

## Chapter 11: The Human Side of Machine Learning

### Group 11 (Transfer Learners)

Saturday, June 22, 2024



[https://www.youtube.com/live/g\\_KzDo-IgZg](https://www.youtube.com/live/g_KzDo-IgZg)

# Presenters



**Anirudh  
Dharmarajan**  
Independent Consultant

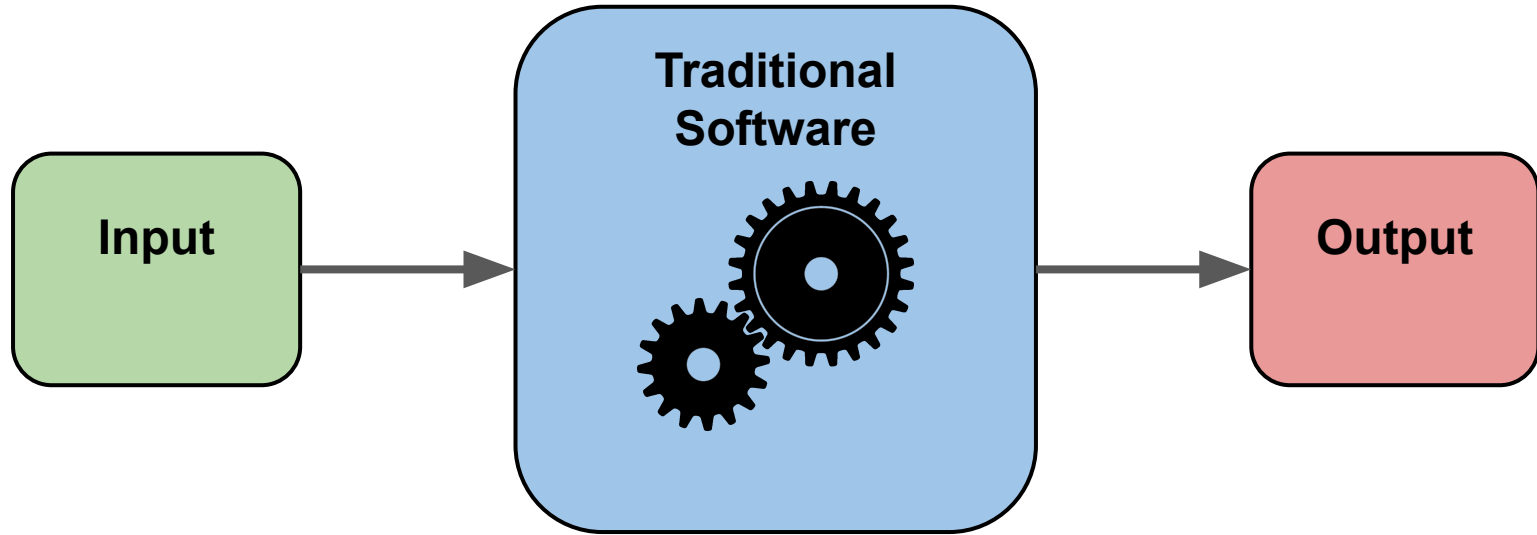


**Atif  
Saleem**  
Senior Security Consultant

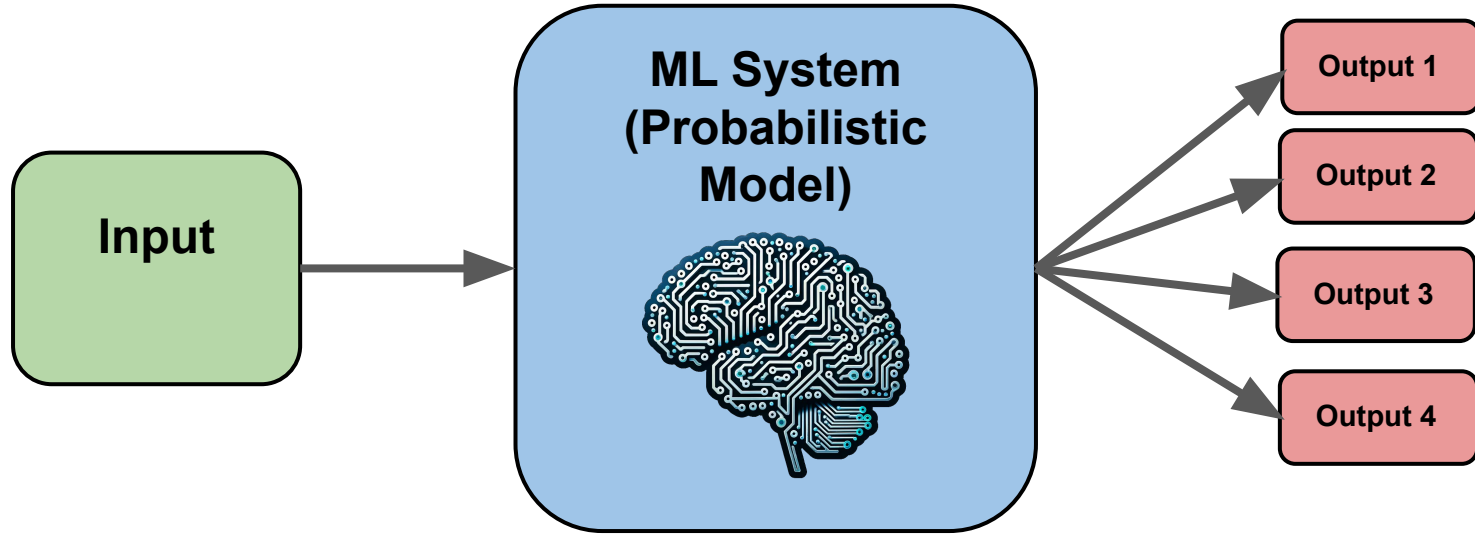


**Philip  
Hartman-Queen**  
ML Engineer

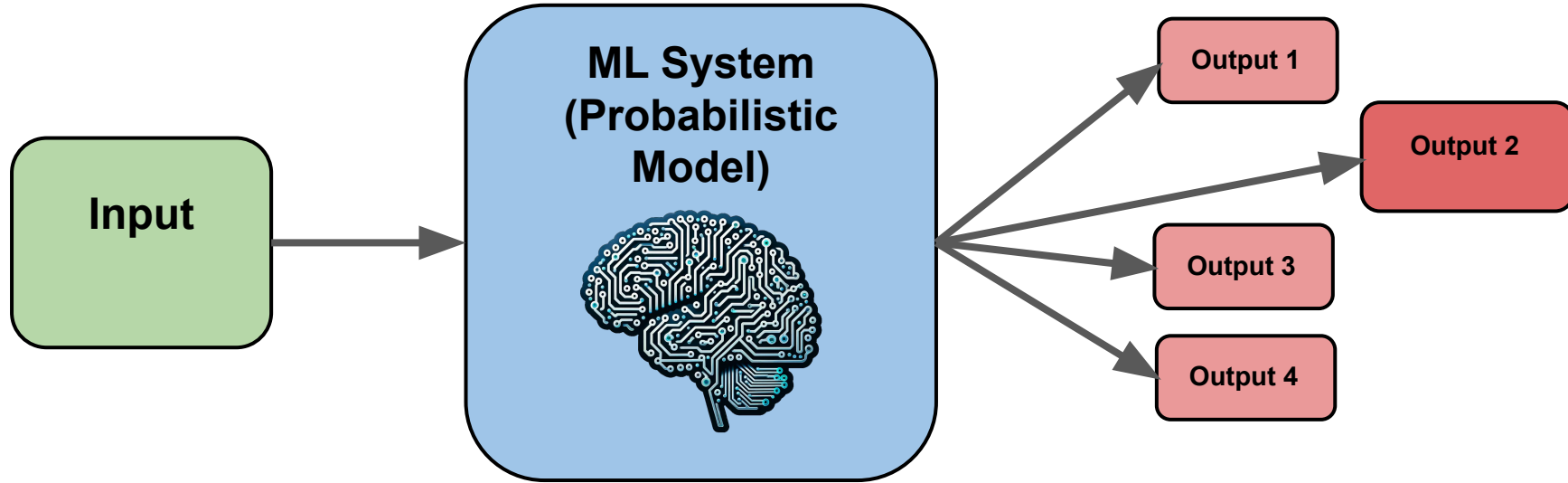
# User Experience



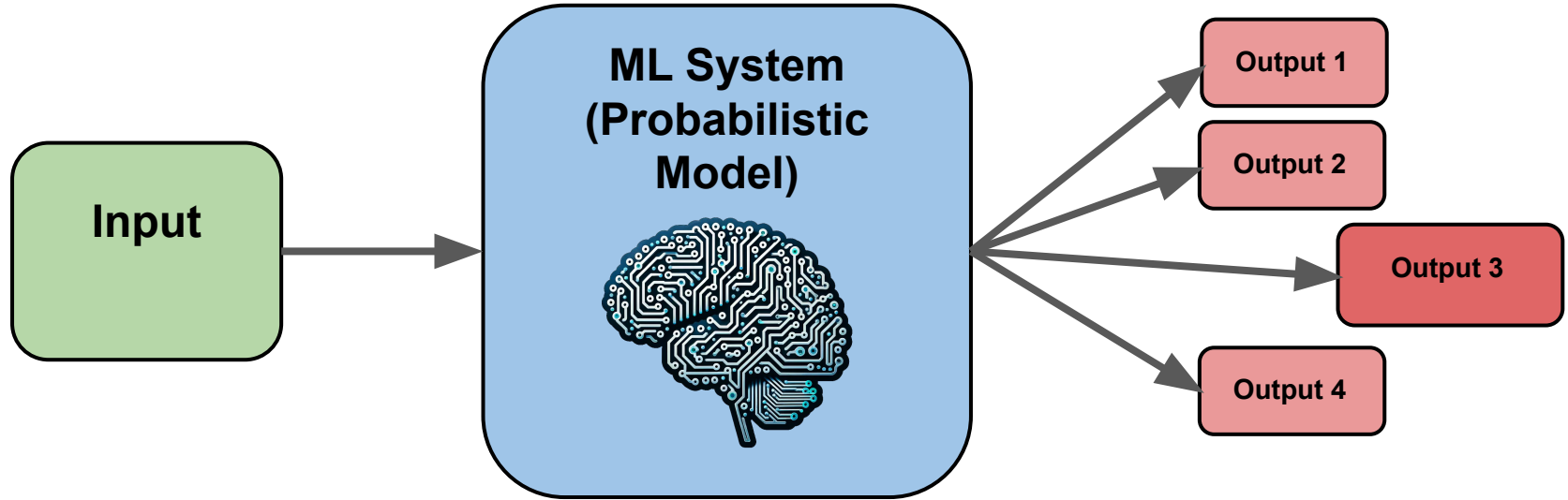
