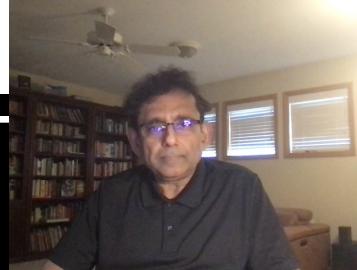


# CONVERSATIONAL COMMERCE FOR THE NEXT 500 MILLION

Ganapathy Krishnan, Ph.D  
VP Artificial Intelligence Center  
Seattle Washington

**Flipkart**



# INDIA & FLIPKART BACKGROUND

- India is diverse country and can be viewed as a “Many Indias”
  - High level of diversity in geography, languages, culture, literacy and income levels
  - One size fits all solutions do not work
  - Credit card customers are only top of the funnel
  - E-commerce market is expected to grow to \$200B by 2026
- Flipkart is building technologies to serve India's diverse customers
  - We have 350 million registered customers
  - We support eleven languages for speech recognition and translation
  - Invested in a tech enabled supply chain to support hundreds of thousands of sellers
  - Highly trusted with Cash on Delivery Offerings



# Voice & Vernac | Increasingly vital to attract and retain customers



Market

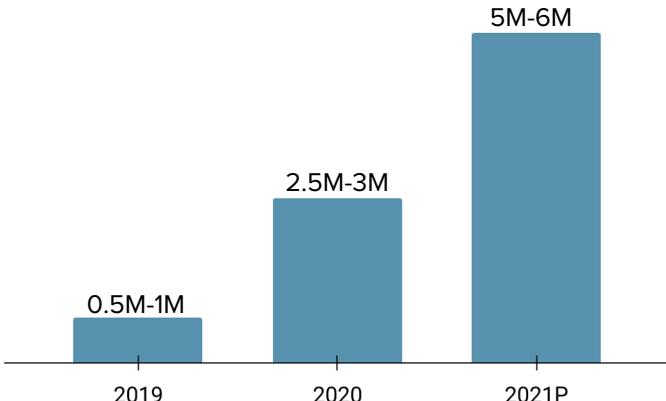
Flipkart

## Voice



**1 in 10** users have tried voice search

MAU for voice assistant apps has increased significantly over the past months



## Vernac



**1 in 3** new e-retail shoppers use vernac interface



50%

**1 in 10** visits via vernac interface

Growth in # of times webpages are translated to vernac language

84%

Online video view vernacular content



# CURRENT ECOMMERCE PARADIGM

Search

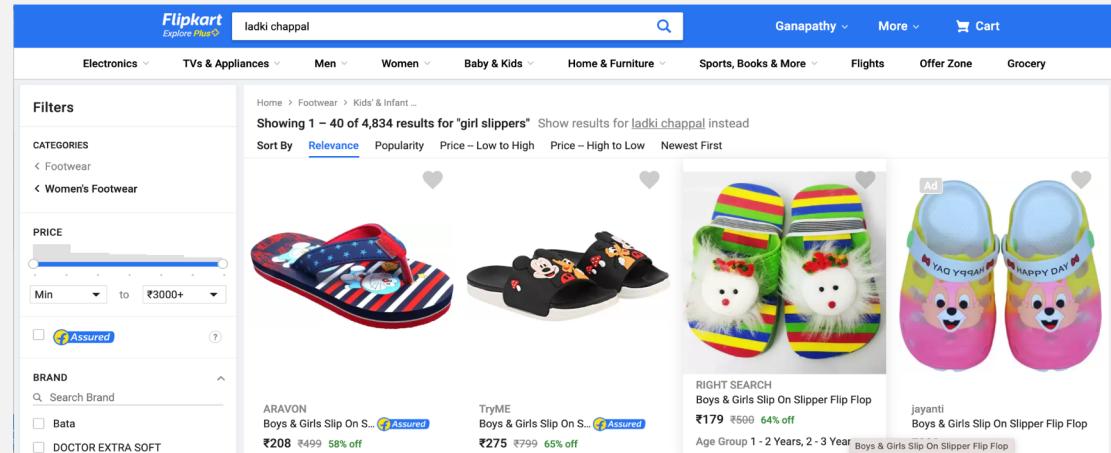
Filter

Product Descriptions

Reviews/FAQ

Price compare

Buy



SIMPLIFYING THE  
SHOPPING JOURNEY WITH  
CONVERSATIONAL  
COMMERCE FOR THE  
NEXT 500 MILLION  
CUSTOMERS



## Search & Discovery

Find the product you want



## Decision

Make sure it meets your needs!



