

DATASCIENCESALON
NEW YORK

Operationalizing Gen AI

BuzzFeed Inc. 



Archi Mitra

Sr Director, Machine Learning
BuzzFeed Inc

Operationalizing Gen AI has led to a constant flow of creative (& successful) scaled content experiments

Jonah Peretti

Founder & CEO



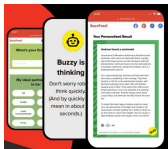
"...We will use AI to create content that wasn't possible before AI..."



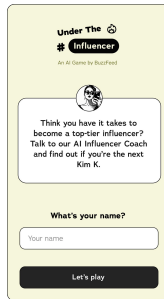
July 2021

Jan 2023

INFINITY QUIZZES



Feb 2023

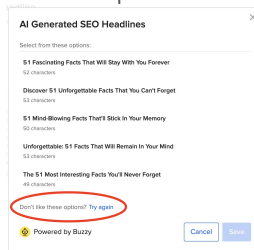


March 2023

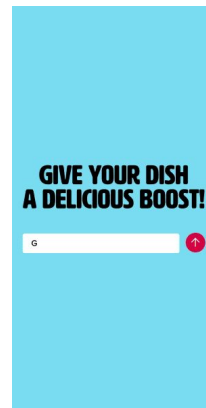
INFINITY QUIZZES



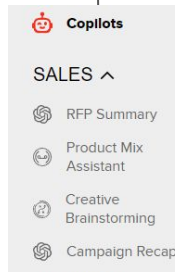
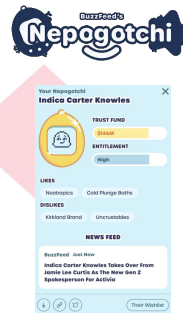
April 2023



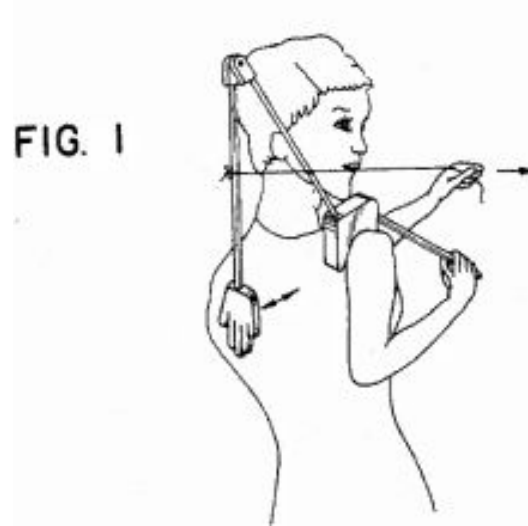
May 2023



June 2023



While it's important to acknowledge what operationalized Gen AI can look like, let's dive into the how



Agenda

- Timeline of Gen AI experiments
- Operationalizing Gen AI at scale -
 1. Democratize x Prototype
 2. Educate x Consult
 3. Scaled Experiment x Build
- Case Study: Botatouille - a domain-aware LLM

1

Empower creatives with reactive (& proactive) prototypes, "Playgrounds" & APIs

Access to playgrounds



Playground



Integrate into existing tools

AI Quiz Save and Close Quiz

Questions **AI Prompt** Preview ⚙️

AI Quiz Mechanics

This prompt will be sent to ChatGPT to generate content using inputs added by quiz takers.

Your Prompt

Use the format [[question_1]] to use the response to Question 1 in the prompt

Result Card

Customize the result card for your AI Quiz here.

Result Title

Example: Presenting [[question_1]]'s Oscar acceptance speech!

Use the format [[question_1]] to use the response to Question 1 in the result title

Powered by Buzzy

Reactive/Proactive prototypes



jin 11:13 AM

Hello! [Here](#) is the notebook you can use to chat with Google's PaLM2, the latest LLM family announced by Google last Wednesday at their I/O event. Feel free to play with it, and let me know if you have any questions. Enjoy! (Make sure you read the guidelines before running the notebook)

Democratize

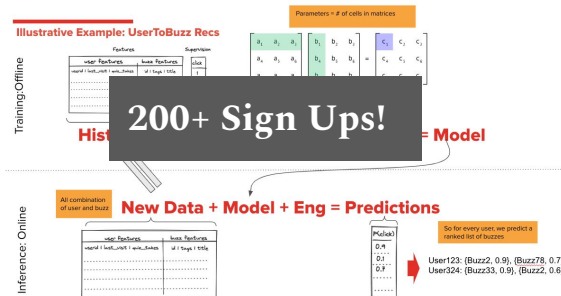
Empower every team to use any API

- Every app emits cost, token use, latency etc to a centralized analytics platform
- Org-wide Gen AI cost tracked in a centralized dashboard

