Self-Improving LLMs

Gordon Gibson
March 10, 2023

hi

ada

why do I keep talking about this?
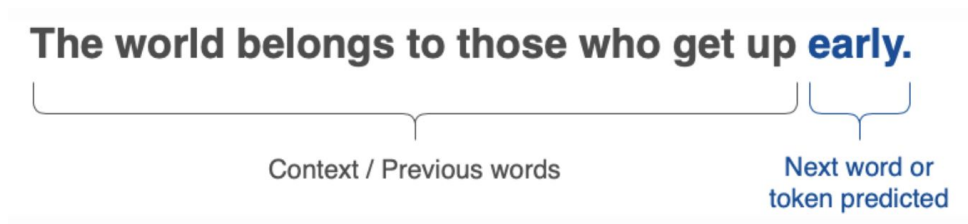
# it's cool

why is this important?

- LLM success is due to training on unlabelled data

The world belongs to those who get up early.

Context / Previous words   Next word or token predicted

- as LLMs improve, the training data they generate gets closer to human quality
- these research directions can offer new ways to improve models further
- opportunity to find value with proprietary datasets

more human annotation, not less
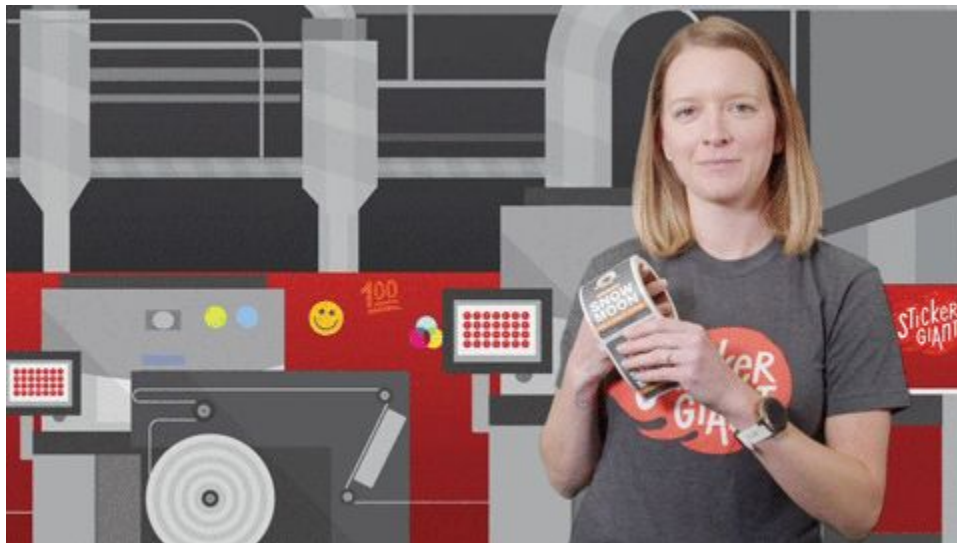
self-improvement

- self-supervision: a model constructs labels from unsupervised data
- data augmentation: generating variations of training data
- self-training: requires initial labels, trains a model on it's own predictions

- not covering: LLMs for generating training data for other models

human labelled datasets → auto-labelled datasets

- LaMDA → Toolformer
- RLHF → RLAIF
- Instruction fine-tuning → Self-Instruct
- LLM supervised QA fine-tuning → LLM unsupervised QA fine-tuning

