Overview of Generative AI Stacks

Soumyadeep Mukherjee

@dementorSam

Why should you care about what I've to say?

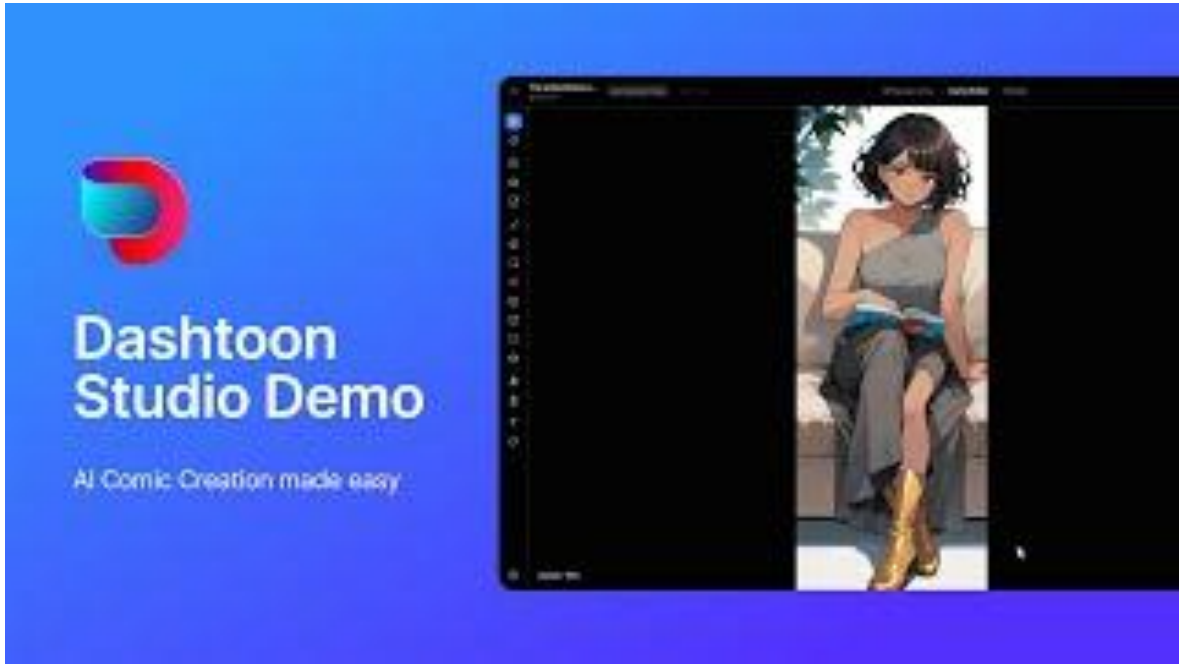
1. Founder and CTO at Dashtoon
   a. building generative ai first company since 2 years
   b. first consumer platform to monetise ai assisted content built in India for US.
   c. 0 to 1mn ARR in 6 months of launch.

2. Built & Scaled AI and Robotics for udaan.com in 2019-2022

Manage India's Largest GenAI Community with Regular meetups since February 2023

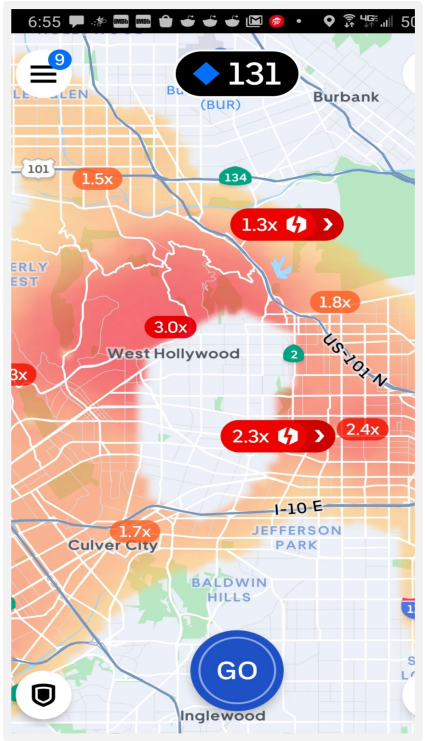
Ping me on https://twitter.com/dementorSam for a group invite.

