BlendX : Complex Multi-intent Detection with Blended Patterns

Yejin Yoon, Jungyeon Lee, Kangsan Kim, Chanhee Park and Taeuk Kim

LREC-COLING 2024 Submitted Paper

Yejin Yoon

Natural Language Processing Lab.,
Hanyang University.

Paper Submission

LREC-COLING 2024

BlendX: Complex Multi-Intent Detection with Blended Patterns

Anonymous submission

Abstract

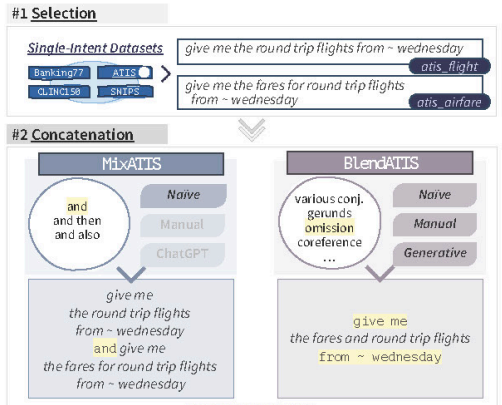
Task-Oriented Dialogue (TOD) systems typically suppose that a user utterance corresponds to a single intent. This assumption may be misaligned with real-world scenarios where users often express multiple intents simultaneously. While there is an emerging interest in Multi-Intent Detection (MID), existing in-domain datasets such as MixATIS and MixSNIPS have limitations in their formulation. To address these issues, we present BlendX, a suite of refined datasets featuring more diverse patterns than their predecessors, elevating both its complexity and difficulty. For dataset construction, we utilize both rule-based heuristics as well as a generative tool–OpenAI's ChatGPT—which is augmented with a similarity-driven strategy for utterance selection. To ensure the quality of the proposed datasets, we also introduce three novel metrics that assess statistical properties of an utterance related to word count, conjunction use, and pronoun usage. Extensive experiments on BlendX reveal that state-of-the-art MID models struggle with the challenges posed by the new datasets, highlighting the need to reexamine the current state of the MID field.

Keywords: Multi-Intent Detection, Task-Oriented Dialogue

1. Introduction

The successful implementation of Task-Oriented Dialogue (TOD) systems begins with the precise recognition of user intents. By accurately discerning the queries embedded in user inputs and routing them to the relevant components, the systems can adeptly respond, thereby effectively fulfilling user requests. In general, such systems are constructed on the assumption that each user utterance correlates exclusively with a single intent, which often diverges from practical scenarios.

Contrary to the conventional setting, the task of **Multi-Intent Detection (MID)** presents a more nuanced and comprehensive challenge for TOD



