

# CBA challenges: Use case of applying Gen AI to Customer Support Tickets

*Lausanne, 22/10/2024*  
*Huynh Thanh Trung*

# Task I – Dataset Insights

- Customer support tickets at electronics store:
  - The dataset contains 8469 tickets, including:
    - Customer Information:
      - Customer Name, Customer Email, Customer Age, Customer Gender
    - Purchased Product:
      - Product Purchased, Date of Purchase
    - Ticket Metadata:
      - Ticket ID, Ticket Type, Ticket Subject, Ticket Description, Ticket Status, Ticket Priority, Ticket Channel
    - Ticket Resolution:
      - Resolution, First Response Time, Time to Resolution, Customer Satisfaction Rating

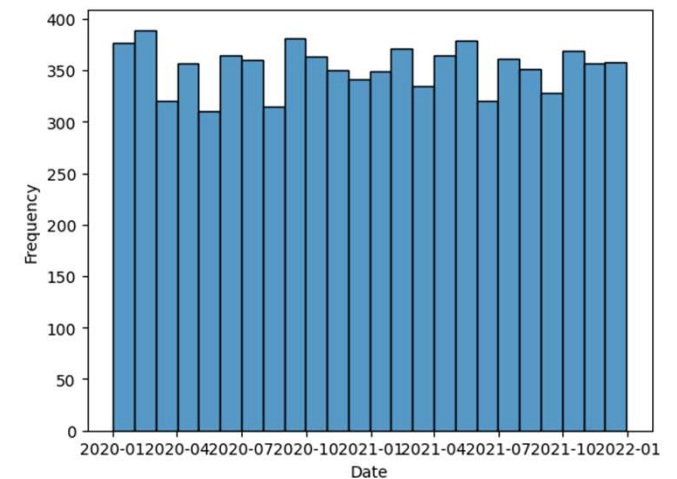
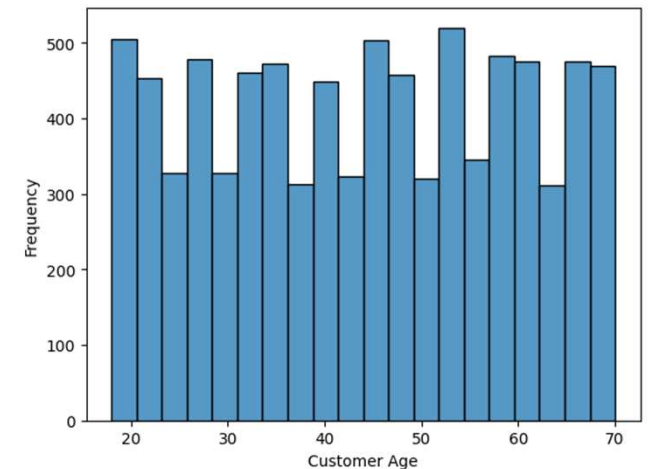
# Task I – Dataset Insights

- Customers demographics:

- Age: 20-70, equally distributed
  - Younger (20-30), Middle (30-45), Elder (45-70)
- Gender: Male (2896), Female (2887), Other (2686)

- Purchased Products:

- The products are electronic devices:
  - Entertainment (e.g., gaming consoles, cameras), work (e.g., laptops, office software), and household purposes (e.g., TVs, vacuum cleaners).
- Purchased from Jan 2020 to Dec 2021

