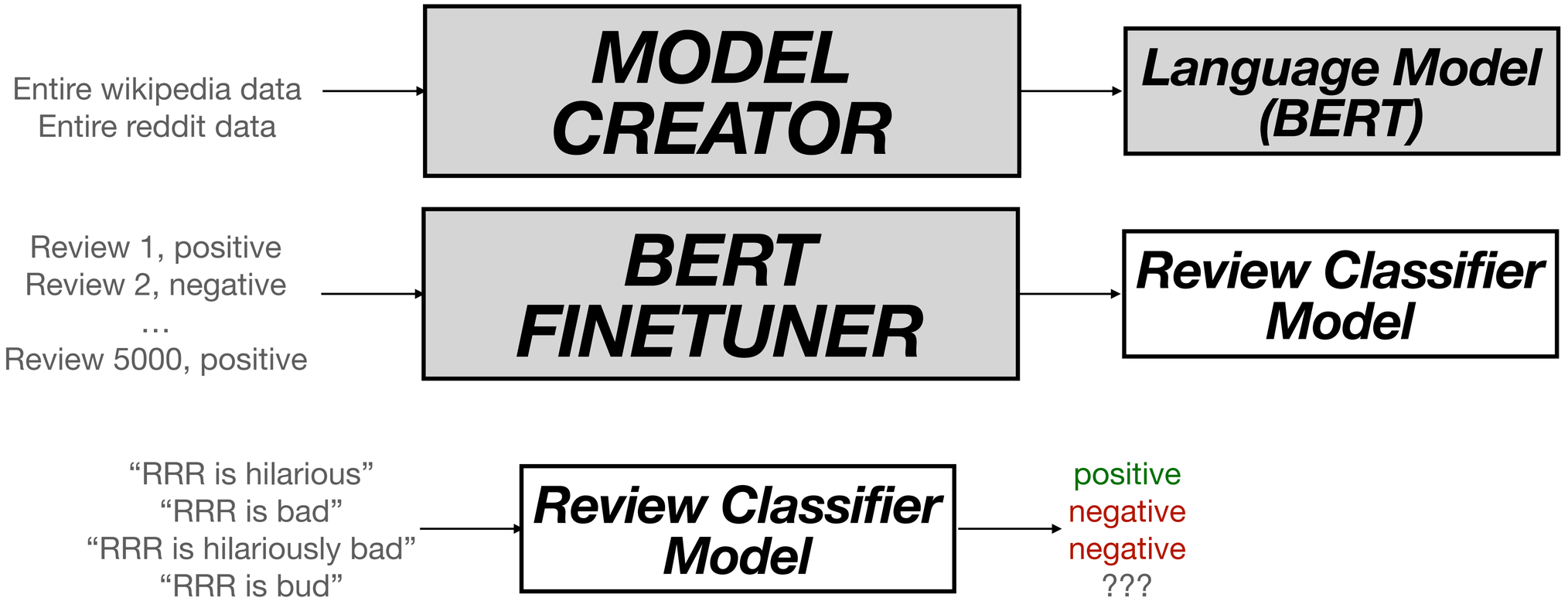
positive
negative
???
positive

→ ***Plant/Animal Classifier*** → Plant


BERTji
from transformers import BertModel

Five years ago…
Deep Learning models

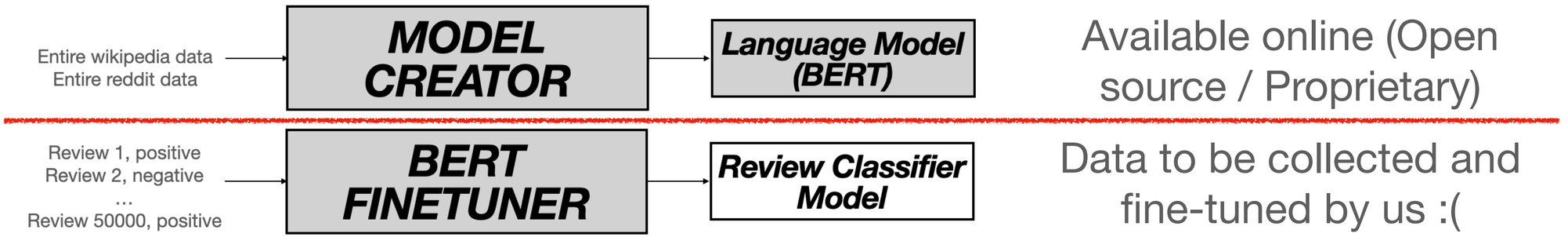
- Example Goal: Read a movie review and tell if it is positive or negative