## Transact

Add to cart and buy product

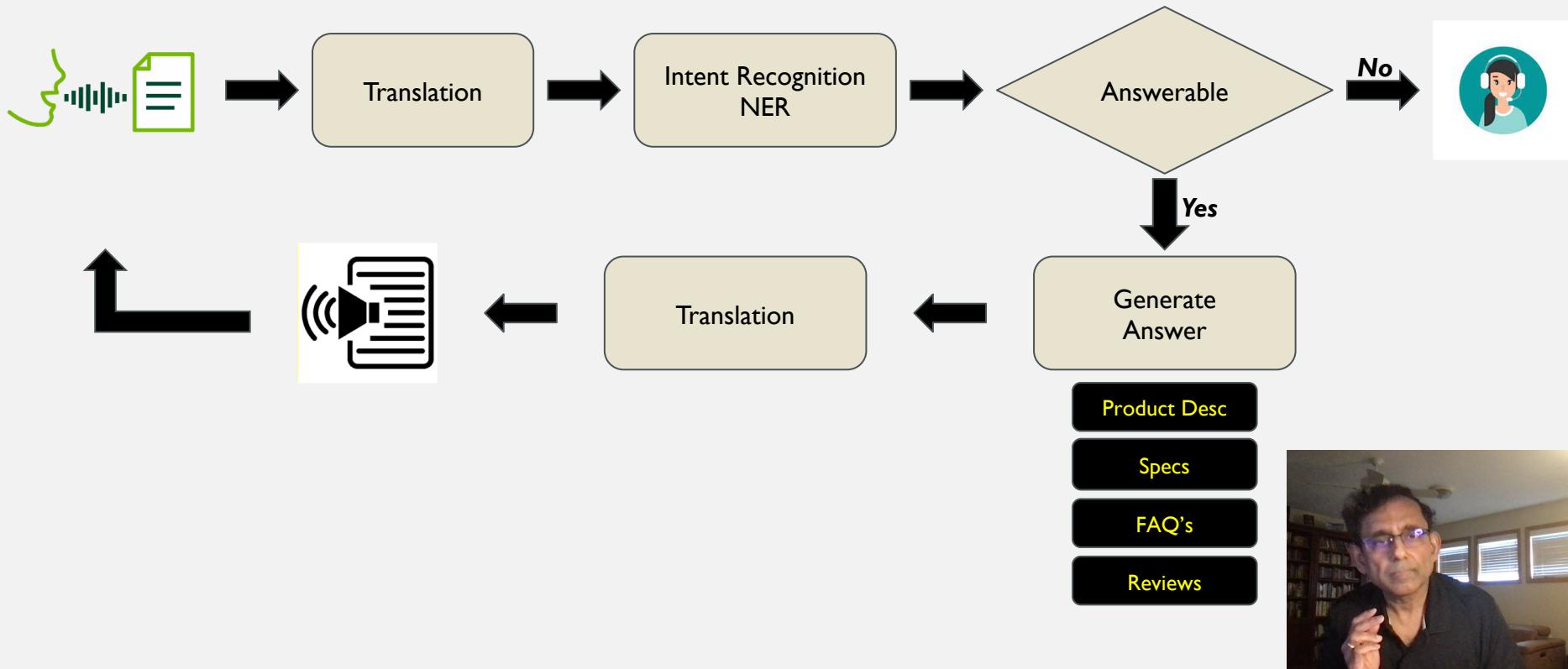


## Customer Service

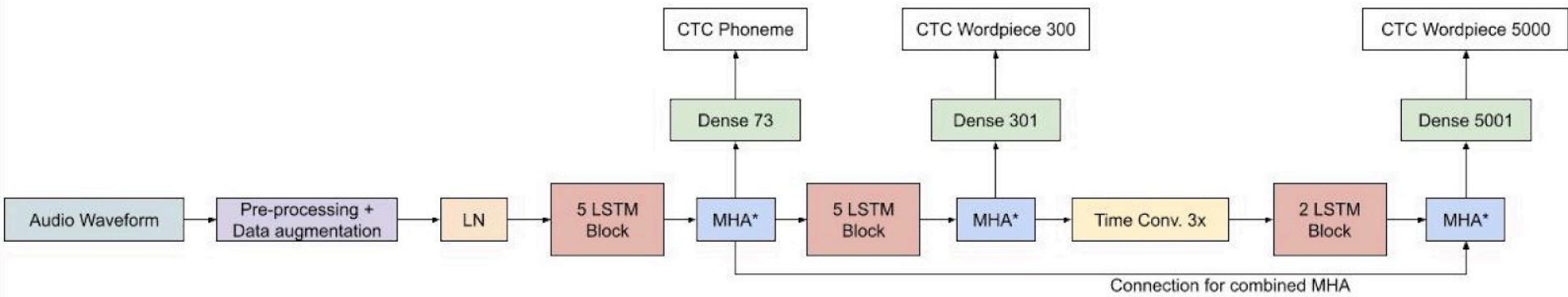
Product delivery  
Exchange  
Returns



# AI COMPONENTS FOR THE DECISION ASSISTANT



# AUTOMATIC SPEECH RECOGNITION



## WHERE IS THE DA AVAILABLE?

- DA is available on our product pages to answer any product related query
  - Rich product descriptions
  - Specifications
  - User generated content
    - Frequently asked questions
    - Product Reviews