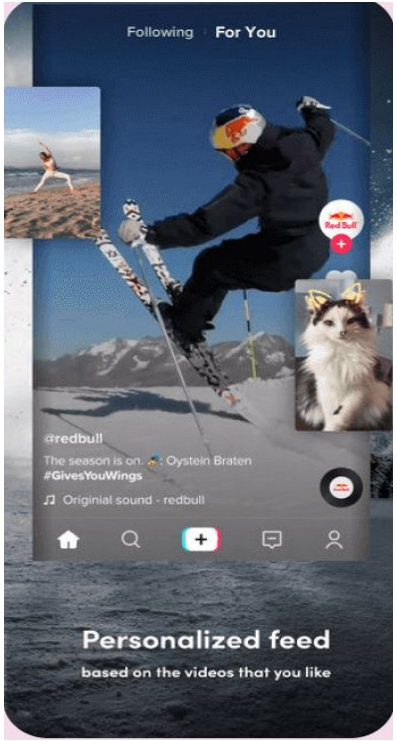
What is Dashtoon?



ML was already everywhere

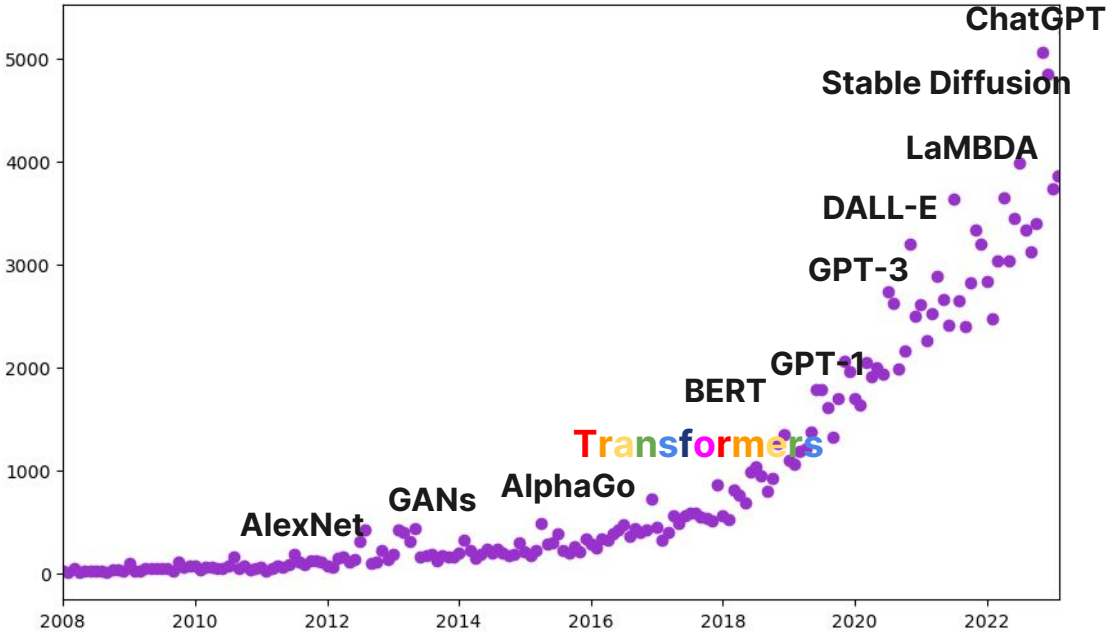
- Ad targeting at Google.
- Pricing at Uber lookalike
- Audiences on Facebook
- Feed at TikTok
- Collaborative filtering at Amazon
- Filters in Snapchat
- Image enhancement on camera
- Email spam detection
- Credit card fraud detection
- Content moderation
- Loan underwriting
- Lead scoring
- Churn prediction.





Accelerating AI research progress

ML+AI Papers on Arxiv Per Month



what made this possible?

2017 landmark Google Brain paper: “Attention is All You Need”

attention = next generation, scalable model architecture (helps decide what parts of the input are important)