Five years ago…

Deep Learning models —> *acquire meaning-making skills!*

- BERT knows that “good, great, nice, very good” have similar meaning!
- Fine-tuning helps to match our requirement
  - Say you want to translate, collect 10k+ samples! “Aap kaise hai -> how are you”, “Bahar kaise jana hai -> how to go outside”



Five years ago…

Deep Learning models -> *acquire meaning-making skills!*

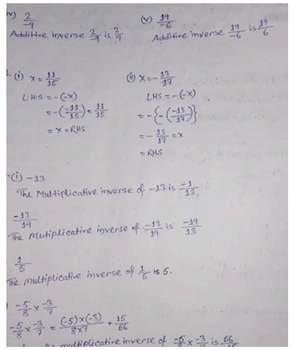
- Example Goal: tell if the image input is handwritten or not-handwritten

Entire wiki commons
image data
Images collected from
Flickr / Insta

→ **MODEL CREATOR** → **Image Model**

Image 1, not-handwritten
Image 2, handwritten
…
Image 10000, handwritten

→ **Image model FINETUNER** → **Handwritten or not?**

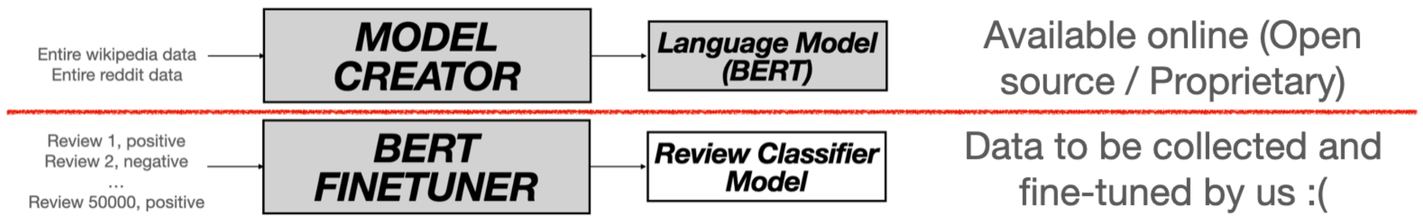


→ **Handwritten or not?** → Handwritten

Five years ago…

Deep Learning models —> *acquire meaning-making skills, but*

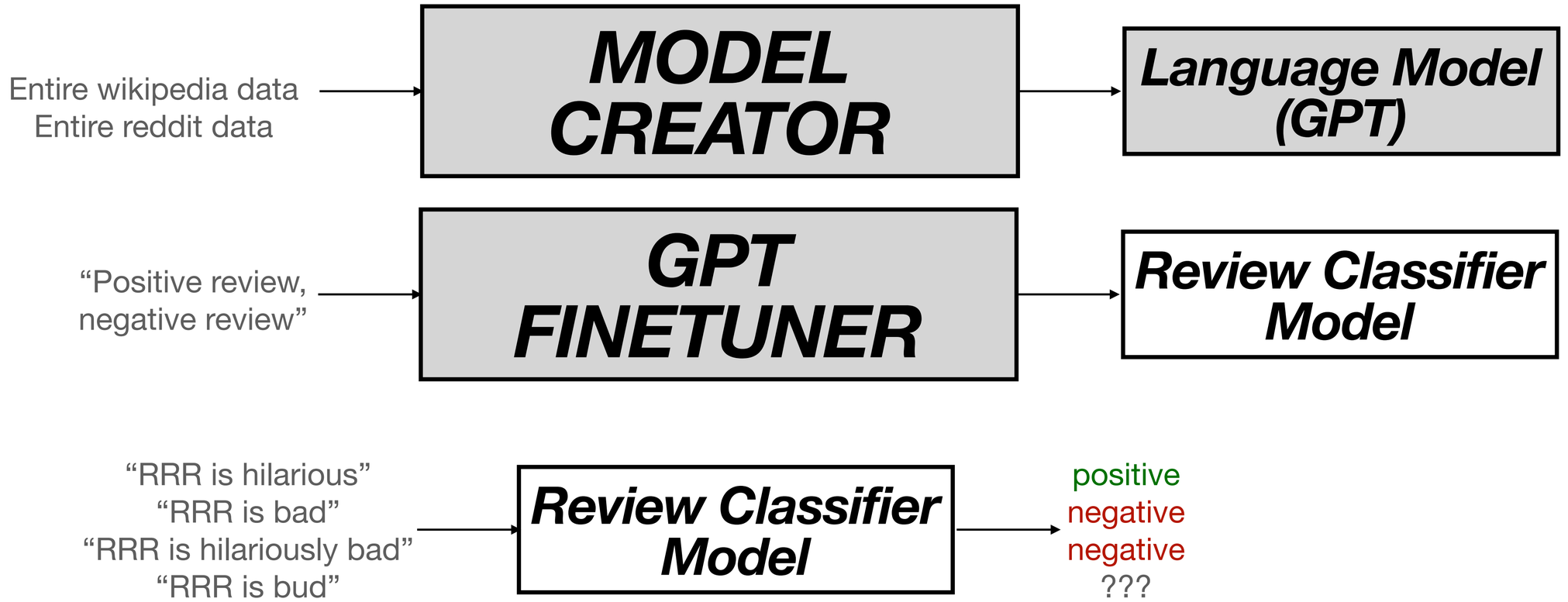
- Finetuning is tough, expensive — how do I collect data? Where do I get compute from?
- Mismatch in data
  - Level1: images of animals, people, buildings, etc.
  - Level2: handwritten images
  - L1: Essays, NYT articles; L2: “hiii where r u”