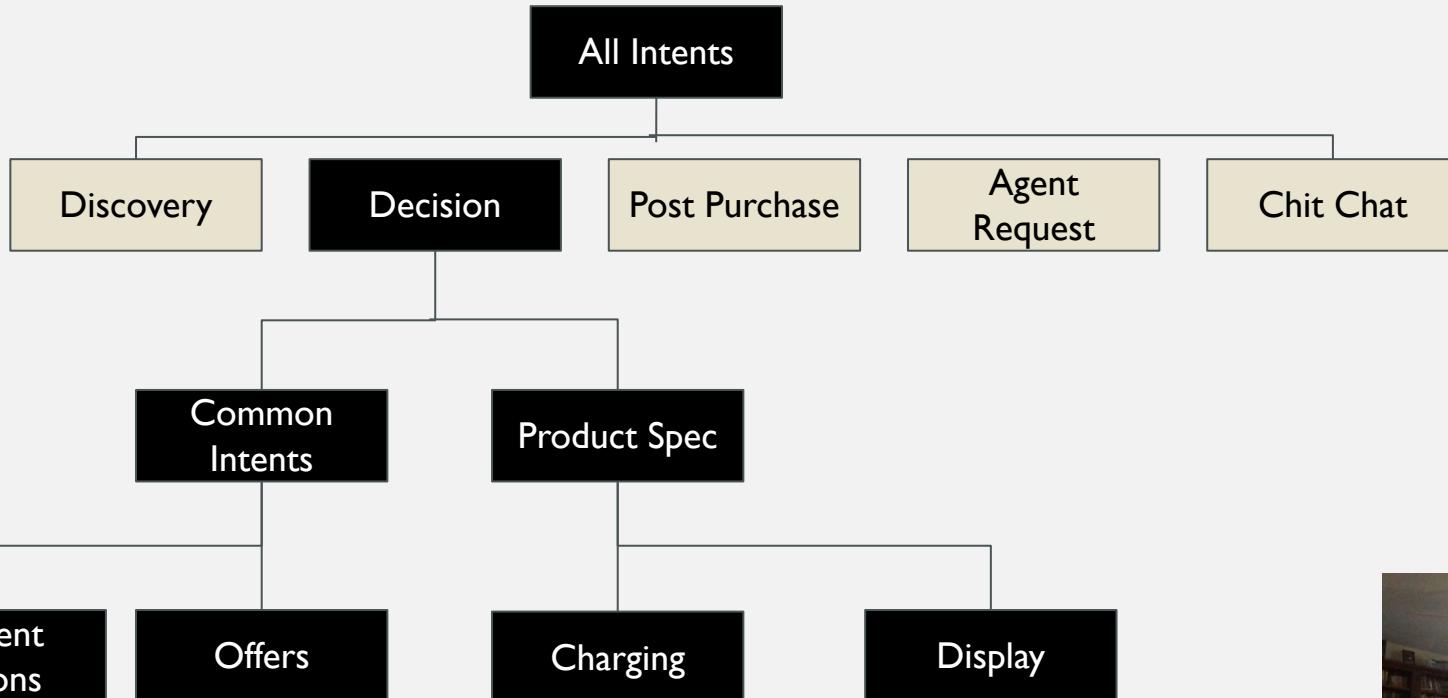


# WHAT DO THE QUESTIONS LOOK LIKE?

- **Q:** "How long can the phone last?"
  - **A:** "6000 mAh battery that offers a standby time of up to 57 days."
- 
- **Q:** "What is the resolution of the screen?"
  - **A:** "1520x720 Pixels".



# INTENT HIERARCHY



# FINE GRAINED NAMED ENTITY RECOGNITION

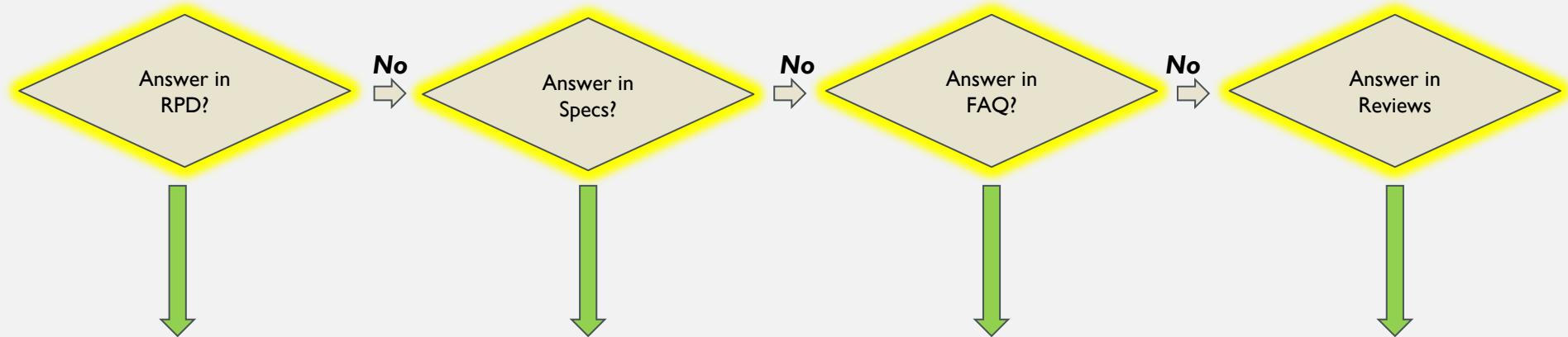
- NER is used to populate API's so we can answer the user
  - Is this shirt available in **blue?** (**Entity is color - value is blue**)
  - Can I get the dress in **PS?** (**Entity is size - value is PS**)
  - Do you have any rebates on the **red [color]** iphone **X [model]**
- NER is also used to mask queries to help ML models generalize better