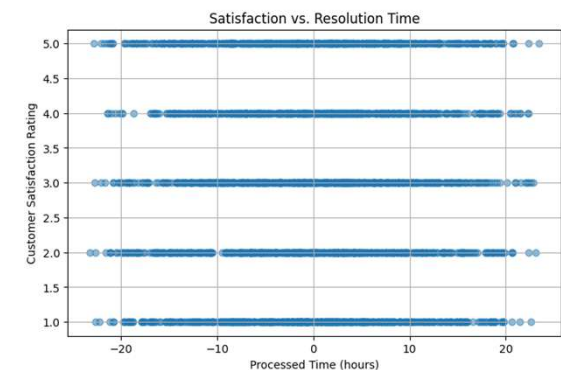
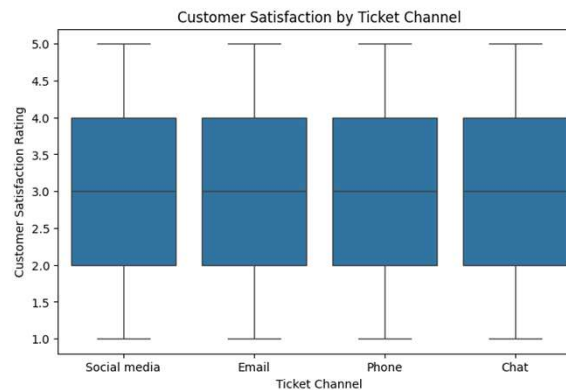
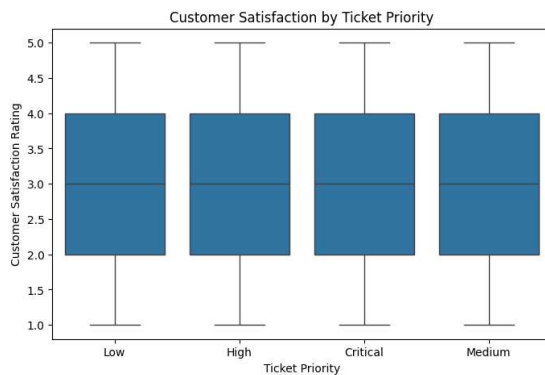


# Task I – Dataset Insights

- Ticket Metadata:
  - Ticket Type:
    - Refund Request, Technical Issue, Cancellation Request, Product Inquiry, and Billing Inquiry
  - Ticket Subject:
    - 16 different subjects (e.g. Refund request, Software bug, Product compatibility)
    - Each subject appears in various data types
  - Ticket Description:
    - The content is brief but lacks sufficient detail if viewed standalone.
      - Ticket subject and ticket type are needed to determine user intent.
    - May contains duplications:
      - E.g. *I'm having an issue with the {product\_purchased}. Please assist.*
    - Some descriptions contain html tags, code injections
  - Ticket Priority:
    - Critical, High, Medium and Low, distributed equally
  - Ticket Channels:
    - Chat, Email, Phone, and Social media, distributed equally

# Task I – Dataset Insights

- Ticket Resolution:
  - Resolution:
    - Provided for closed tickets, but the entries in the dataset lack meaningful content.
  - Customer Satisfaction Rating:
    - Ordinal data: 1-5, equally distributed
    - No significant difference in rating distribution is observed across customers of different age groups, genders, ticket priorities, or channels.
  - First time response vs Time to Resolution
    - The Time to Resolution can occur either before or after the First Response Time
    - The interval between them does not significantly impact customer rating.



# Task I – Ticket Descriptions and Subjects

- From "Ticket Subject" and "Ticket Description" columns, how would you identify common queries from customers?
  - Preprocess the Ticket Description
    - Remove duplicate descriptions, extra spaces, non-informative phrases
  - Use embedding to transfer the sentence into latent representation spaces
    - Sentence-T5
  - Calculate the centroid for each Ticket Subject clusters
  - Select top-k sentences whose embeddings are the closest to the centroid

Typical tickets for Product inquiry

```
- i'm working on this and am quite frustrated with the state of it. please be as helpful as possible. (1) remove that {product_purchased} i'm not sure if this issue is specific to my device or if others have reported similar problems.
- please contact support(at)least(as soon as possible) if it is urgent. if you have any problems please contact me. purchasing i've noticed that the issue occurs consistently when i use a specific feature or application on my {product_purchased}.
- i'm facing a problem with my {product_purchased}. the {product_purchased} is not turning on. it was working fine until yesterday, but now it doesn't respond. after sending something out a certain way, but i've reviewed the troubleshooting steps on the official support website, but they didn't resolve the problem.
- i have an issue with the {product_purchased}. how do my credit cards work? how do my credit cards work? this problem started occurring after the recent software update. i haven't made any other changes to the device.
- if you know of other problems with your purchase contact me by email and i will try to fix the issue before sending it back. this is an older version i've checked for any available software updates for my {product_purchased}, but there are none.
```