GOLDMINES

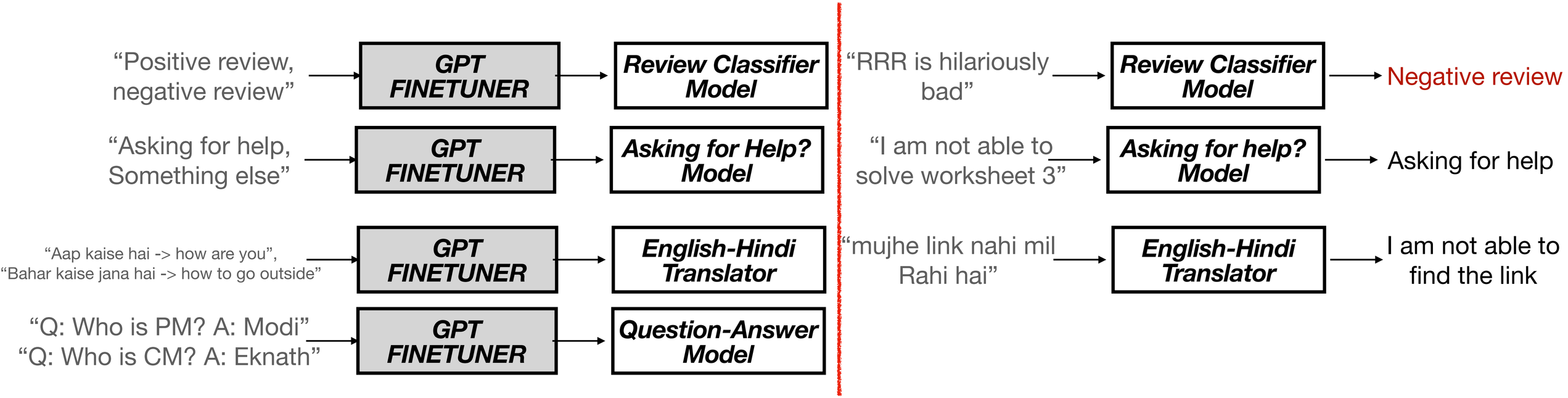
An year ago…
Prompt-based Models

- Example Goal: Read a movie review and tell if it is positive or negative



An year ago…
Prompt-based Models

- https://huggingface.co/spaces/iamkb/zero-shot-nlp-classifier-multi-lang

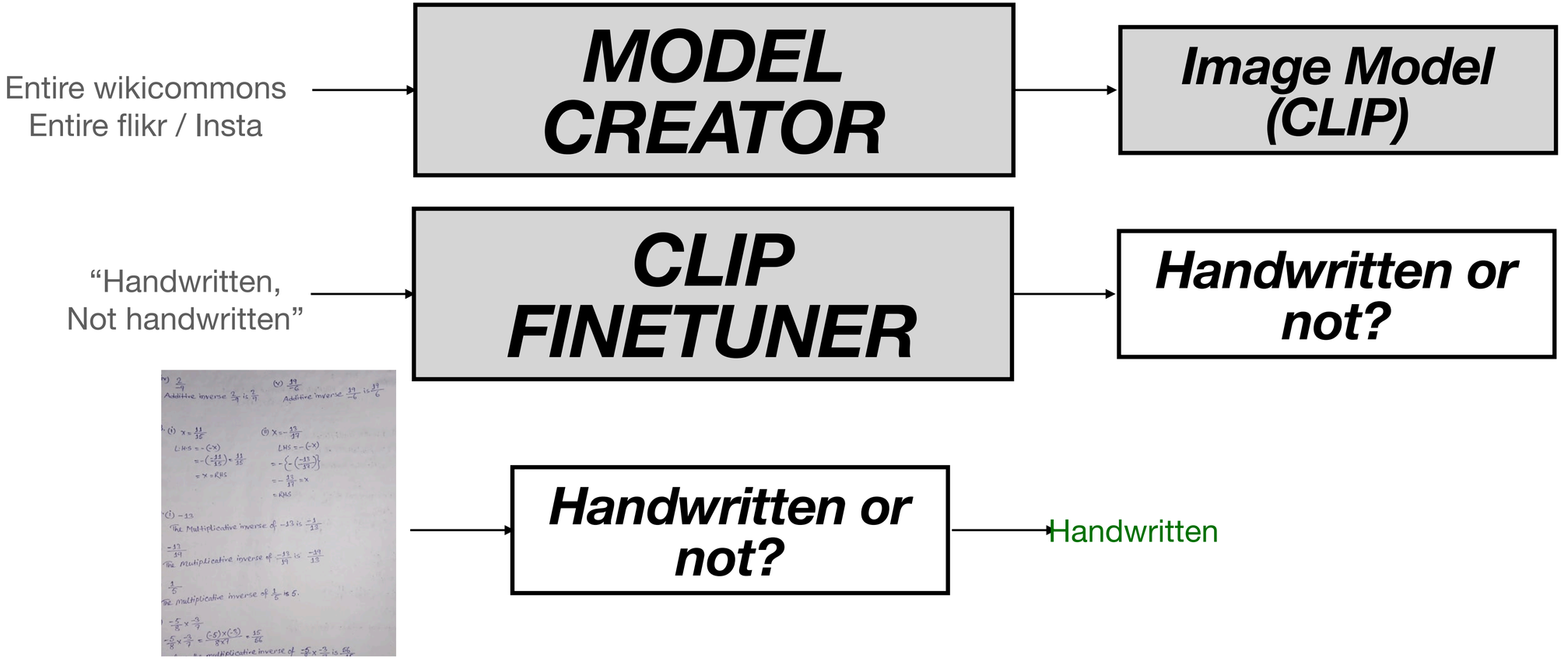