# HIGH LEVEL SYSTEM FLOW



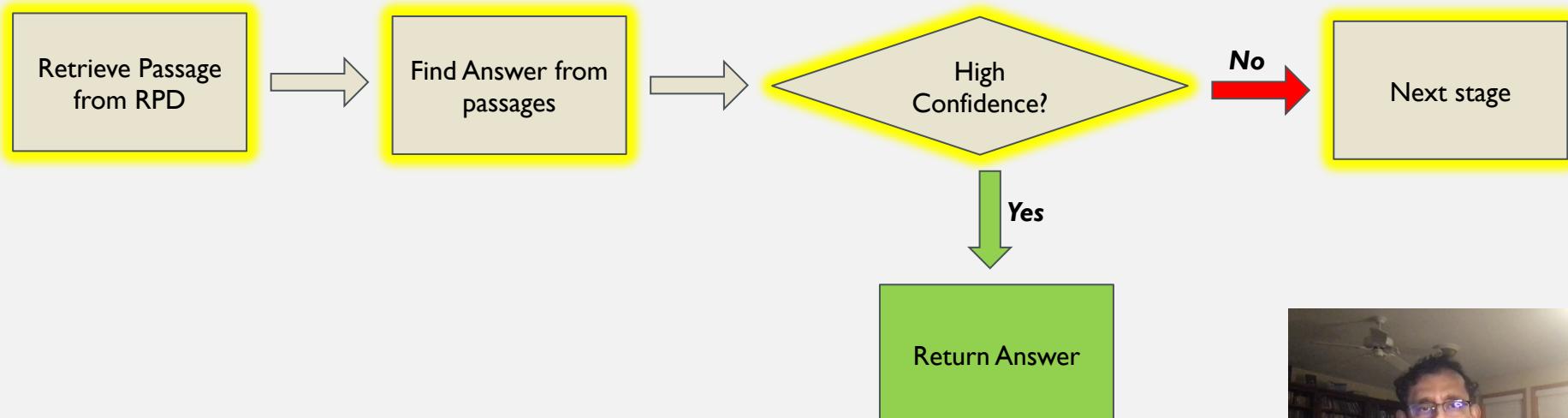
No



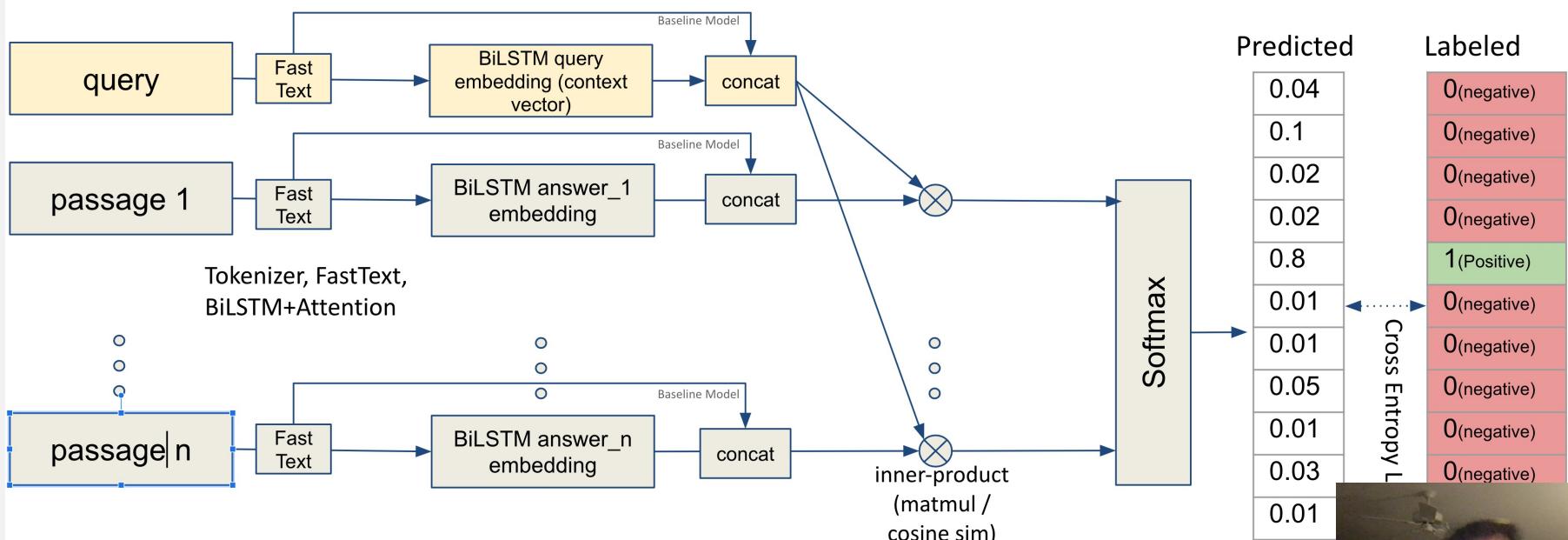
Answer found



# SEARCH FOR ANSWER IN RICH PRODUCT DESCRIPTIONS



# TRAINING THE MODEL FOR PARAGRAPH RETRIEVAL



$$p = \text{softmax}(q \cdot AT)$$



# RPD BATCH TRAINING

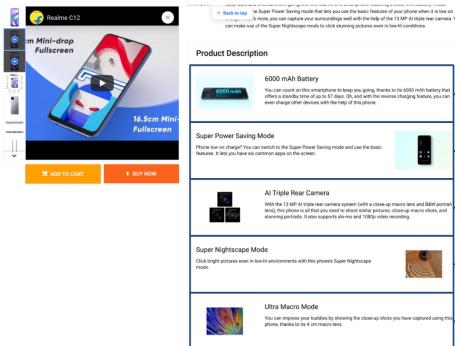
question	passage	intent
Does it have fingerprint unlock?	Unlocking the Redmi Y2 is as easy as it gets. You can either use the fingerprint sensor or just use the Face Unlock feature.	decision__product_spec__unlock__finger_print_unlock
what does amoled display offers in k20	Whether you're watching videos or playing games, enjoy a stunning visual experience with the Redmi K20 which comes with a 16.23-cm (6.39) AMOLED Display, 91.9% screen-to-body ratio, HDR range and a 100 DCI-P3 colour gamut.	decision__product_spec__display__display_type
What smart features does the Nokia smartphone have?	This Nokia smartphone comes with multiple smart features, including Google Assistant, Digital Wellbeing, Adaptive Battery, App Actions, Google Lens, and Visual Snapshot. The Google Assistant effectively simplifies your life	decision__product_spec__processor__processor