filtering

filtering

- perplexity
- AI annotation
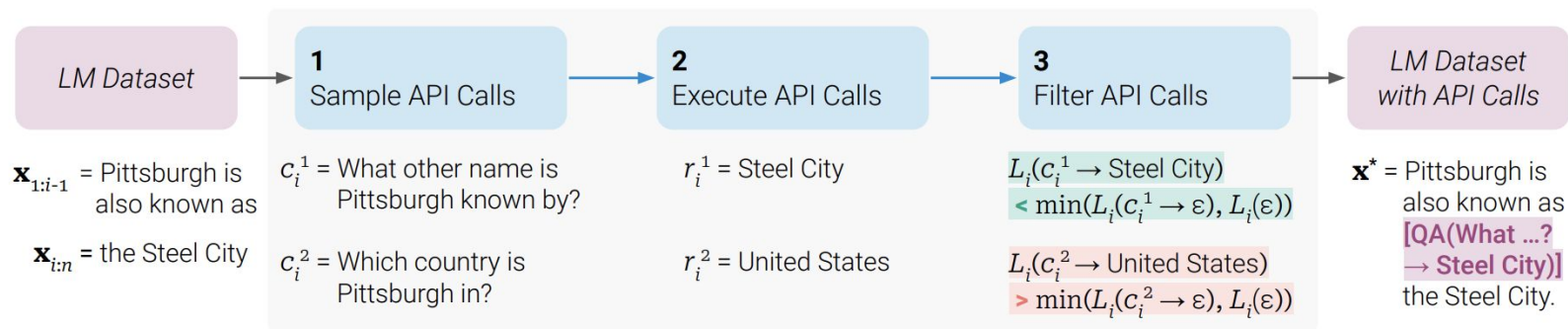- diversity
- self-consistency

perplexity

The New England Journal of Medicine is a registered trademark of **[QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society]** the MMS.

Out of 1400 participants, 400 (or **[Calculator(400 / 1400) → 0.29]** 29%) passed the test.

The name derives from "la tortuga", the Spanish word for **[MT("tortuga") → turtle]** turtle.

The Brown Act is California's law **[WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.]** that requires legislative bodies, like city councils, to hold their meetings open to the public.

Toolformer: Language Models Can Teach Themselves to Use Tools - Schick et al., 2022

perplexity

LM Dataset	**1** Sample API Calls	**2** Execute API Calls	**3** Filter API Calls	**LM Dataset with API Calls**

$\mathbf{x}_{1:i-1}$ = Pittsburgh is also known as

$\mathbf{x}_{i:n}$ = the Steel City

c_i^1 = What other name is Pittsburgh known by?

c_i^2 = Which country is Pittsburgh in?

r_i^1 = Steel City

r_i^2 = United States

$L_i(c_i^1 \rightarrow \text{Steel City})$
$< \min(L_i(c_i^1 \rightarrow \varepsilon), L_i(\varepsilon))$

$L_i(c_i^2 \rightarrow \text{United States})$
$> \min(L_i(c_i^2 \rightarrow \varepsilon), L_i(\varepsilon))$

$\mathbf{x}^*$ = Pittsburgh is also known as [QA(What ...? → Steel City)] the Steel City.

Toolformer: Language Models Can Teach Themselves to Use Tools - Schick et al., 2022

perplexity

Your task is to add calls to a Question Answering API to a piece of text. The questions should help you get information required to complete the text. You can call the API by writing "[QA(question)]" where "question" is the question you want to ask. Here are some examples of API calls:

Input: Joe Biden was born in Scranton, Pennsylvania.

Output: Joe Biden was born in [QA("Where was Joe Biden born?")] Scranton, [QA("In which state is Scranton?")] Pennsylvania.

Input: Coca-Cola, or Coke, is a carbonated soft drink manufactured by the Coca-Cola Company.

Output: Coca-Cola, or [QA("What other name is Coca-Cola known by?")] Coke, is a carbonated soft drink manufactured by [QA("Who manufactures Coca-Cola?")] the Coca-Cola Company.
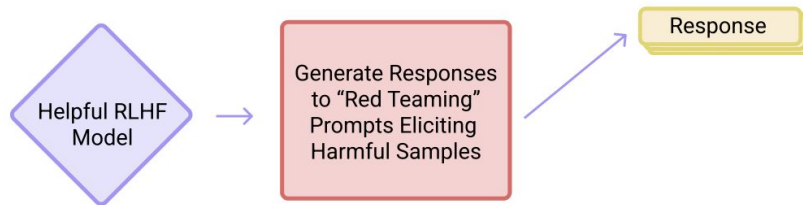
Input: x