transformers can do text, speech, vision, music, video too

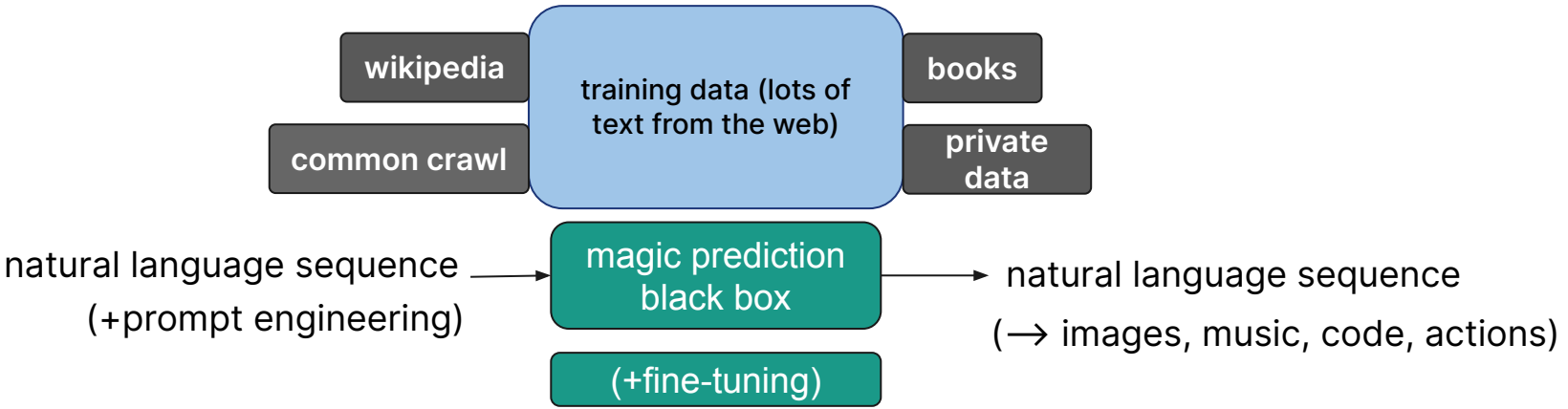
vaswani, parmar, shazeer et al

what is a large language model (LLM)?

model: input-output algorithm that learns (adjusts its internal variables) by "trial and error" from data

language: takes as input a window of text, generates new text by predicting the next word, repeatedly

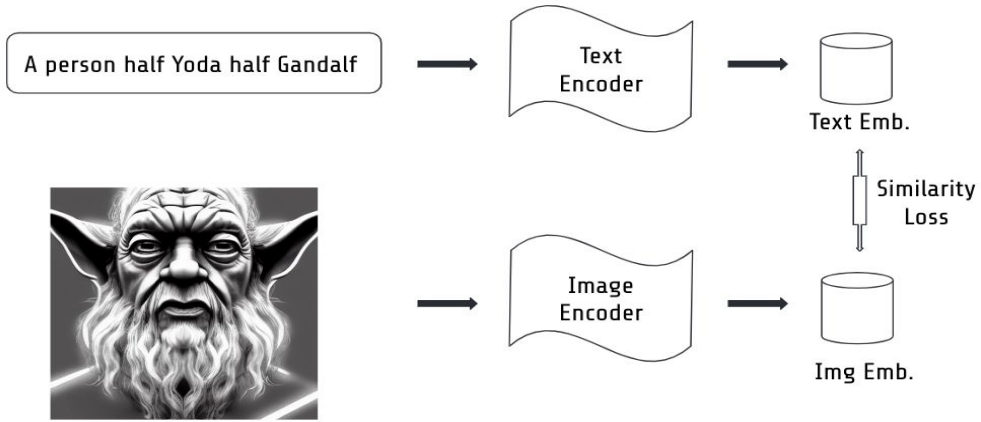
large: many variables, trained on a huge dataset of examples



what is a diffusion model (text2image)?

model: input-output algorithm that learns (adjusts its internal variables) by gradually refining predictions over iterations

diffusion: generates data by reversing a process of adding noise, progressively reducing noise to create high-quality outputs



What became possible?

text2image	inpaint	animatediff	RAG
lipsync	textTo3d	text2speech	img2img
image2text	text2music	segment	reference

Product Photography

Product Photography Workflow

Green indoor background with coral circular base

prompt

text2image

product photos

inpaint

final output

AI Anchor

AI Anchor Workflow

Indoor newsroom with breaking news written in the background.

background prompt

text2image

final output

"Welcome to the talk"

Anchor text

text2speech

Anchor Image

lipsync

inpaint

Comic Creation

Comic Panel Workflow

A girl angry at her boyfriend.