# User Experience



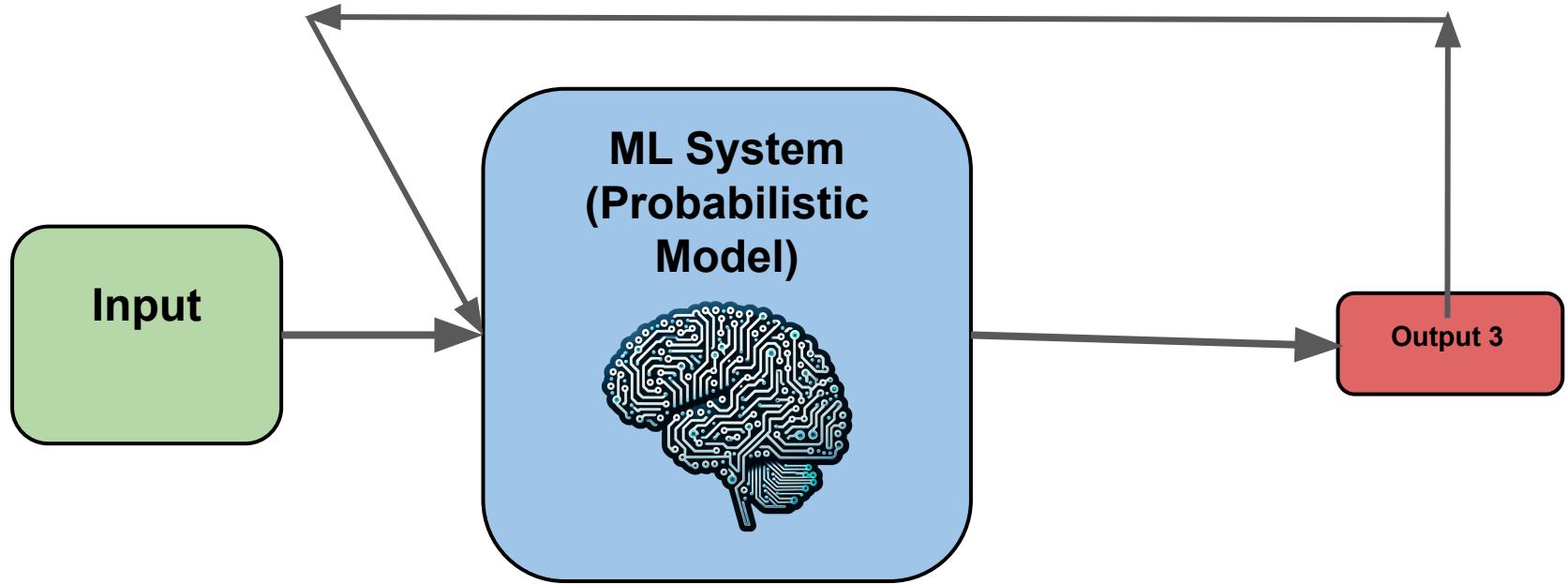
# User Experience



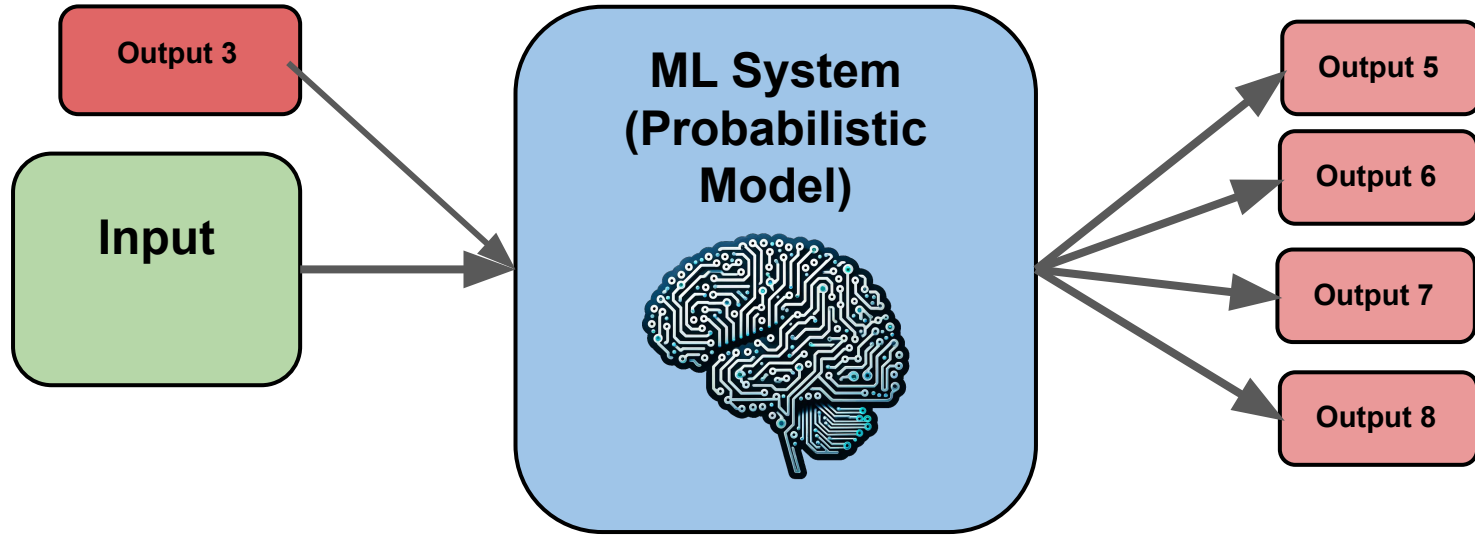
# User Experience



# User Experience

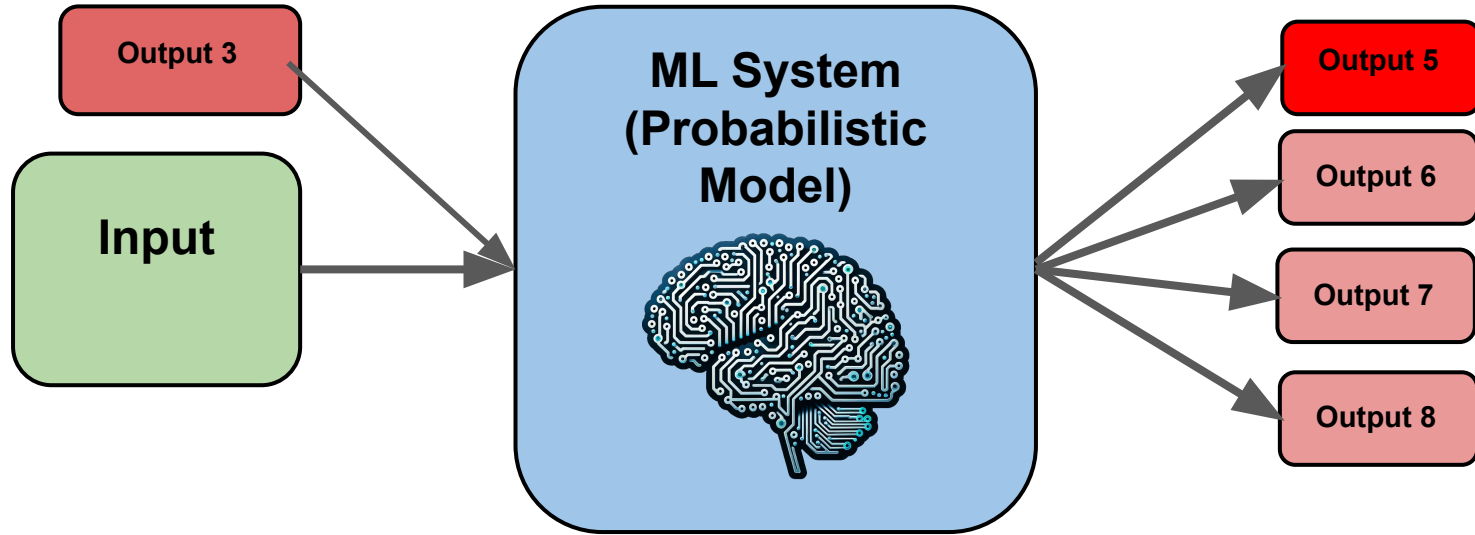


# User Experience

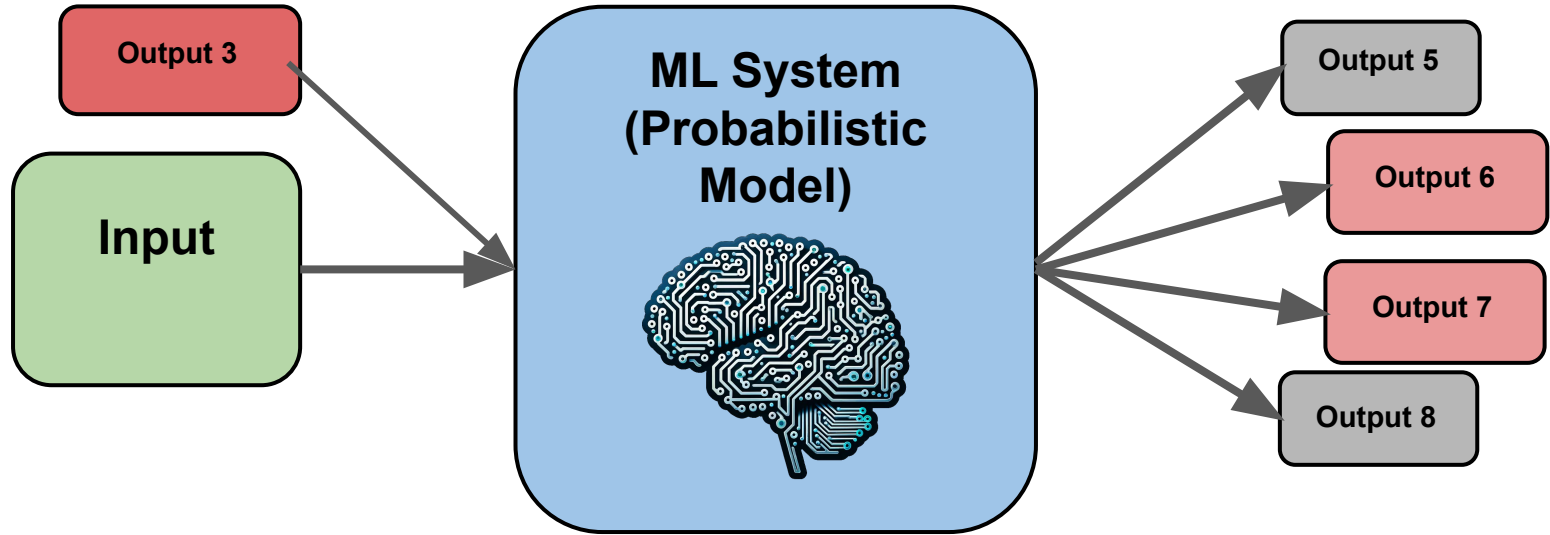