# Task I – Ticket Descriptions and Subjects

- Can you explain how you would use the "Ticket Type" field to categorize customer inquiries into broad categories or high-level categories?
  - Ticket Type:
    - Refund Request, Technical Issue, Cancellation Request, Product Inquiry, and Billing Inquiry
  - Tickets of each ticket type contains all 16 ticket subjects
    - It is impossible to build a hierarchy
  - Proposal: Define 2 high-level categories (based on semantic)
    - “Service Help”: [Technical issue, Product inquiry, Billing inquiry]
    - “Cancellation and Refund”: [Refund request, Cancellation request]
  - The proposed high-level categories matches the clustering result on Ticket Description semantics

# Task I – Ticket Descriptions and Subjects

- Given the "Customer Satisfaction Rating" in our dataset, how can we pre-processing this feature and how can we potentially use this feature in the models you are developing in Part II?
  - The "Customer Satisfaction Rating" is an ordinal variable ranging from 1 to 5
  - Only available for closed tickets.
  - Since it is already numerical and reflects an inherent order, we can retain the original values without transformation.
  - In Part II, this feature can be used to explore patterns
    - Which customer segments tend to give higher ratings
    - How factors such as resolution style, time to resolution, first response time influence customer satisfaction.
    - It may also serve as a target variable for predictive modeling.



# Task II – User Intents

- Design an AI or GenAI solution to categorize customer inquiries into different intent categories.
  - Since user intents are not available, first I define a list of predefined user intents based on the dataset.
  - Then, I model the categorization problem as Natural Language Inference (NLI) task:
    - Determine if the hypothesis that the ticket content relates to the user intent
    - Select the intent with highest score on Entailment label
    - Implement by using zero-shot classification on facebook-bart-mnli

Ticket Type: Refund request, Ticket Subject: Display issue, Ticket Description: I've recently set up my {product\_purchased}, but it fails to connect to any available networks. What steps should I take to troubleshoot this issue? If you're the user that's using Microsoft Windows Media Encoder I've tried different settings and configurations on my {product\_purchased}, but the issue persists.

User intent: Request Refund

Ticket Type: Technical issue, Ticket Subject: Display issue, Ticket Description: There were three or four people in the room, and they all took the keys, and he was completely quiet... it looked like in the very background. I've tried troubleshooting steps mentioned in the user manual, but the issue persists.

User intent: Fix Display Issue

Ticket Type: Cancellation request, Ticket Subject: Cancellation request, Ticket Description: I've accidentally deleted important data from my {product\_purchased}. Is there any way to recover the deleted files? I need them urgently. I've accidentally deleted important data from my {product\_purchased}. Is there any I've performed a factory reset on my {product\_purchased}, hoping it would resolve the problem, but it didn't help.