2

Educate, engage & consult in Gen AI native content & tools ideas

Demystifying AI talks



AI Office Hours



Gen AI Workshops/Guidelines

- Led by super users of Gen AI within their respective content, business, sales & tech teams
- Internal guides on do's/don'ts collaboratively created by Tech + Content teams

#ai-help slack channel

3

Build, Buy & Integrate Gen AI into existing scaled tech stack with tight feedback loop with creators

Model alignment/
prompt testing path


Content/Editorial
teams



Product teams



ML teams
colab

Develop low effort way to iteratively align models by soliciting/incorporating bidirectional feedback from domain experts

**Scaled
Experiment**

Production build path

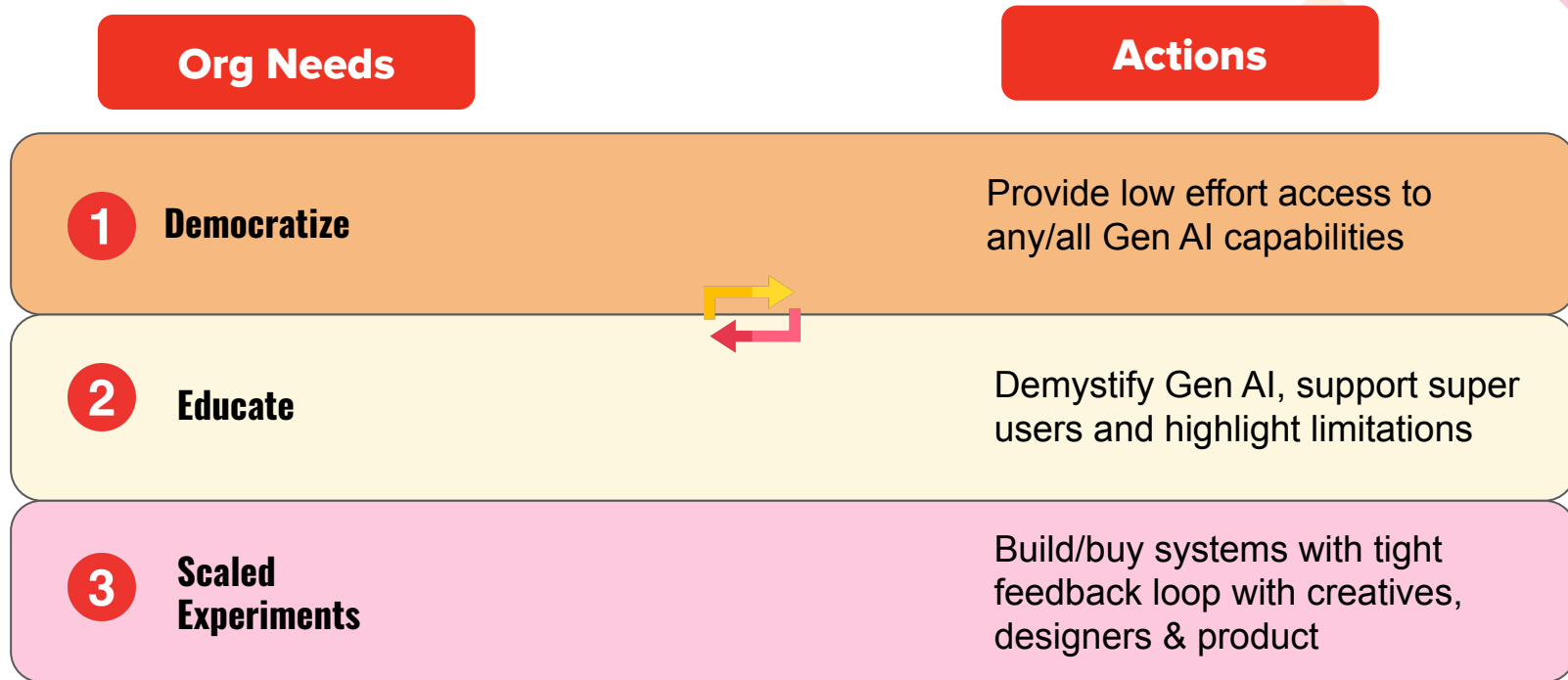
Develop flexible
system design

Incorporate
models/logic

Load test & emit
org-wide metrics

Red-teaming

3 easy hard (l)earned steps to operationalizing Gen AI



Botatouille: Operationalizing a domain-aware LLMs

1 Democratize



Got access to GPT, it recommends Tasty recipes!! But links are all wrong and recs are old.