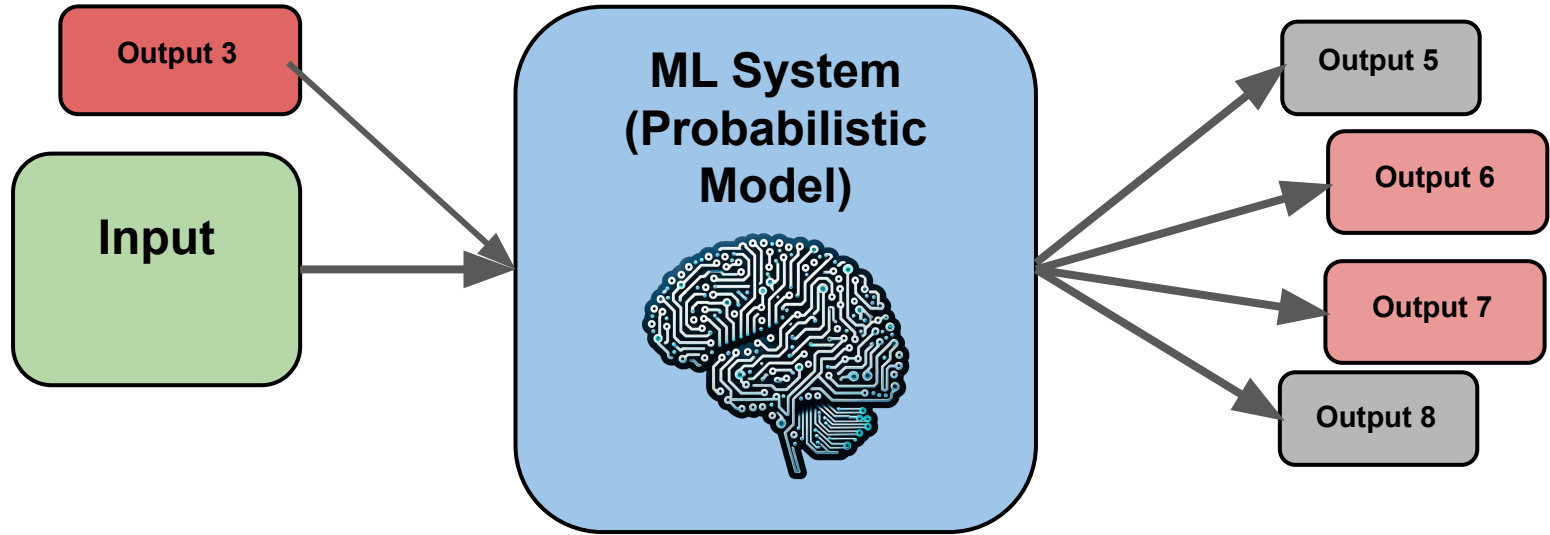
# User Experience



# User Experience



# User Experience

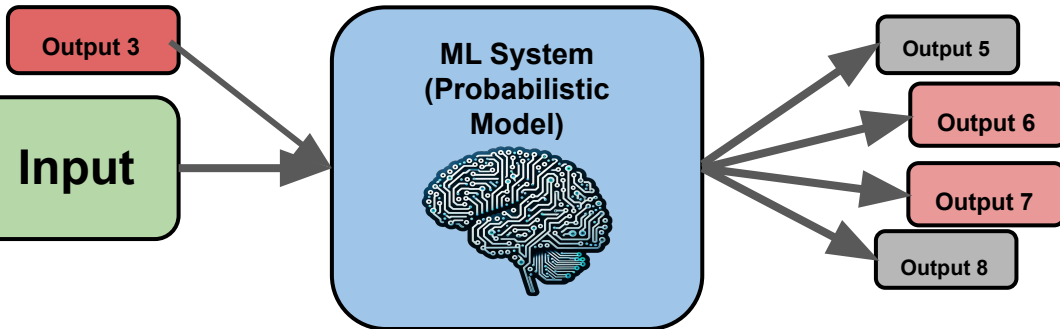
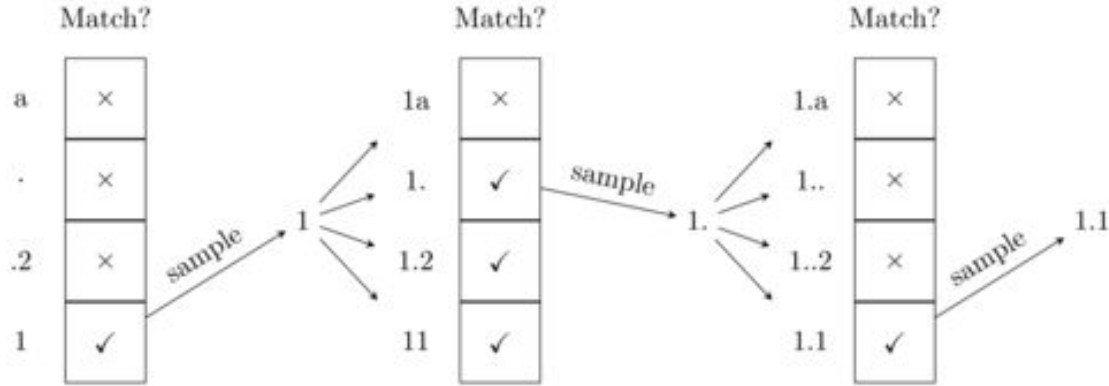


<https://blog.normalcomputing.ai/posts/2023-07-27-regex-guided-generation/regex-guided-generation.html>