User intent: Request Cancellation

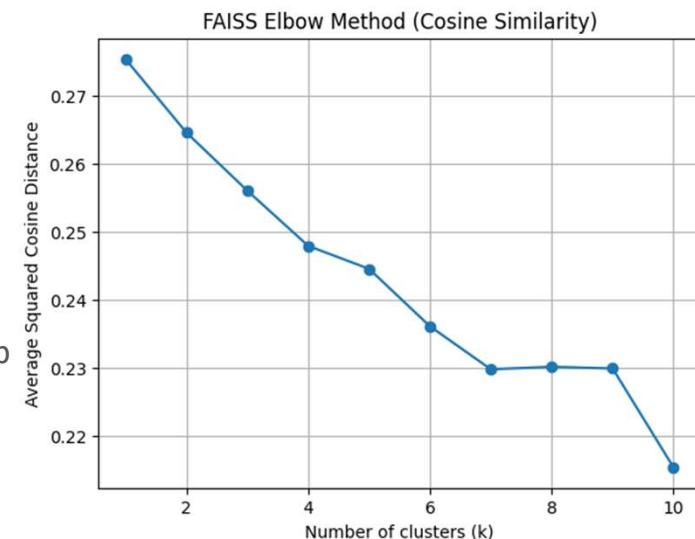
```
user_intents = [  
    "Help with Product Setup",  
    "Check Peripheral Compatibility",  
    "Resolve Network Problem",  
    "Recover Account Access",  
    "Report Data Loss",  
    "Fix Payment Issue",  
    "Request Refund",  
    "Battery Draining Too Fast",  
    "Need Installation Support",  
    "Report Software Bug",  
    "Report Hardware Issue",  
    "Ask for Product Recommendation",  
    "Report Delivery Problem",  
    "Fix Display Issue",  
    "Request Cancellation",  
    "Check Product Compatibility"  
]
```

# Task II – User Intents

- How would you design an unsupervised learning model or generative AI model to discover new categories of customer inquiries, given that we don't have a clear categories list?
  - Understand Inquiries: Use ticket content to capture the semantics of user inquiries.
  - Semantic Clustering: Embed the ticket content and apply clustering to group similar inquiries.
  - Label Generation: Determine a representative title for each cluster by analyzing the top-k most typical tickets in the group.

## Process:

- Preprocess the ticket content
  - Preprocess and merge type, subject, description
- Forward ticket content to pretrained embedding model
  - Use Sentence-T5 for embedding
- Use K-mean clustering on the embeddings
  - Elbow method to determine the optimized number of K
- Analyze the clusters with top-k typical tickets from each group
  - Use Gen AI/human intuition to label the intent for each ticket group



## Task II – User Intents

- Given the "Customer Age" and "Customer Gender" fields, how might you personalize the AI model's responses to different customer segments?
  - Categorize the customers into 3 customer age group: Younger, Middle, Elder
  - Propose to use the rule-based as follows:
    - Younger: use casual tone
    - Middle: use balanced tone; with empathic tone for woman, efficient tone for men and neutral tone for other
    - Elder: use formal, respectful tone
  - Then, give the tone rule as instruction to the Gen AI agent
- Implementation: Prompt engineering with GPT-based agent

```
{'product': 'Microsoft Surface', 'ticket_content': "Ticket Type: Product inquiry, Ticket Subject: Refund request, Ticket Description: I'm unable to access my {product_purchased} account. It keeps displaying an 'Invalid Credentials' error, even though I'm using the correct login information. How can I regain access to my account? Solution 1 I'm unable to find the option to perform the desired action in the {product_purchased}. Could you please guide me through the steps?", 'intent': 'Request Refund', 'tone': 'Respond in a casual and friendly tone suitable for younger audiences.'}
```

Assistant: Hey there! I'm sorry to hear you're having trouble accessing your Microsoft Surface account. It sounds like you're getting an 'Invalid Credentials' error, even though you're using the right login info. Let's get this sorted out for you!

First, let's try resetting your password. You can usually do this by clicking on the 'Forgot Password' link on the login page. If that doesn't work, you can reach out to Microsoft Support for further assistance. They'll be able to help you regain access to your account.

If you're having trouble finding the right option on your Microsoft Surface, don't worry! I can guide you through the steps. Just let me know where you're getting stuck, and I'll do my best to help you out. We'll have you back up and running in no time! 😊