Presentation Agenda

1 **PRE-REQUISITE**
Key Terms

2 **INTRODUCTION**
Problem States / Background

3 **DISCUSSION #1**
Data Construction

4 **DISCUSSION #2**
Experiment and Analysis

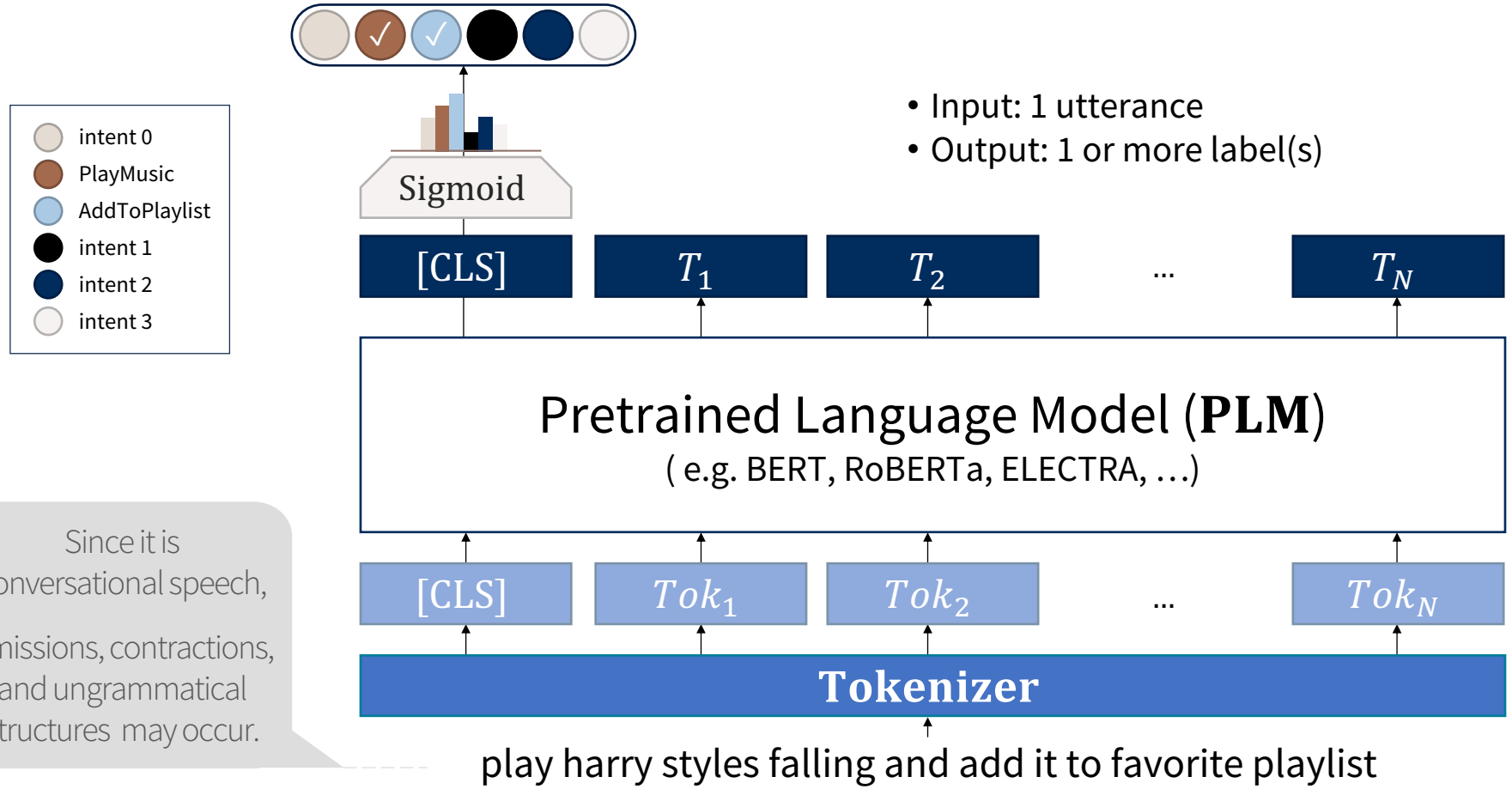
5 **CONCLUSION**
Contribution

PRE-REQUISITE

~~# Task-oriented Dialogue System~~

Multi-intent Detection

Multi-intent Detection (MID)