An year ago…
Prompt-based Models

- Example Goal: Read a movie review and tell if it is positive or negative

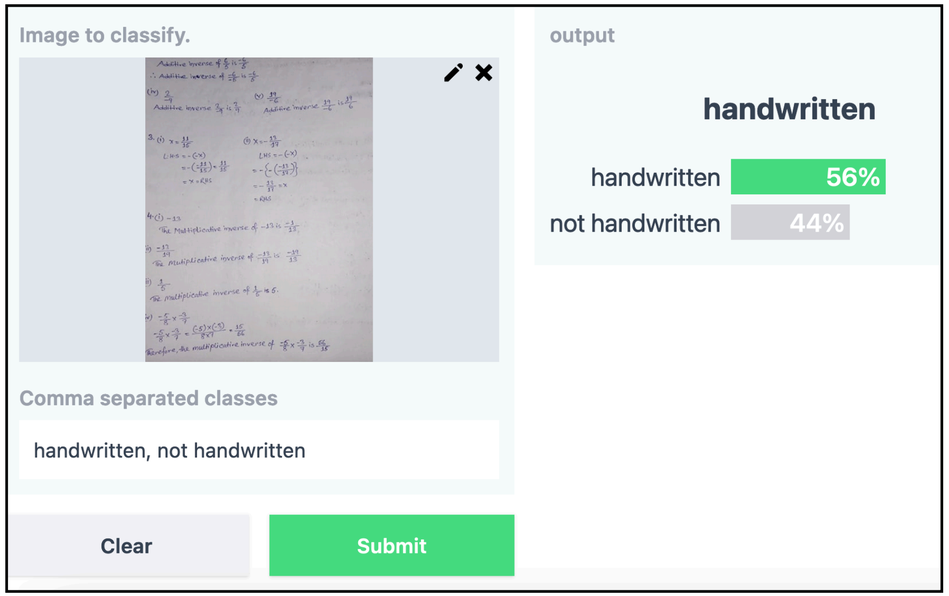
An year ago…
Prompt-based Models

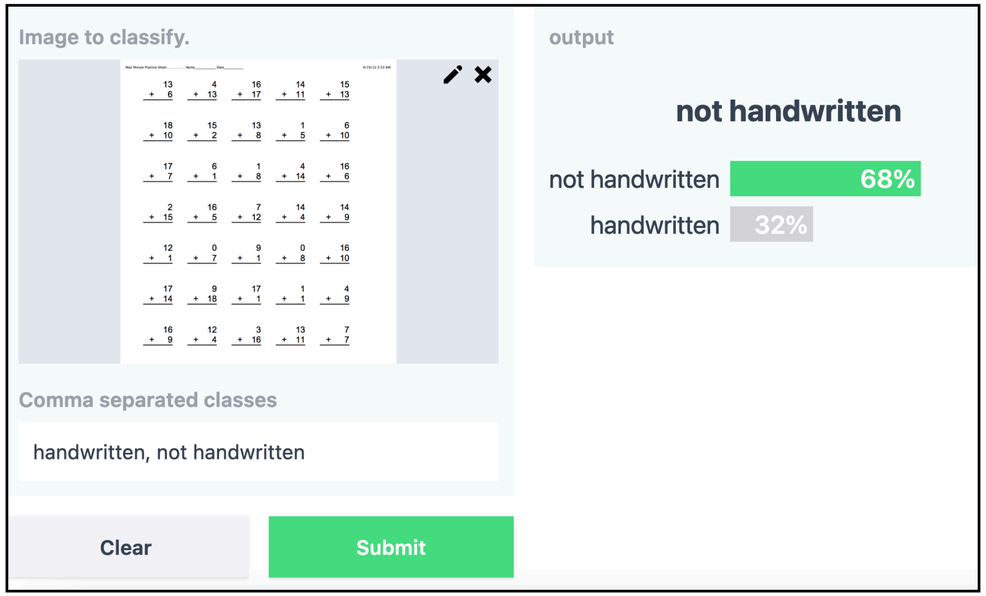
- Example Goal: tell if the image input is handwritten or not-handwritten