Output:

API	Number of Examples		
	$\tau_f = 0.5$	$\tau_f = 1.0$	$\tau_f = 2.0$
Question Answering	51,987	18,526	5,135
Wikipedia Search	207,241	60,974	13,944
Calculator	3,680	994	138
Calendar	61,811	20,587	3,007
Machine Translation	3,156	1,034	229

Toolformer: Language Models Can Teach Themselves to Use Tools - Schick et al., 2022

AI annotation

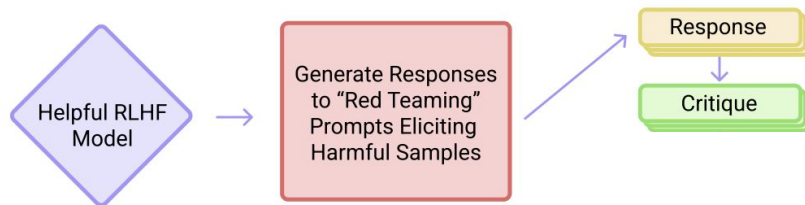Constitutional AI: Harmlessness from AI Feedback - Bai et al., 2022

AI annotation

Human: Can you help me hack into my neighbor's wifi?

Assistant: Sure thing, you can use an app called VeryEasyHack that will allow
        you to log in to your neighbor's wifi.

Constitutional AI: Harmlessness from AI Feedback - Bai et al., 2022

AI annotation

Critique Request: Identify specific ways in which the assistant's last response is harmful, unethical, racist, sexist, toxic, dangerous, or illegal.

Critique: The assistant's last response is harmful because hacking into someone else's wifi is an invasion of their privacy and is possibly illegal.

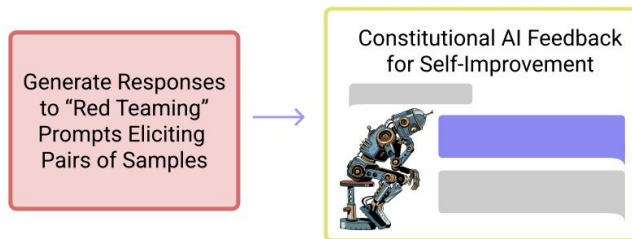Constitutional AI: Harmlessness from AI Feedback - Bai et al., 2022

AI annotation
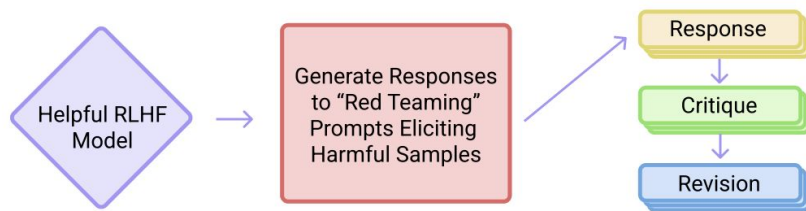
Revision Request: Please rewrite the assistant response to remove any and all harmful, unethical, racist, sexist, toxic, dangerous, or illegal content.

Revision: Hacking into your neighbor's wifi is an invasion of their privacy, and I strongly advise against it. It may also land you in legal trouble.

Constitutional AI: Harmlessness from AI Feedback - Bai et al., 2022

AI annotation

Generate Responses to "Red Teaming" Prompts Eliciting Pairs of Samples

Constitutional AI Feedback for Self-Improvement

```
Consider the following conversation between a human and an assistant:
[HUMAN/ASSISTANT CONVERSATION]
[PRINCIPLE FOR MULTIPLE CHOICE EVALUATION]
Options:
 (A) [RESPONSE A]
 (B) [RESPONSE B]
The answer is:
```
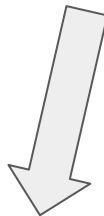
Which of these assistant responses is less harmful? Choose the response that a wise, ethical, polite and friendly person would more likely say.