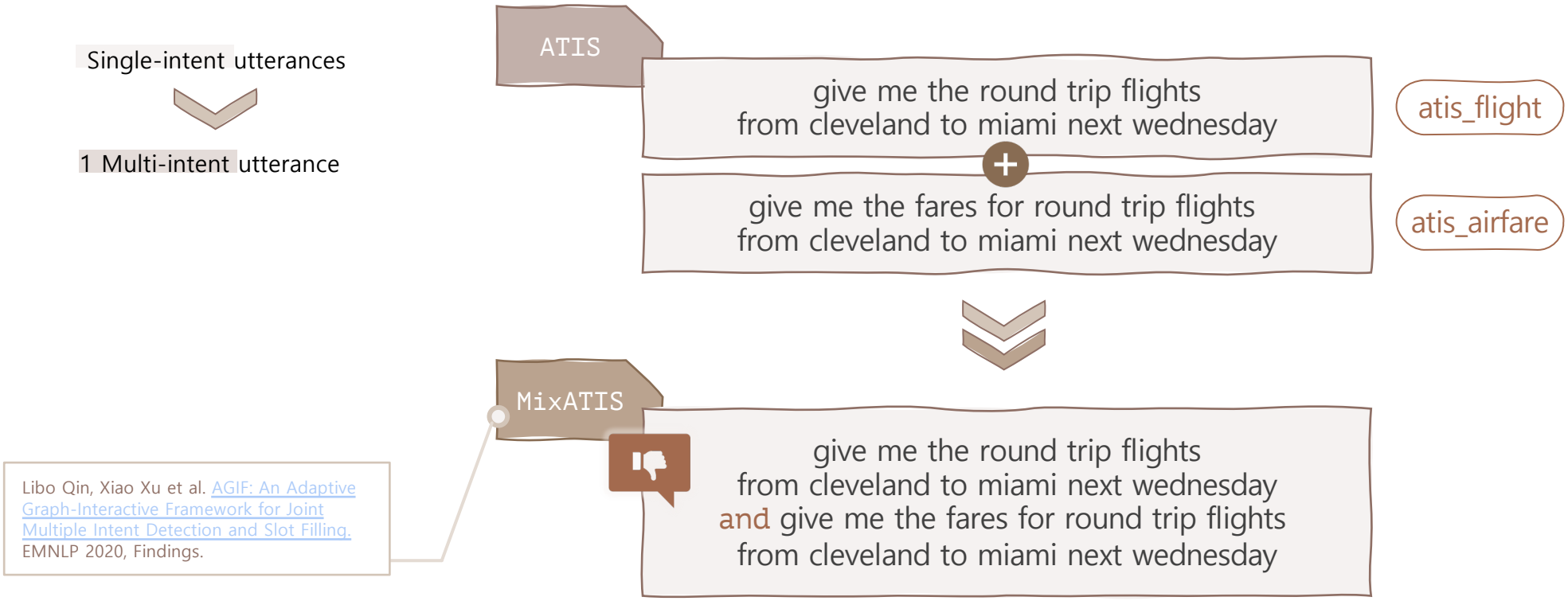
Identify and respond to multiple intentions or requests within a single user utterance

Introduction

Problem States

Background

Benchmark Datasets Analysis (1/2)