how many mAh battery it has?	Its 4230 mAh battery offers up to 22 hours of browsing, up to 14 hours of playing videos on YouTube, and up to 35 hours of usage when the screen is on standby	decision__product_spec__battery__battery_capacity
what is the rear camera resolution?	It has a high-definition 8 MP rear camera with a powerful LED flashlight that lets you take clear images even in low-light conditions.	decision__product_spec__camera__camera_primary_quality
How much ROM does this phone have?	Powered by an up to 1.65 GHz octa-core processor, this smartphone ensures easy multitasking. Also, it has 3 GB of RAM and 32 GB of ROM to ensure seamless performance and to store images, videos, and much more.	decision__product_spec__memory__memory

↔ ↔ Positive Question-Answer Pair

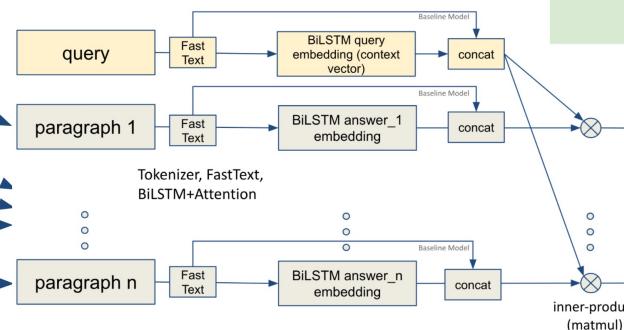
↔ → Negative Question-Answer Pair



# INFERENCE



"What kind of battery does it have?"