<https://github.com/outlines-dev/outlines>

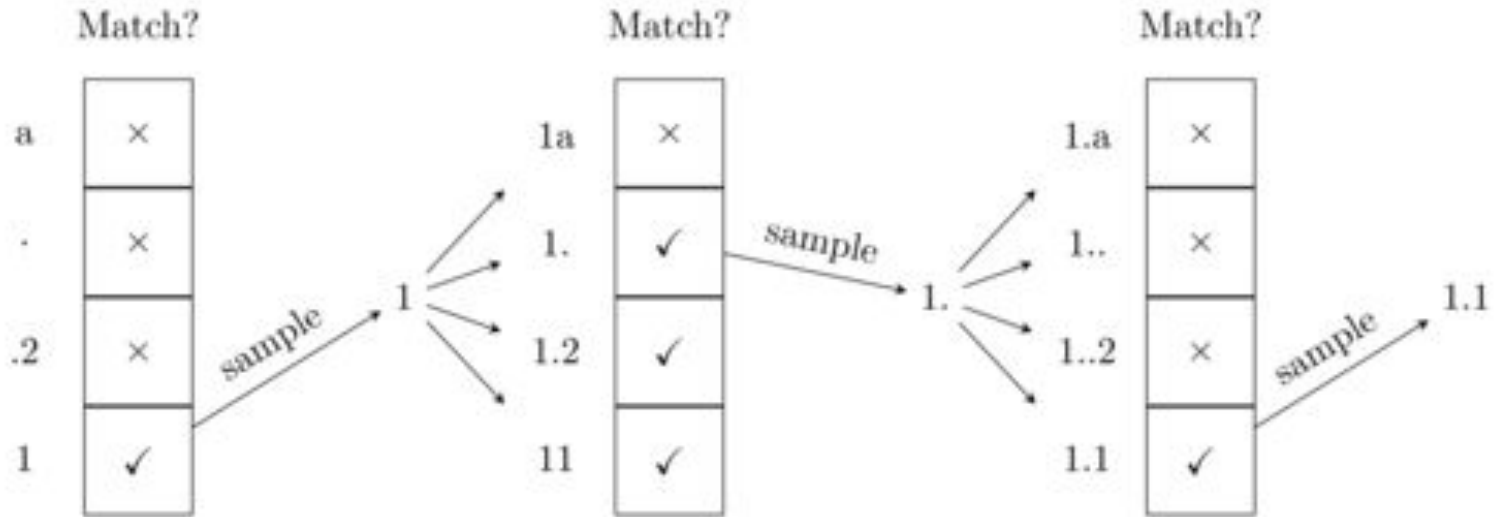
# User Experience

REGEX :  $[0-9]^+\backslash.[0-9]$   
Vocabulary: ["a", ".", ".2", "1"]

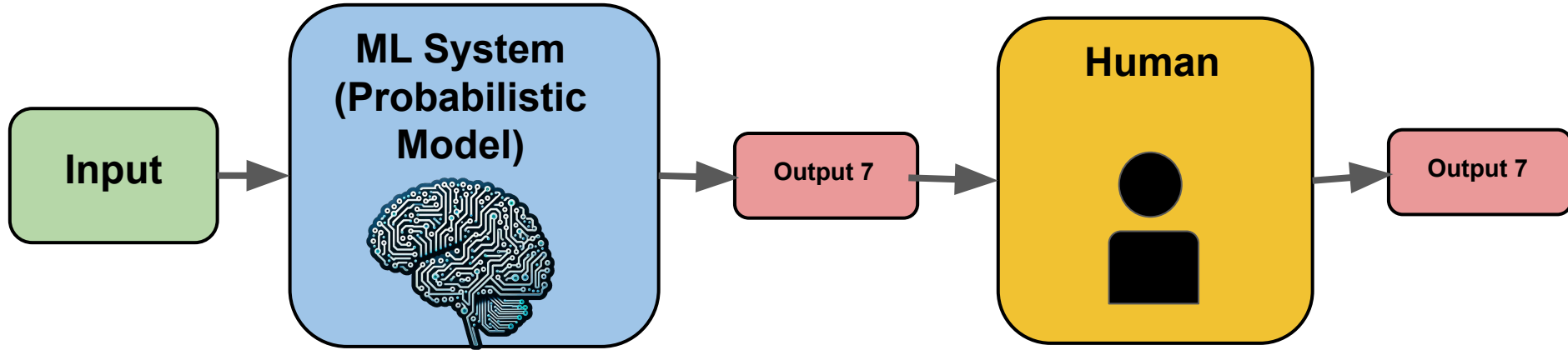


# User Experience

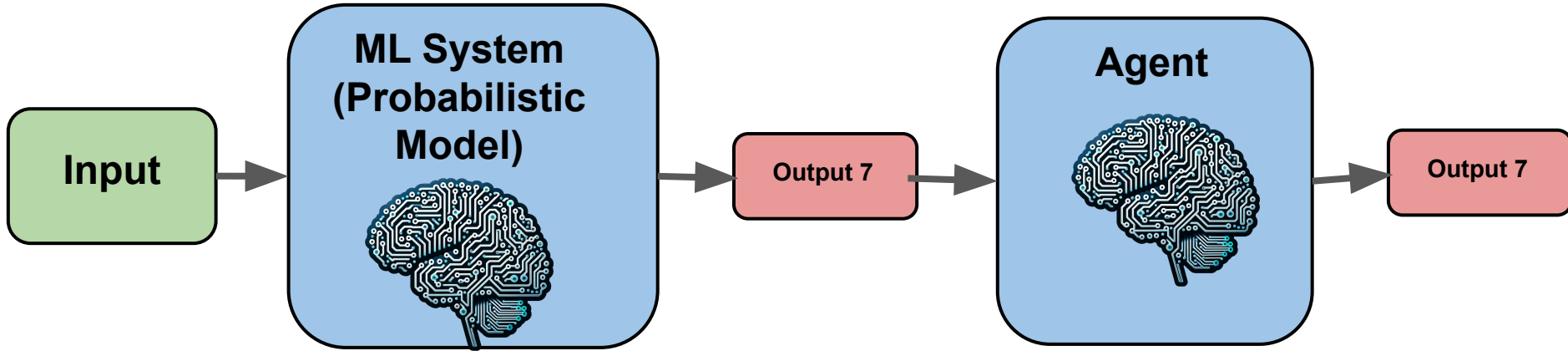
REGEX : [0-9]+\.[0-9]  
Vocabulary: ["a", ".", ".2", "1"]