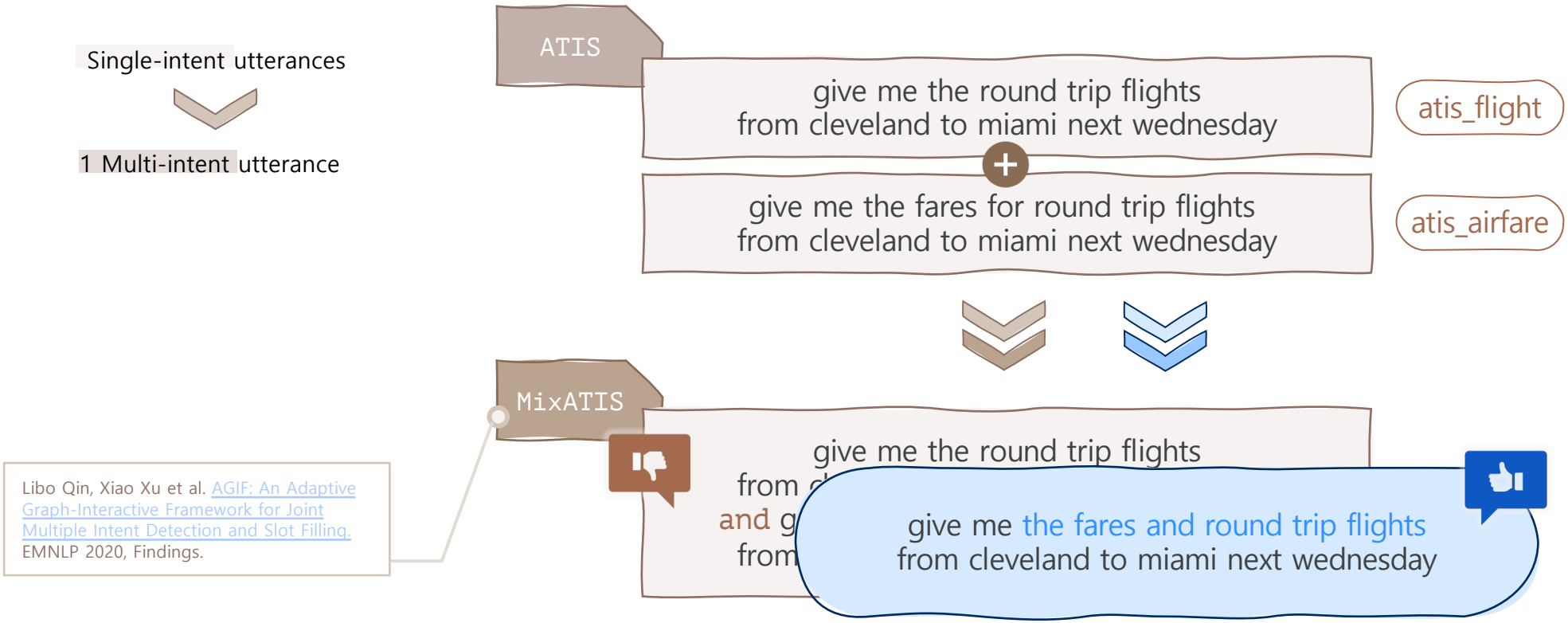


The dataset relies on only a few specific connectors ("and", "and then", "and also") when concatenating 2 or more single-intent utterances.

→ Real-world conversations often involve **more varied and complex ways of combining intents**

Benchmark Datasets Analysis (2/2)

EMNLP2022 Findings / AGIF/ Qin et al.



We are focused on constructing our own set that better mirrors natural language usage
to provide more **challenging** and **realistic** resources
for training and testing multi-intent detection models.