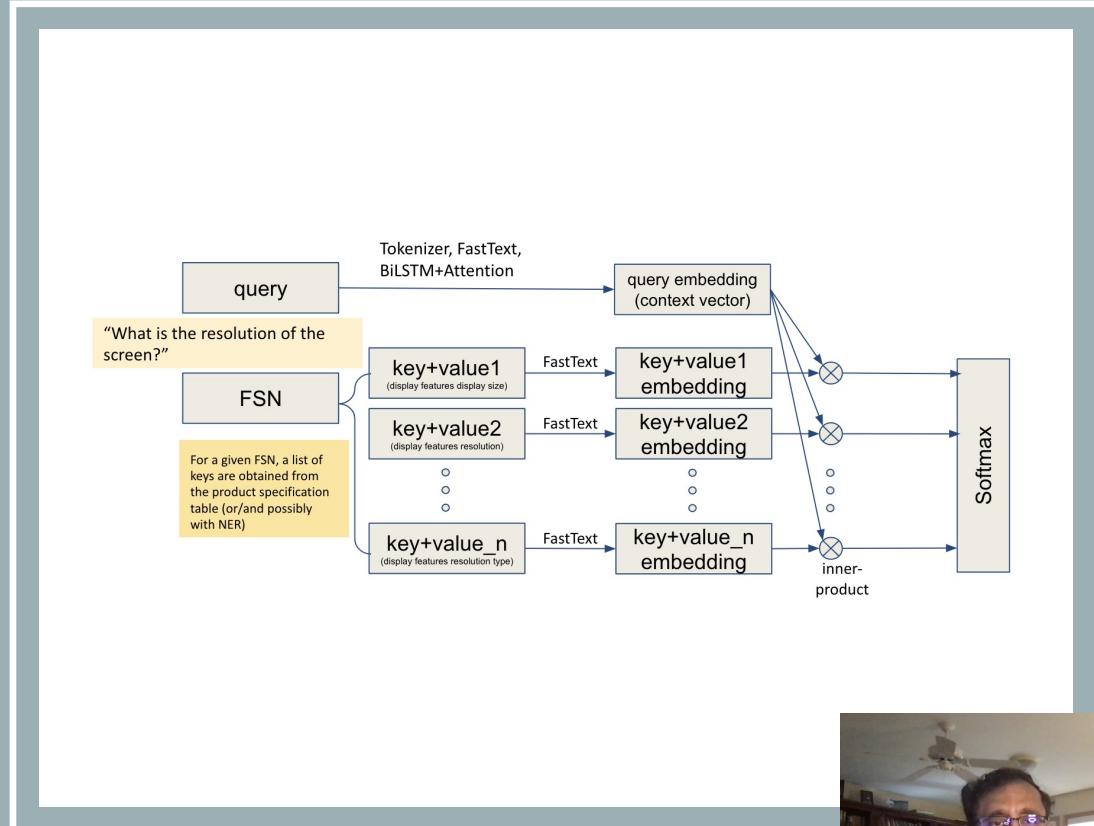


The top-k answer candidates are selected and passed to the RPD QnA model.

Similarity Scores  
0.8  
•  
•  
•  
•  
0.1



# KEY VALUE RETRIEVAL

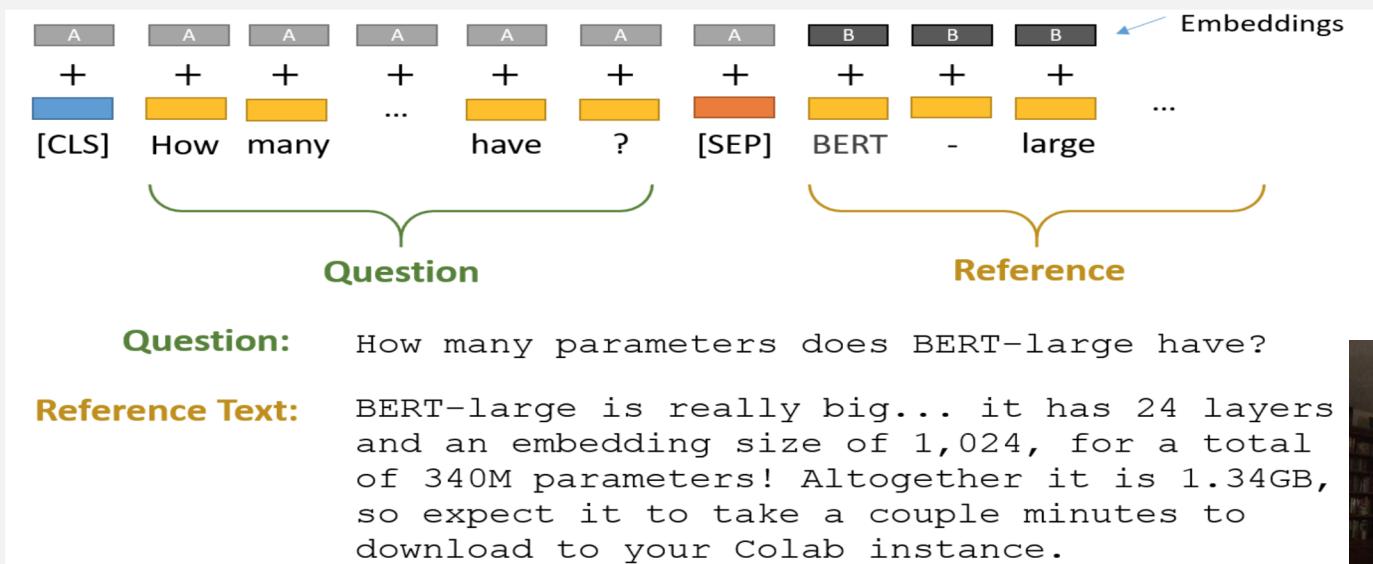


# QnA (Question And Answer)

Our QnA implementation  
is extractive.

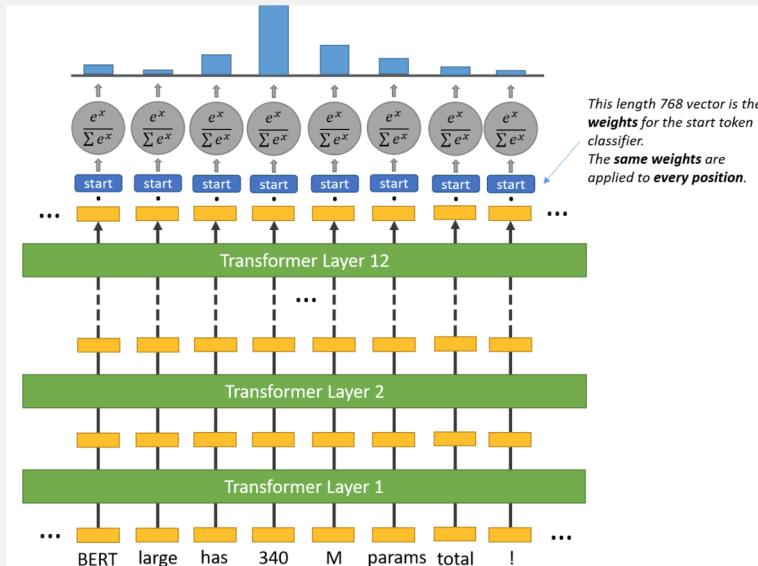
Question And Answer task  
is basically mapping:

- a sequence of text (query)
- to a sequence of text (answer)
- in a paragraph where the answer  
is thought to be in (Context)

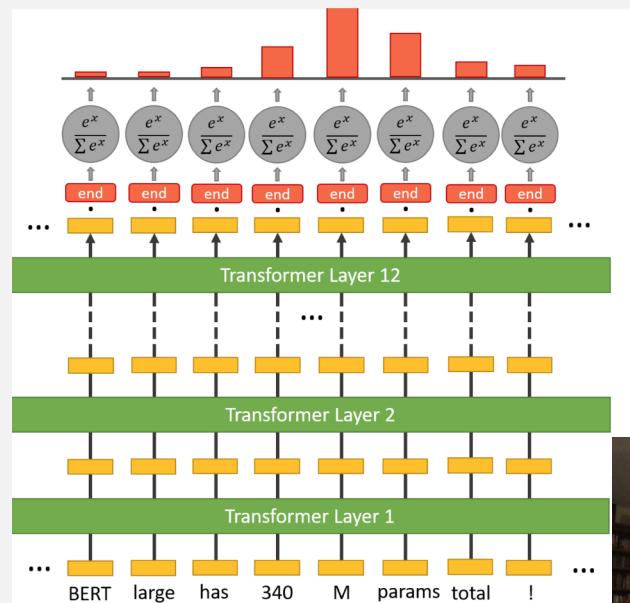


# QnA (Question And Answer)

Each token in the text is scored for the Start of an Answer and End of an answer



We take the maximum scoring token for the Start and End and reconstruct the result between start and End tokens inclusively.



Pretrained Base Model

Train QnA Task on  
SQuAD V2 (English)

Fine tune on RPD  
Labeled Data

Fine tune on KV  
Labeled Data

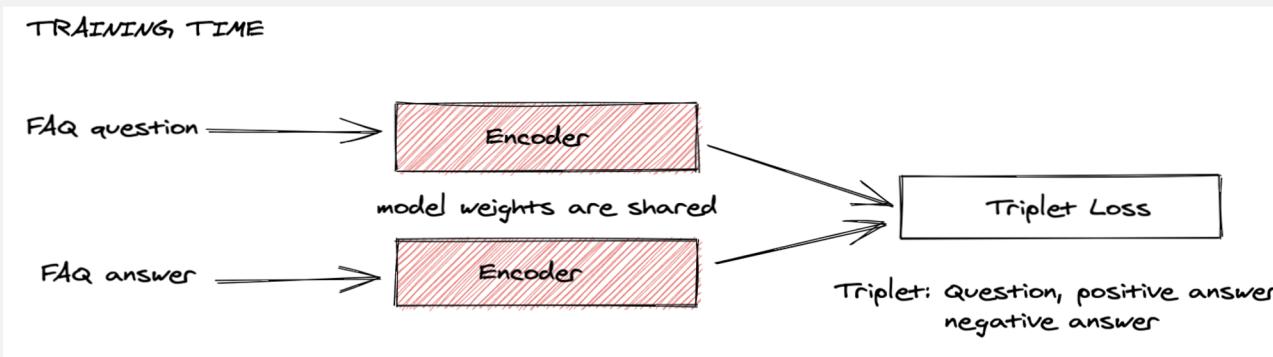
## OUR IMPLEMENTATION DETAILS

- The model architecture was chosen to be ALBERT - A Lite BERT by Google (2019) as a start
- Why ALBERT?
  - Much fewer parameters than BERT
  - Better tokenization algorithm (SentencePiece instead of WordPiece)
  - No NSP -> Better Task: Sentence Order (SO)
  - No Dropout -> Data Augmentation provided better results
  - And does exceptionally well in QnA Tasks as a single model
- We used corpuses that have unanswerable questions given the context.
- Our model is thus trained to return “**Can’t Answer**” (**Null**) if the query cannot be answered from the context it is paired with.



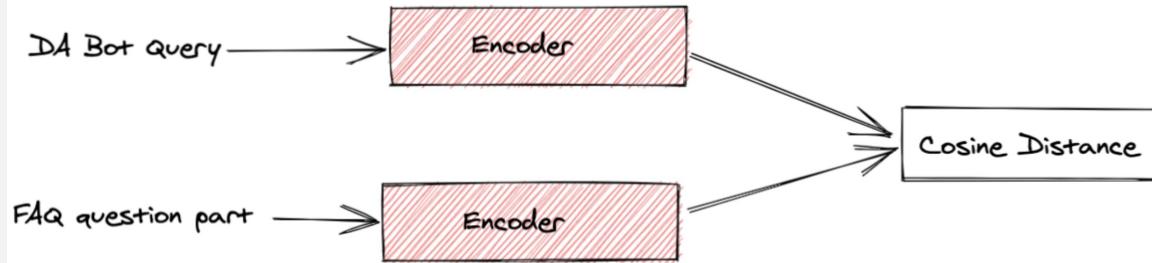
# LEVERAGING UGC

- Encoder converts text into numeric embeddings
- **Encoder used:** [sentence-transformers/stsb-distilbert-base](#)
  - Siamese distilBERT based encoder with shared weights (single model)
  - Fast & performant on sentence data
- **Training data for fine tuning:**
  - UGC QnA (FAQ) Question & Answer pairs
  - Top 200 verticals by frequency of questions