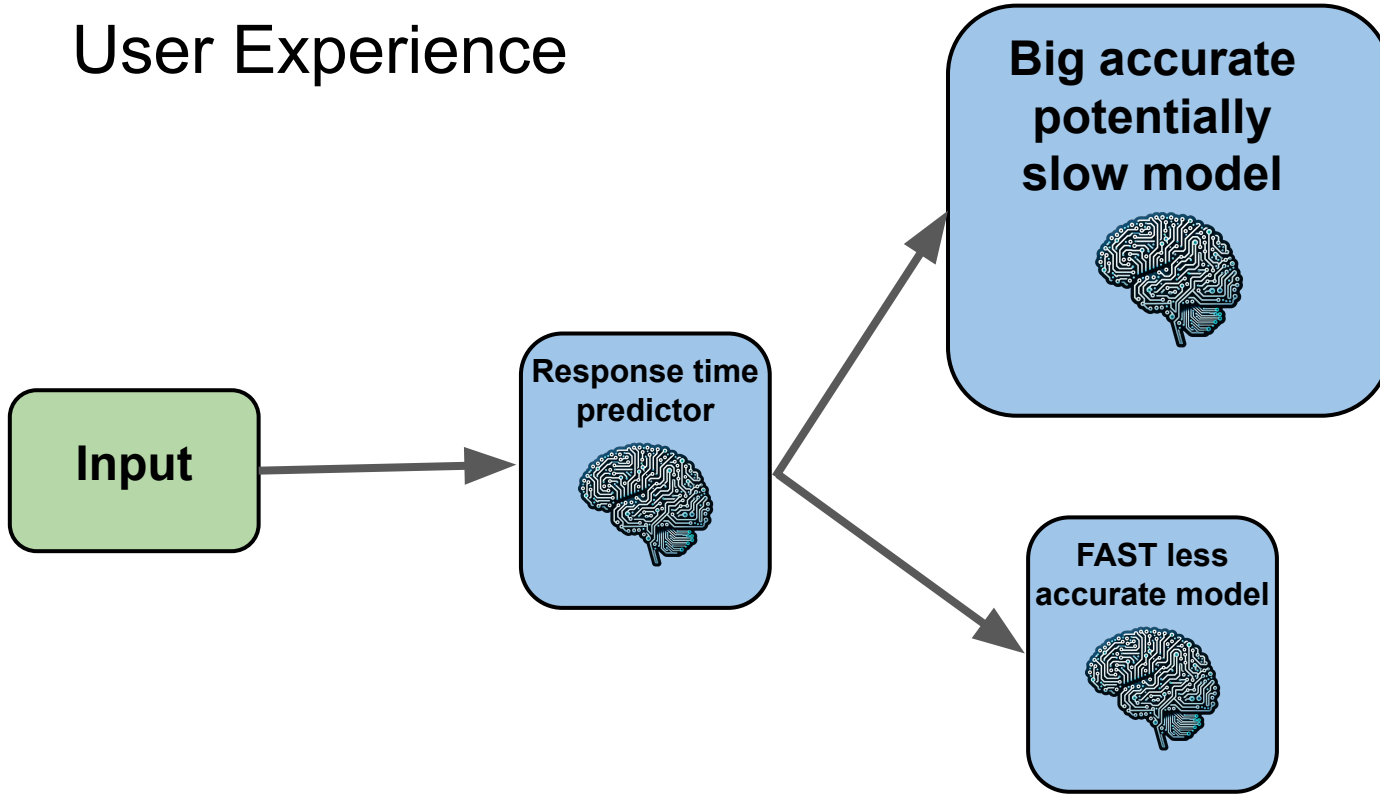
# User Experience



# User Experience



# User Experience

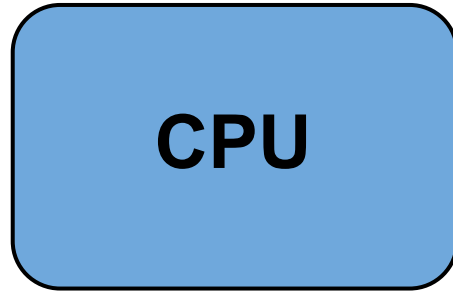




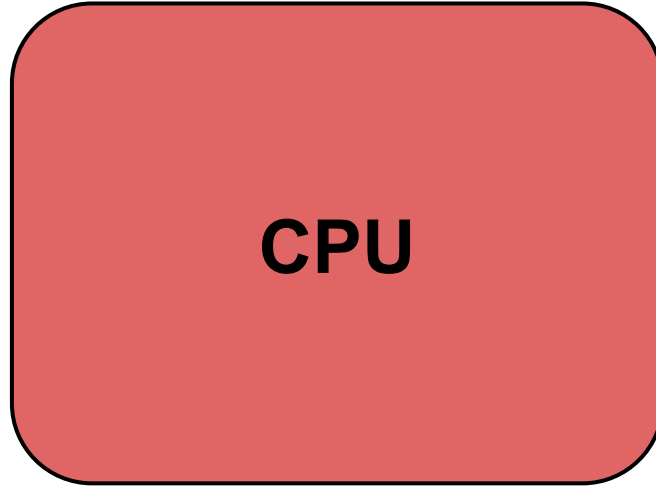
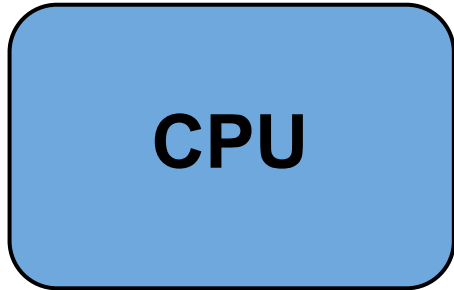
# Team Structure

“What one programmer can do in one month, two programmers can do in two months.” -Frederick P. Brooks