story

LLM

panel text

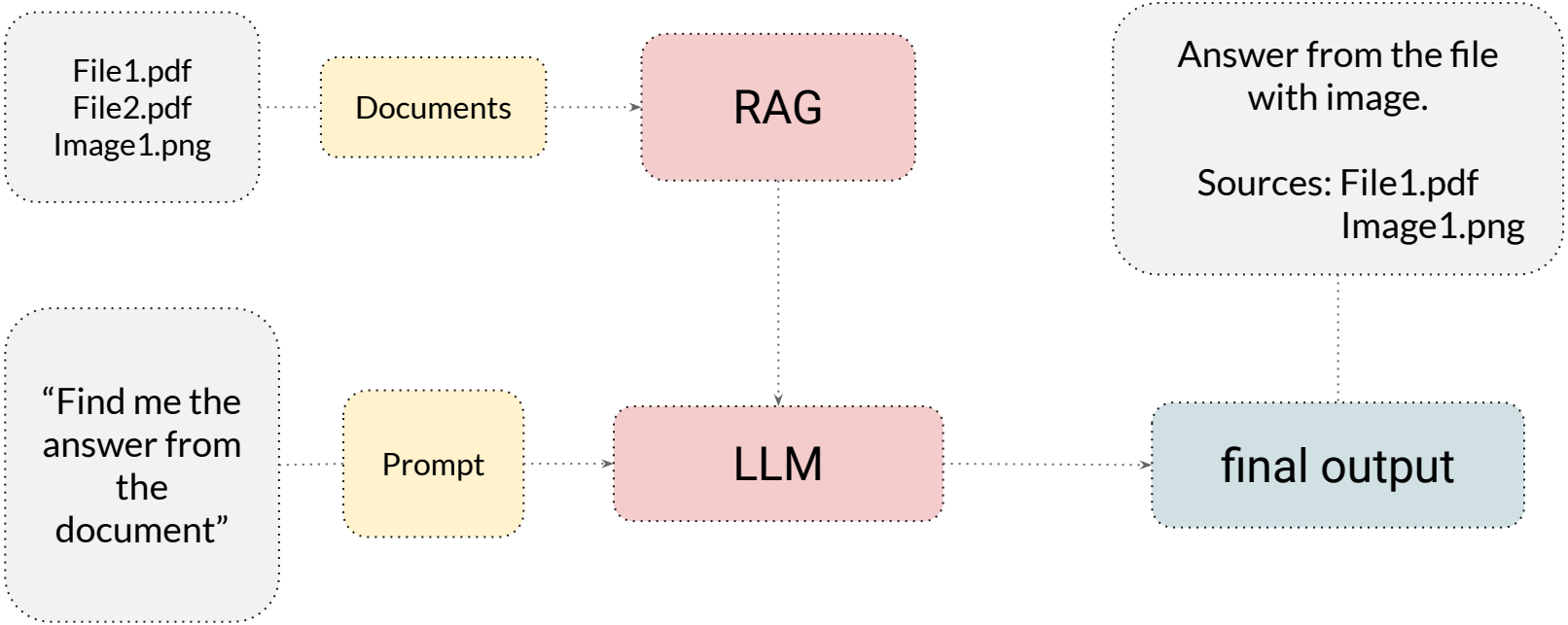
character

text2img

final output

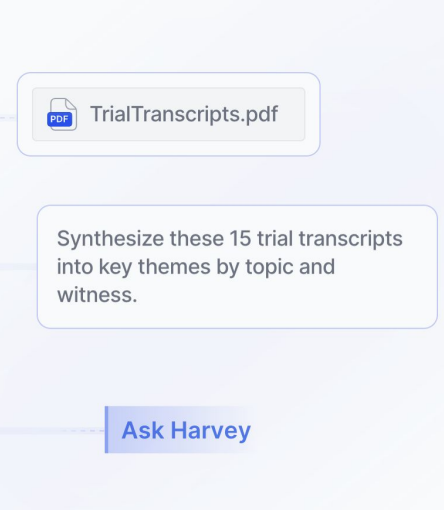
Q&A with Source

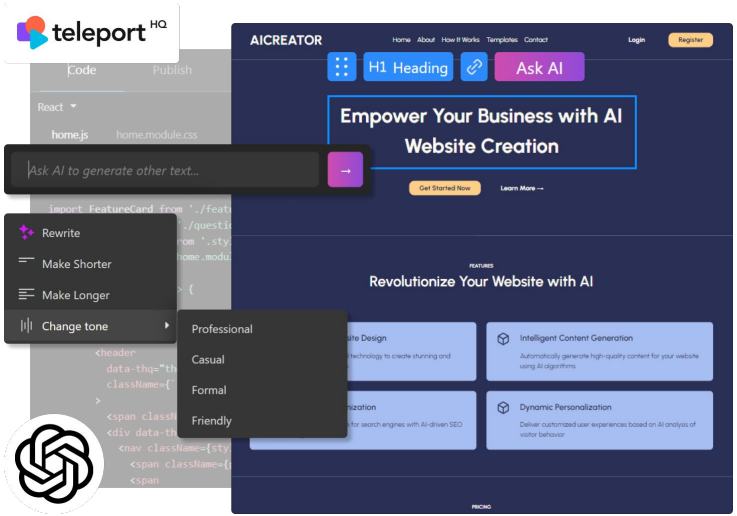
Q&A with source:
Retrieval-augmented generation (RAG)



Many More

Legal AI





Website Generation

Clothes Try-On

Challenges

- Lack of consistency
  - multiple attempts with same input can give varying results
- Controllability is far from non-generative workflows
- Hallucinations
  - Can create information out of thin air
- Bias
  - If data is biased, output will be biased
- Safety
  - Ethics
  - Privacy
  - misuse

questions

@dementorSam