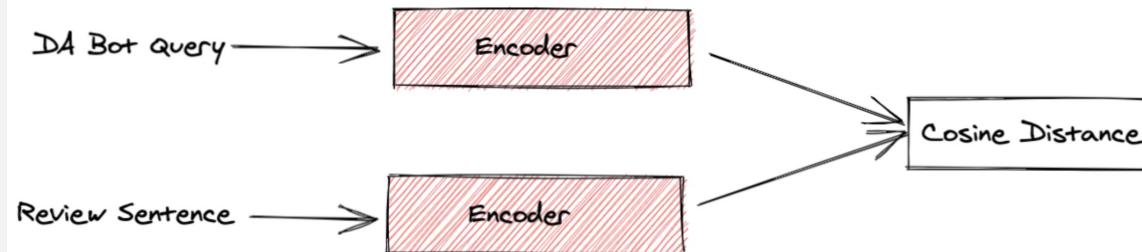


# UGC MODELS IN PRODUCTION

INFERENCE TIME: FAQ



INFERENCE TIME: Reviews



## Downstream Task: Semantic Textual Similarity

- FAQ: Comparing similarity between user query and FAQ Question part
- Reviews: Comparing similarity between user query and Review Sentence



# USE CASES AND AI TECHNOLOGIES

Word Embeddings (Fasttext) -  
Used as inputs, represent  
words as vectors.



Sequence Models (BiLSTMS)  
for low latency applications  
such as Intent Recognition.

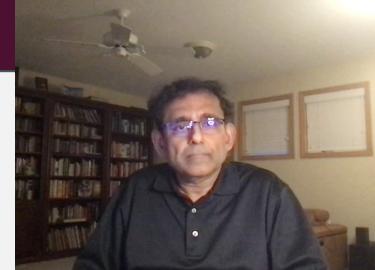
Sequence Models (Bi-LSTMS +  
CRF) for Named Entity  
Recognition

Sentence Embeddings (BERT  
based Siamese Networks) -  
Used as inputs and represent  
the query as vectors.



Transformer based models for  
complex tasks like span based  
answer extraction (Albert)

Generative transformer  
models for inference and  
answer generation (e.g. T5)

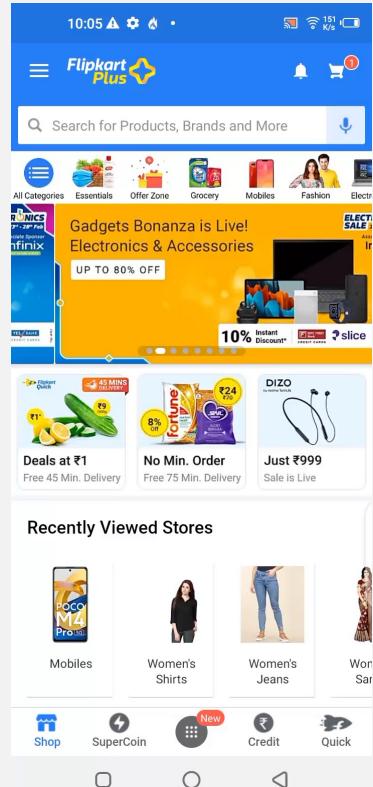


# METRICS

- Models
  - Precision & Recall
  - Word Error Rate
  - BLEU (translation)
- Overall System
  - Bounce Rate
  - Latency
  - Answerability
  - Session Success Rate
  - Transfer to agent
  - CSAT Ratings
- Business
  - Conversion
  - Repeat Customer
  - NU -> NC

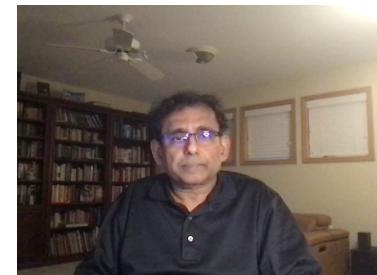


# DEMO BY ANANTH V – SR. ARCHITECT



# LEARNINGS

- Whenever possible, use lightweight models.
- A good MLOps infrastructure saves you a lot of time and lets you experiment more frequently
- It is really important to have the right tools for labelling and ensuring that the data quality is high
- Active learning helps reduces labelling cost
- Synthetic data generation also helps reduce labelling cost
- It is really important to have great analytics so you can tune and fix your models



## SOME NEW AREAS WE ARE WORKING ON

- Incorporating high quality chit chat
- Zero shot or few shot learning
- Incorporating Video reviews
- Integration between Decision, Discovery, Post Purchase, Chit chat
- Abstract summarization
- Generative models for answers



THANK YOU  
FOR  
LISTENING

Reference this talk and connect on LinkedIn  
<https://www.linkedin.com/in/ganapathykrishnan>