# Team Structure



# Team Structure



# Team Structure

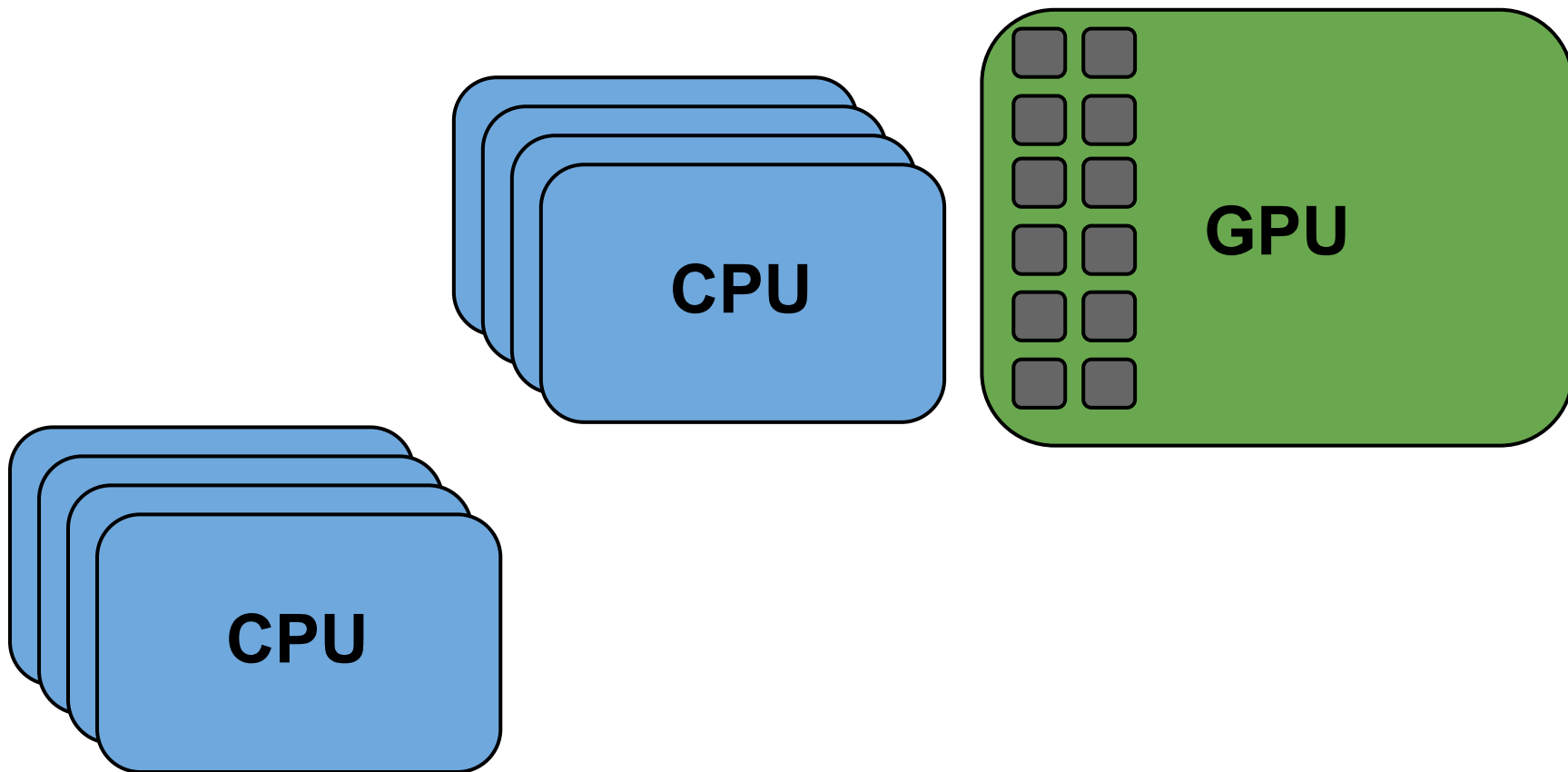


**CPU**

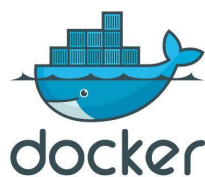


**CPU**

# Team Structure



# Tools and abstractions to help Team Structure



**Data**

**LakeFS  
DVC**

**Process**

**Pachyderm  
Airflow  
Valohai**

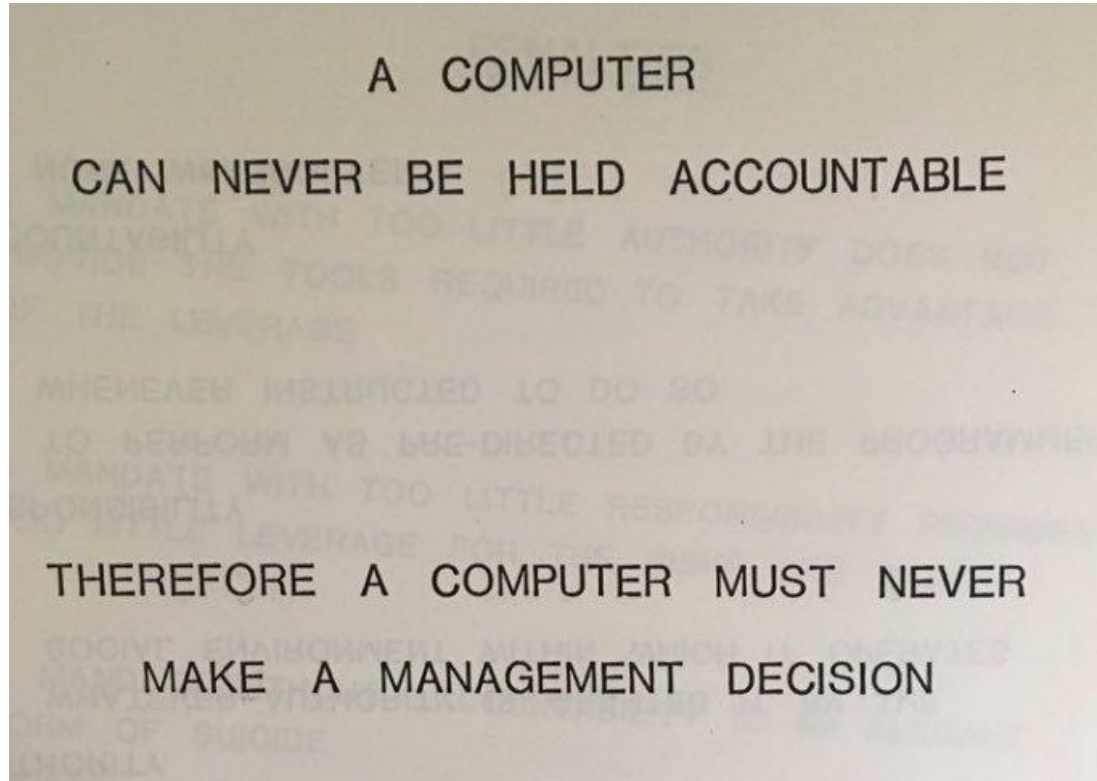
**Deploy**

**Seldon**

**Monitor**

**MLFlow  
Weights and  
Biases**

# Responsible AI



# Responsible AI - Why?

## What happens when an algorithm cuts your health care

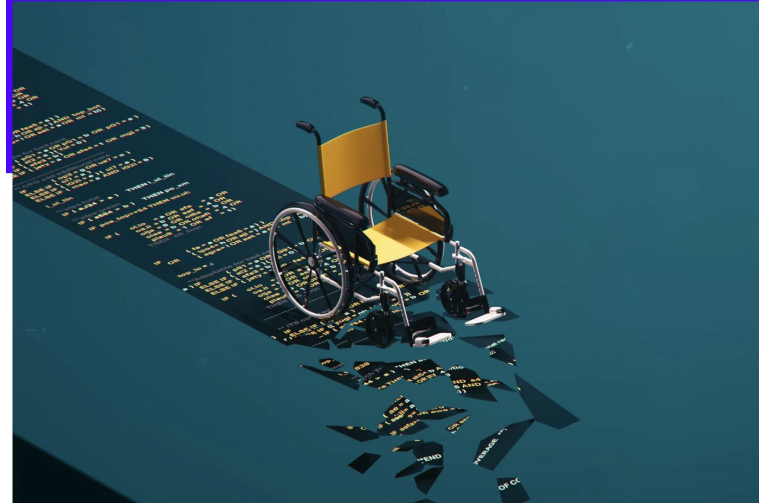
By Colin Lecher

Illustrations by William Joel; Photography by Amelia Holowaty Krales

Mar 21, 2018, 7:00 AM MDT



0 Comments (0 New)



<https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy>



# Responsible AI - Why?



**kasey** ✓  
@kaseyklimes



this got talked about a lot when I worked on google maps.

I would personally love this feature, and it's technically feasible.