LLMs can be domain-aware that draws on our Tasty RecSys. Let's put together a quick playground to test it.



colab ~2 weeks

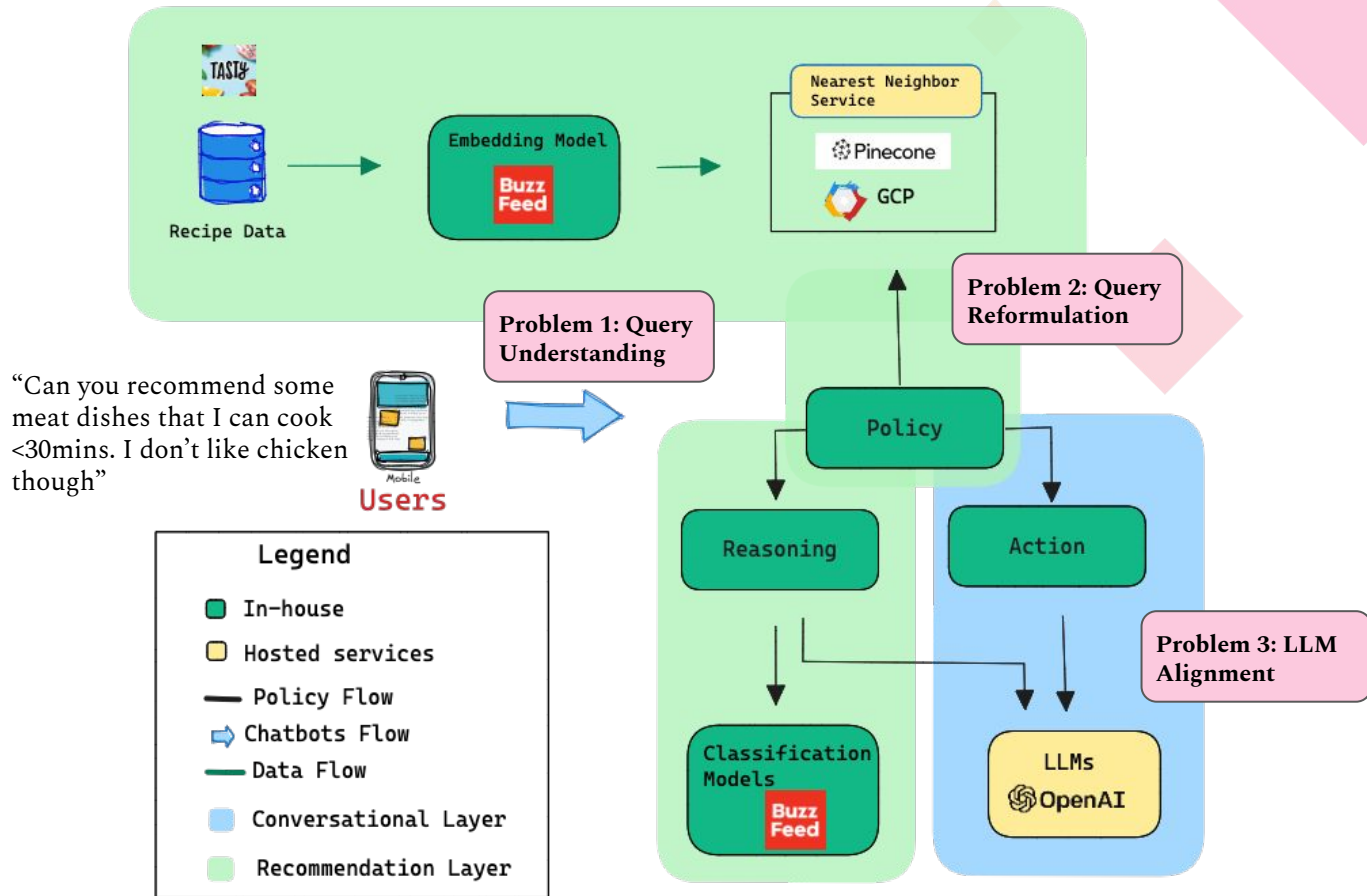
Playground works really well. This will be very compelling feature for Tasty users. It recommends a recipe for every prompt though

2 Educate

- LLMs' have a data recency problem
- LLMs' love to hallucinate
- **Reason & Act + Semantic Search** is framework that tries solve these two problems but introduces it's limitations
- Each reasoning (or thinking steps) increases latency and 3rd party LLM tokens -> the smarter you want it to be the more you have to "pay".
- Concept of red-teaming and model alignment

Botatouille: Existing ML problems masquerading as LLM problems?

3 Scaled Experiment



Thanks for joining. 