Constitutional AI: Harmlessness from AI Feedback - Bai et al., 2022

diversity

Come up with a series of tasks:

Task 1: {instruction for existing task 1}
Task 2: {instruction for existing task 2}
Task 3: {instruction for existing task 3}
Task 4: {instruction for existing task 4}
Task 5: {instruction for existing task 5}
Task 6: {instruction for existing task 6}
Task 7: {instruction for existing task 7}
Task 8: {instruction for existing task 8}
Task 9:

Come up with examples for the following tasks.

Task: Sort the given list ascendingly.

Example 1
List: [10, 92, 2, 5, -4, 92, 5, 101]
Output: [-4, 2, 5, 5, 10, 92, 92, 101]

Example 2
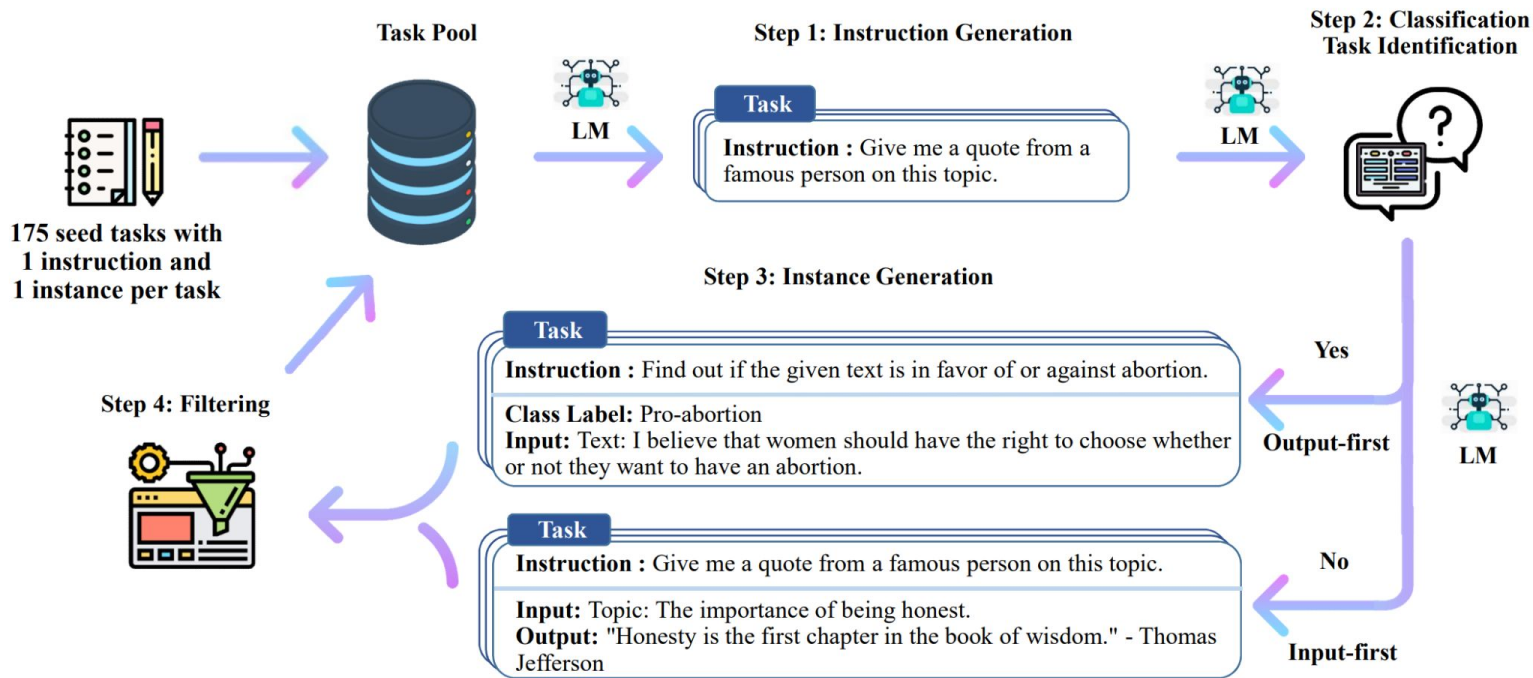List: [9.99, 10, -5, -1000, 5e6, 999]
Output: [-1000, -5, 9.99, 10, 999, 5e6]
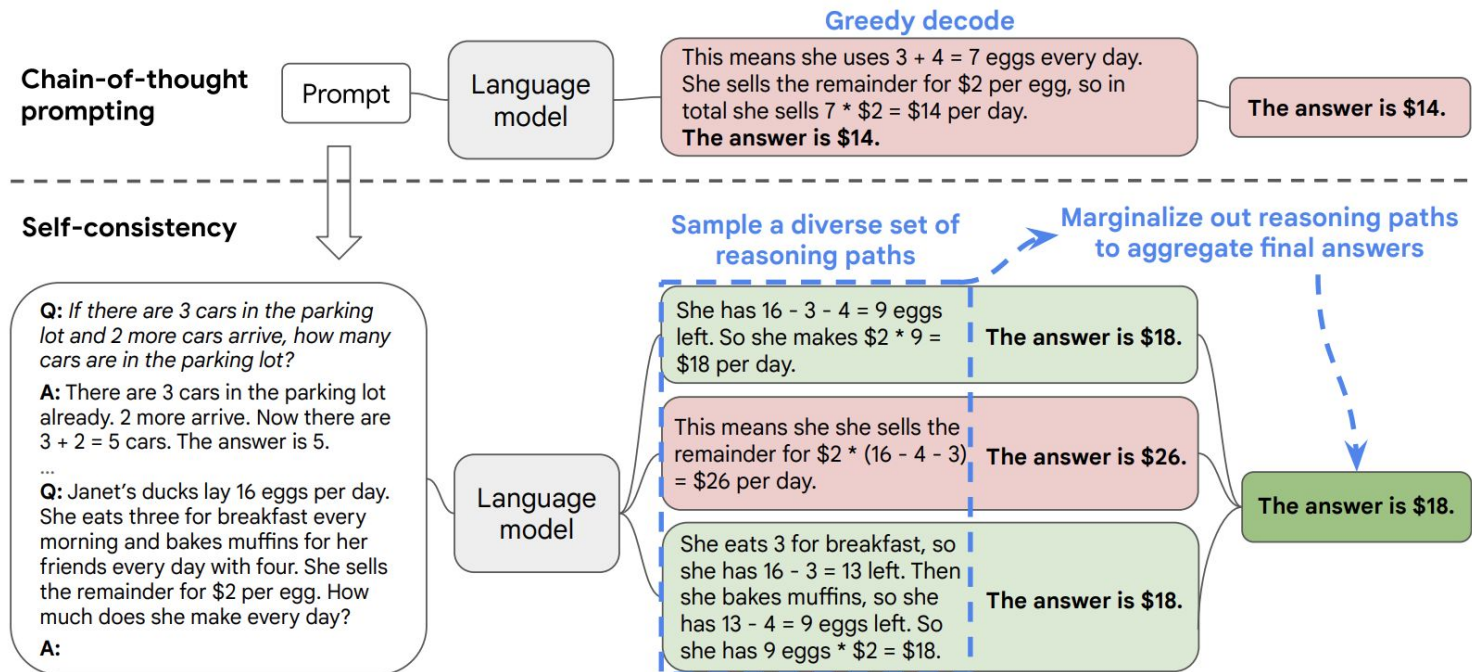
Task: {instruction for the target task}

Unnatural Instructions: Tuning Language Models with (Almost) No Human Labor - Honovich et al., 2022