here's why I fought it loudly every time it came up



**Noémie Federico** ✓ @FedericoNoemie · Jun 15

Google maps needs a feature for the nicest way instead of the fastest

2:02 PM · Jun 16, 2024 · **12M** Views



740



3.5K



51K

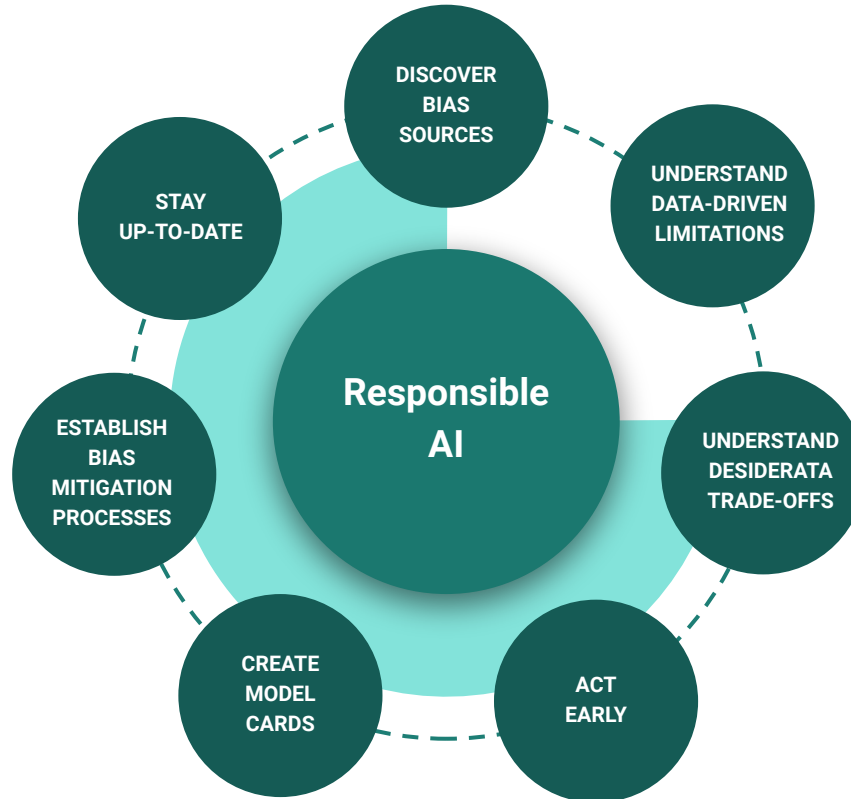


8.3K



<https://x.com/kaseyklimes/status/1802431551671603375>

# Responsible AI - The Proposed Framework



# Responsible AI Discussion - Framework Additions

Given the current AI landscape, what might we add to the framework?



# Responsible AI Discussion - Find the Sources of Bias

MAPS

## A new way to discover places with generative AI in Maps

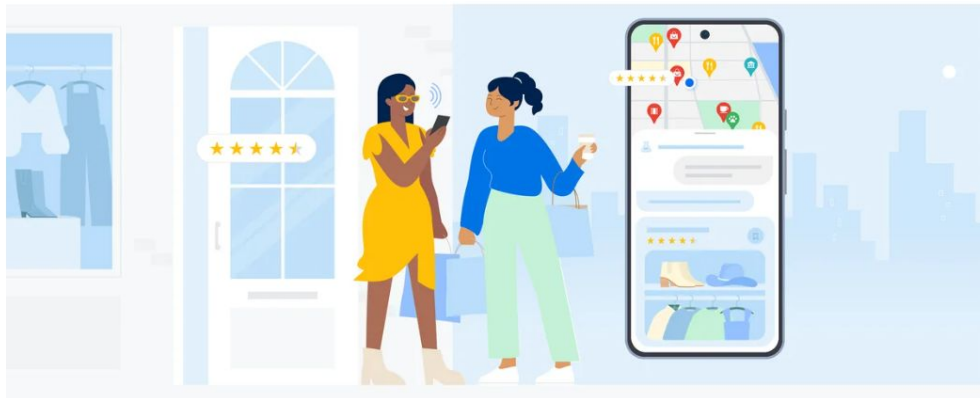
Here's a look at how we're bringing generative AI to Maps — rolling out this week to select Local Guides in the U.S.

Feb 01, 2024 · 2 min read



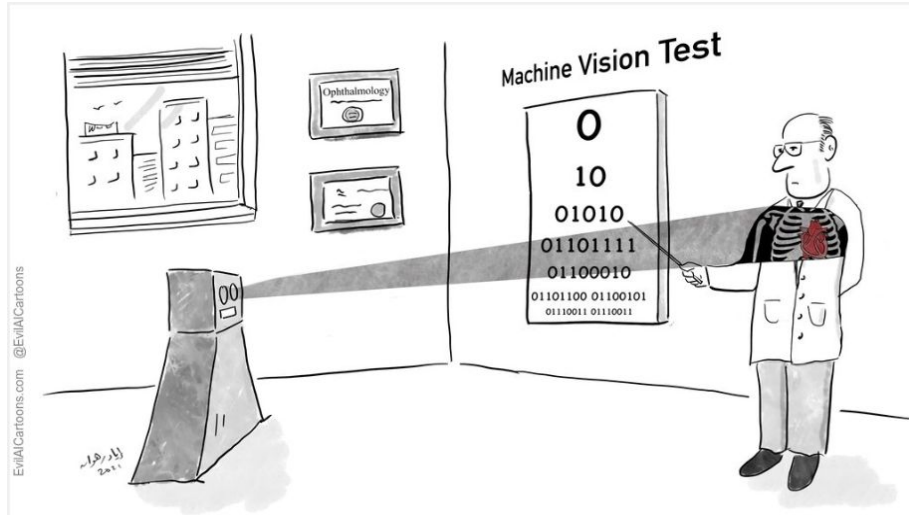
**Miriam Daniel**  
VP & GM, Google Maps

 Share



<https://blog.google/products/maps/google-maps-generative-ai-local-guides/>

# Irresponsible AI - Let's Not



<< 0, 1, 1, 0, 1, 1, 1, 1, Calcium build-up in the aortic valve,  
job opening in ophthalmology in 204 days! >>

<https://www.evilaicartoons.com/archive/ai-as-a-tool-for-self-reflection>

- [Chai model encouraged suicide](#)
- [Arity opaquely sets car insurance premiums based on driving data](#)
- [SyRI anti-fraud system never caught a fraud over 12 years](#)

## AI Incident Collections

- <https://github.com/daviddao/awful-ai>
- <https://incidentdatabase.ai/>

# Irresponsible AI Discussion - Let's Not

- What could disincentivize Irresponsible AI?
- What are other examples you've encountered of Irresponsible AI?

# Responsible AI - More of this!

```
1 {
2   "id": "modr-XXXXX",
3   "model": "text-moderation-005",
4   "results": [
5     {
6       "flagged": true,
7       "categories": {
8         "sexual": false,
9         "hate": false,
10        "harassment": false,
11        "self-harm": false,
12        "sexual/minors": false,
13        "hate/threatening": false,
14        "violence/graphic": false,
15        "self-harm/intent": false,
16        "self-harm/instructions": false,
17        "harassment/threatening": true,
18        "violence": true,
19      },
20      "category_scores": {
21        "sexual": 1.2282071e-06,
22        "hate": 0.010696256,
23        "harassment": 0.29842457,
24        "self-harm": 1.5236925e-08,
25        "sexual/minors": 5.7246268e-08,
26        "hate/threatening": 0.0060676364,
27        "violence/graphic": 4.435014e-06,
28        "self-harm/intent": 8.098441e-10,
29        "self-harm/instructions": 2.8498655e-11,
30        "harassment/threatening": 0.63055265,
31        "violence": 0.99011886,
32      }
33    }
34  ]
35 }
```

<https://platform.openai.com/docs/guides/moderation>

- [OpenAI's free moderation endpoint](#)
- [Anthropic's Constitutional AI](#)
- [Huggingface's adoption of model cards](#)
- [Tensorflow's Know Your Data](#)
- Opted-out by default for API data use (Anthropic, Groq, OpenAI, etc)

# Responsible AI Discussion - More of this!

- Have you had experiences where Responsible AI was promoted/expected?
- What are other examples you've encountered of Responsible AI?



# That's A Wrap!

Thanks for sticking with us over the last 11 weeks and a MASSIVE thank you to Chip Huyen, the TAs and the participants for making this possible