# Task II – Human Escalation

- How would you come up with new features to determine which queries should be escalated to human operators?
  - Beside the priority, user intent can be used to decide if the query should be escalated to human operator
  - Also, we can use content-based feature: e.g. if the ticket is really negative?
- As the "Human Escalated" labels are not available, I propose a heuristic to label the tickets using ticket sentiment, user intent and ticket priority
  - User intent in `escalated_intents`
  - Negative ticket with high score ( $> 0.95$ )
  - Priority is Critical or High
  - Around 30% of the ticket is escalated to human
- I trained a classifier using ticket content using the

```
escalated_intents = [  
    "Recover Account Access",      # Critical for user access  
    "Report Data Loss",           # Potentially irreversible loss  
    "Fix Payment Issue",          # Involves financial sensitivity  
    "Request Refund",             # May involve policy exceptions  
    "Report Hardware Issue",      # Needs diagnostics or replacement  
    "Report Software Bug",        # Technical issue, may need logging/debugging  
    "Request Cancellation",       # Retention-sensitive, may need confirmation  
    "Report Delivery Problem"     # Time-sensitive and frustrating for users  
]  
  
non_escalated_intents = [  
    "Help with Product Setup",    # Routine onboarding  
    "Resolve Network Problem",    # Medium priority troubleshooting  
    "Battery Draining Too Fast",  # Performance-related but not critical  
    "Fix Display Issue",         # Visual/UX issue, typically fixable  
    "Need Installation Support",  # Basic guidance  
    "Check Peripheral Compatibility", # Informational query  
    "Check Product Compatibility", # Informational query  
    "Ask for Product Recommendation" # Low-stakes query  
]
```

# Task III – Open Questions

- Design an AI or GenAI approach to predict customer satisfaction (using the "Customer Satisfaction Rating" field) based on other features in the dataset.

Process:

- Feature use:
  - Customer Age group: categorical
  - Customer Gender: categorical
  - Ticket Content: text
  - Resolution: text
  - Ticket Sentiment: categorical
  - User Intent: categorical
  - Target: Rating [1-5]
- Pipeline:
  - Preprocess Ticket Content and Resolution
  - Use pretrained model for text embedding, and train embedding for categorical features
  - Apply MLP on the concat embeddings, then using CrossEntropyLoss for multi-label classification

## Task III – Open Questions

- Describe an approach to detect potential fraudulent activities, high-risk situations or complaint using the available data fields. How would you integrate this into the AI model's decision-making process for escalation to human operators?
  - Fraud Detection:
    - The dataset seems to contains ticket with code injection and text with html tags
    - Solution: Build filters for such cases as 2 additional features: has\_html and has\_code\_injection
  - Complaint:
    - Use sentiment analyzer as an additional feature
    - Record the ticket with Negative labels and high score (e.g. > 0.95)

Some examples:

I'm having an issue with the {product\_purchased}. Please assist. <script src="../../../libs/products/touches/touches.js"></script>

If we make our "instructions" I've noticed that the issue occurs consistently when I use a specific feature or application on my {product\_purchased}.

-----  
I'm having an issue with the {product\_purchased}. Please assist. <script type="text/javascript">(function() { var c1 = document.getElementById('Product')[1].parentNode; c1. The issue I'm facing is intermittent. Sometimes it works fine, but other times it acts up unexpectedly.  
-----