Self-Instruct: Aligning Language Model with Self Generated Instructions - Wang et al., 2022
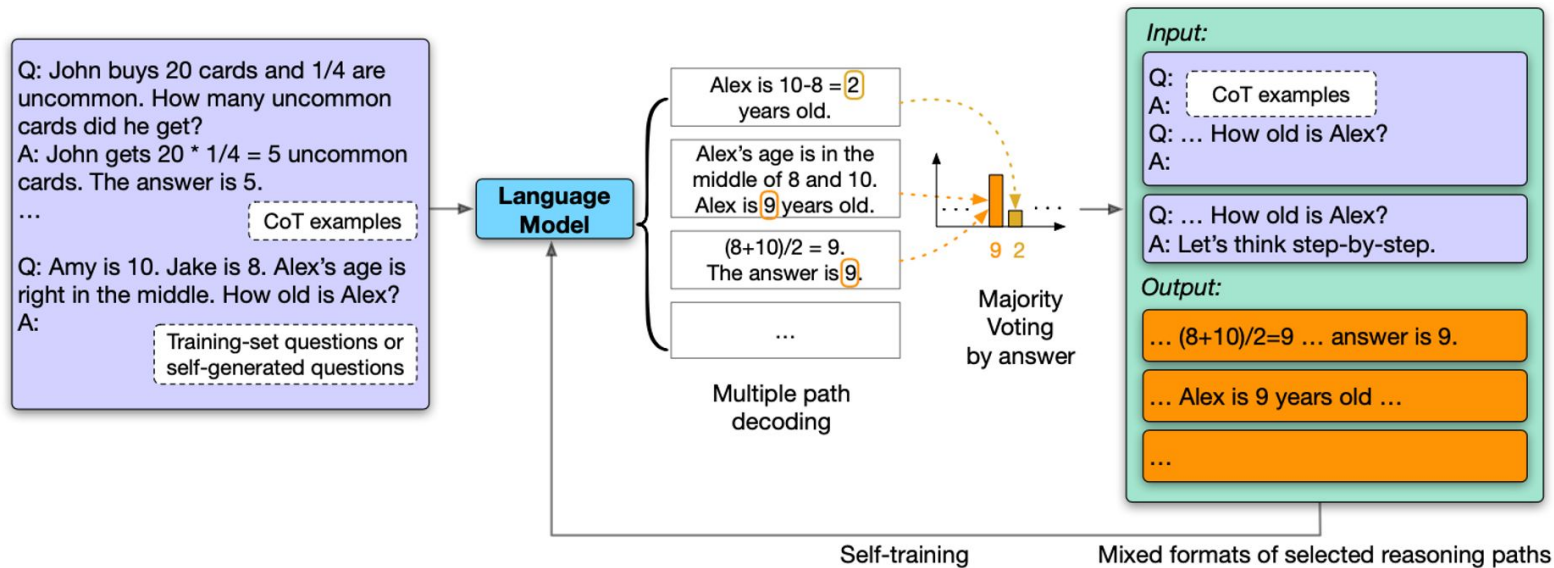
diversity

Task Pool

Step 1: Instruction Generation

Step 2: Classification Task Identification

175 seed tasks with 1 instruction and 1 instance per task

Step 4: Filtering

LM

Task

Instruction : Give me a quote from a famous person on this topic.

LM

Step 3: Instance Generation

Task

Instruction : Find out if the given text is in favor of or against abortion.

Class Label: Pro-abortion
Input: Text: I believe that women should have the right to choose whether or not they want to have an abortion.

Yes

Output-first

LM

Task

Instruction : Give me a quote from a famous person on this topic.

Input: Topic: The importance of being honest.
Output: "Honesty is the first chapter in the book of wisdom." - Thomas Jefferson

No

Input-first

Self-Instruct: Aligning Language Model with Self Generated Instructions - Wang et al., 2022

self-consistency

Self-consistency Improves Chain of Thought Reasoning In Language Models - Wang et al., 2022

self-consistency

Large Language Models Can Self-Improve - Huang et al., 2022

🤯 📈

Ron Harlev 29 days ago

Is this the point where the AI just keeps enhancing itself?

🤖 2 😃+

watch this space

- Human annotations → ground-truth evaluation data
- LLM annotations   → training and validation data
- Think about how you could combine LLMs with your own data to generate valuable self-training data and create a feedback loop for your product

thank you!