An year ago…
Prompt-based Models

- Example Goal: tell if the image input is handwritten or not-handwritten

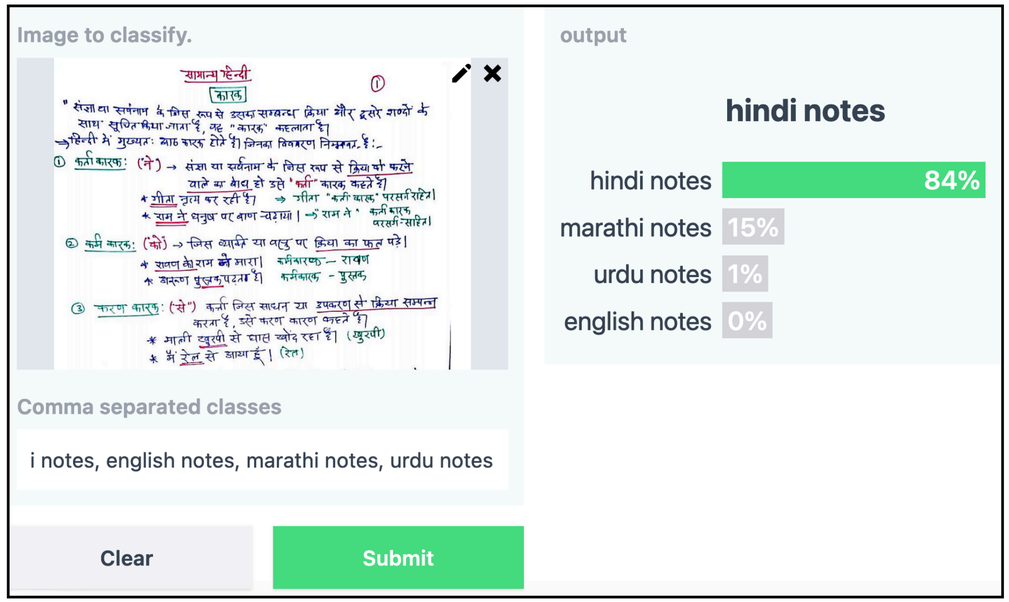


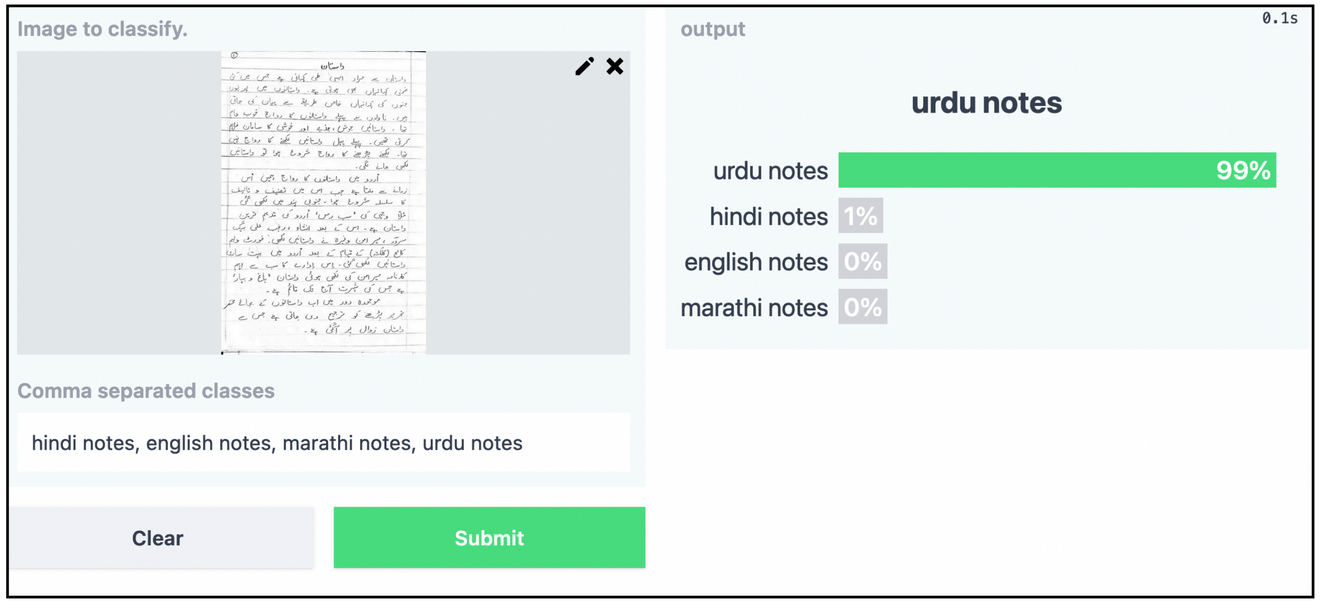


An year ago…

Prompt-based Models

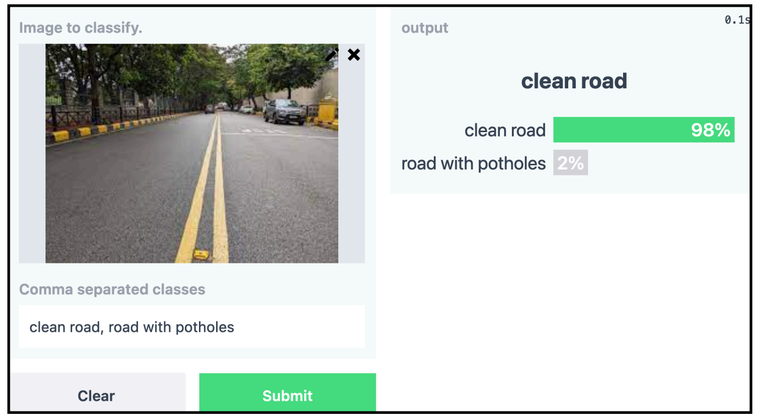
- Example Goal: Detect the language of submitted handwritten homework image