## Task III – Open Questions

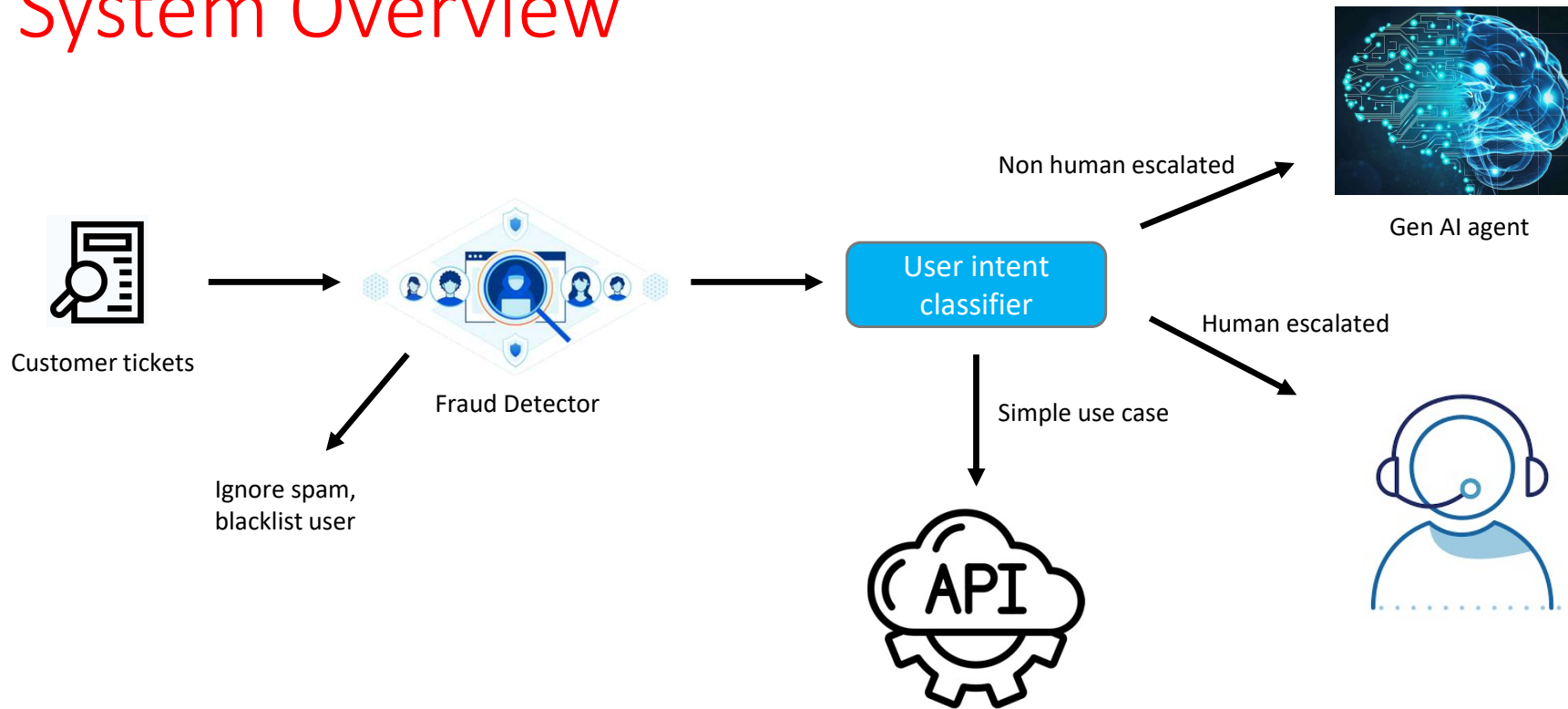
- Explain how you would implement a system that continuously learns and improves its responses over time, using the "Ticket Description", "Resolution", and "Customer Satisfaction Rating" fields.
  - First, from the given dataset, construct 2 sets of resolved tickets with high rated resolution and low rated resolution
  - SFT a strong based-model (e.g. Mistral 7B) with high-rated resolution to instruct the model to generate good resolution
  - Further DPO the model on the two sets with good and bad resolution to have the model learn how to generate good resolution and avoid bad resolution
  - Periodically update the DPO training with new good and bad resolution samples to have the continuously learn and improves it responses over time

# Task III – Open Questions

- How would you design a hybrid system that combines rule-based approaches for simple queries with more advanced machine learning models for complex tasks? Discuss the pros and cons of this approach.
  - Fraud Detection rule-based filter
    - Ignore ticket that has fraud patterns (e.g. high-risk email domain, contain code injection, contains spam keywords)
  - Design some rule-based response from some basic user intent
    - E.g. Help with product setup, ask for recommendation
- Pros:
  - Efficiency: time and resource saving
  - Flexibility: can modify each module easily
- Cons:
  - Misclassification at router level may delay resolution.
  - Static rules need regular maintenance.
  - Hybrid flow may obscure what went wrong and where.



# System Overview



Thank you for your attention !