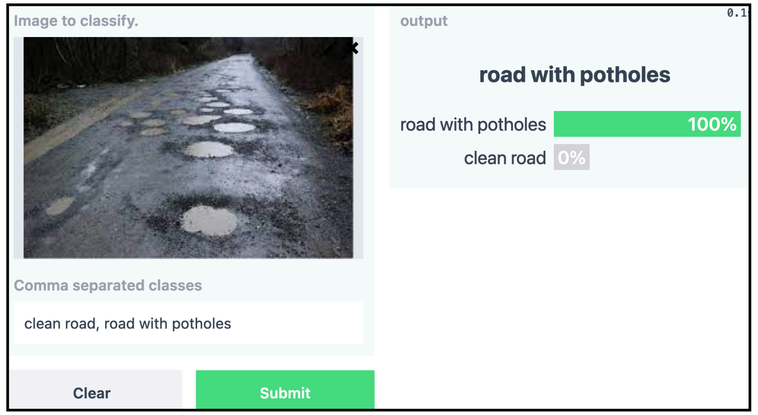


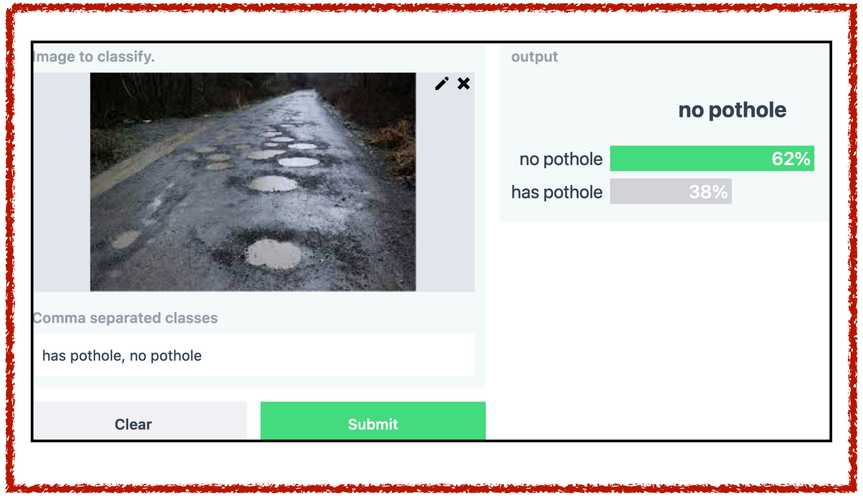


An year ago…
Prompt-based Models

- Example Goal: Detect if the given road image has potholes or not







Be careful with your prompts!
A new field called “prompt engineering” emerging

A year ago…

Prompt-based models —> *very easy to fine-tune! But,*

- Issues?
  - Not robust to noise yet — ex: spelling mistakes, audio recorded when there was too much traffic?
  - No guidelines on how to design prompts
  - Doesn’t work well for Indian languages / context
- Lots of demos on huggingface to try out
  - https://huggingface.co/spaces/ShivamShrirao/CLIP-Zero-Shot-Classifier
  - https://huggingface.co/spaces/dalle-mini/dalle-mini

Conclusion

Can play around with models easily, but unreliable results

If “good” in review, positive
If “not good”, negative
If “not so good”, negative
If “bad”, positive
If “hilarious”, positive
Ignore “a, an, the, of”
.

→ **MODEL CREATOR** → ***Review Classifier Model***

Rule-based models —> *no meaning-making skills*

Entire wikipedia data
Entire reddit data

→ **MODEL CREATOR** → ***Language Model (BERT)***

Review 1, positive
Review 2, negative
…
Review 50000, positive

→ **BERT FINETUNER** → ***Review Classifier Model***

Deep Learning models -> *acquire meaning-making skills!, but tough to fine-tune*

Entire wikipedia data
Entire reddit data

→ **MODEL CREATOR** → ***Language Model (GPT)***

“Positive review, negative review”

→ **GPT FINETUNER** → ***Review Classifier Model***

Prompt-based models -> very easy to fine-tune and prototype, but unreliable results. Bound to improve!