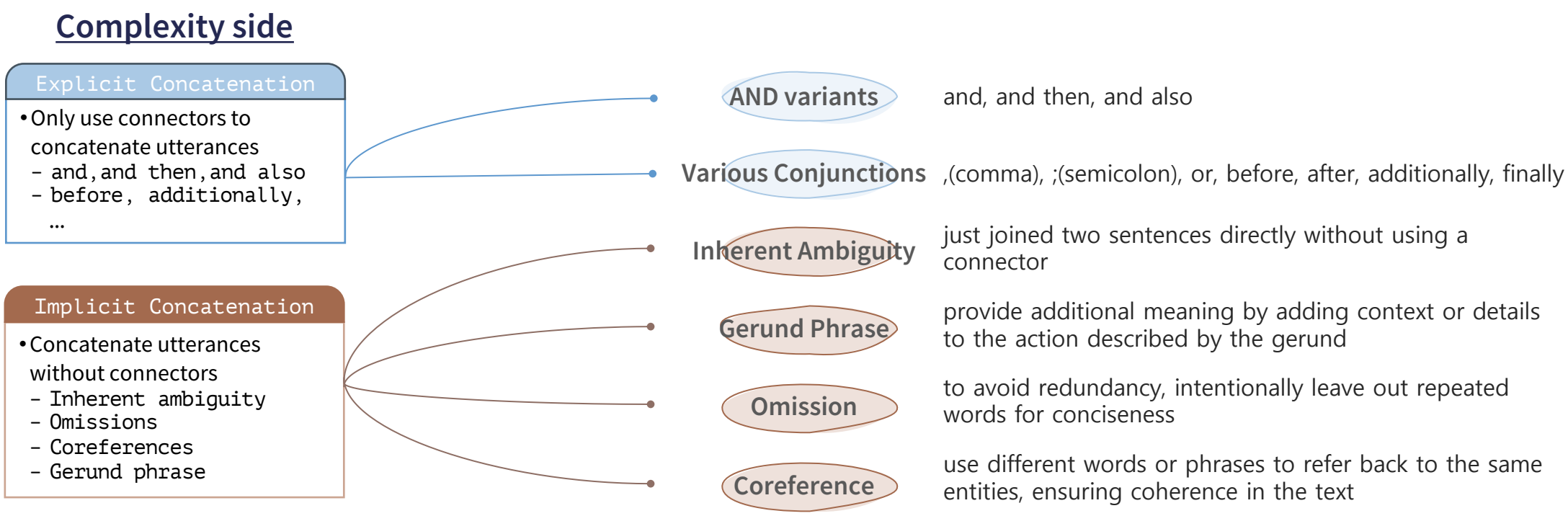
Discussion 1

Concatenation: Single- to Multi- intent utterance

Utterance Selection

Categories of Concatenation Complexity-side

- **Complexity side**
  - Explicit Concatenation: use connectors during concatenation
    → AND variants / Various Conjunctions
  - Implicit Concatenation: do NOT use connectors during concatenation
    → Inherent Ambiguity / Gerund Phrase / Omission / Coreference



(Intuitive) ChatGPT in Concatenation (1/2)

- **Prompt Engineering for ChatGPT Concatenation**

```
You are a native English speaker.
[Task Definition] Combine 2 or 3 sentences as one single sentence.
[Goal] The focus is on creating a single sentence that captures the essence of both ideas without unnecessary redundancy.
[Instructions] - Avoid adding just punctuation.
               - Don't paraphrase.
               - Don't compromise the meaning of each sentence.
               - Don't capitalize all characters.
               - Don't replace numbers with radix.
               - Maintain the intent of each sentence.
               - Don't forget that if a sentence starts with a verb, it's a statement.
               - Do NOT use conjunctions like 'and'
               - Don't print '[Good Answer]'
               - Don't print intent directly.
[Example 1]
         My dog is playful (dog's feature) + My dog loves chasing balls (dog's feature)
          [Good Answer] My playful dog loves chasing balls
          [Bad Answer] My dog is playful, and my dog loves chasing balls
          [Bad Answer] My dog is playful, and also loves chasing balls.
[Example 2]
         They finished the project(project done) + They had time(taking time)
          [Good Answer] With time on their hands, they finished the project
          [Bad Answer] They finished the project, and they had time
          [Bad Answer] They finished the project and had time
[Example 3]
         She answered the phone (answering)) + She was making dinner (cooking)
          [Good Answer] While answering the phone, she was making dinner
          [Bad Answer] She answered the phone, and She was making dinner
Combine the following sentences naturally. Inside the parentheses is the intent of each sentence. :
{utt1} (intent: {intents[0]})  + {utt2} (intent: {intents[1]})
```

Returning results that don't follow the explicit constraints we gave ChatGPT

Few-shot setting

(Intuitive) ChatGPT in Concatenation (2/2)

- **Failure of Using ChatGPT**
  - Over 90% of the 100 randomly sampled instances were connected by 'and'
  - Failed to maintain intent in at least 15 instances
    - Intent removed / changed
    - Fail to merge the utterances

	Utterance 1	Utterance 2	Concatenation	Implication
failed	weather for frankfort (`GetWeather`)	3 out of 6 for the last album (`RateBook`)	frankfort's weather gets a 3 out of 6 rating for the last album	Intent removed
failed	what day of the week do flights from nashville to tacoma fly on (`atis_day_name`)	flight number from houston to dallas (`atis_flight_no`)	flights from nashville to tacoma fly on what day of the week and what is the flight number from houston to dallas	Intent changed
failed	what is mci (`atis_abbreviation`)	list la (`atis_city`)	combine the sentences: "what is mci?" and "list la".	Failed to merge
succeeded	play the playlist funtime activity (`AddToPlaylist`)	add adele onto funtime activity playlist (`PlayMusic`)	play the playlist, adding adele to the funtime activity playlist	Gerund phrase
succeeded	how do i freeze my bank account (`account_blocked`)	do you know why my bank account is frozen (`freeze_account`)	how do i freeze my bank account, and do you know why it is frozen	Coreference

Table 1: Failed (Top) and succeeded (Bottom) results of the **Generative Approach** and their implications.

Only few cherry-picked examples were, brilliantly, implicitly concatenated, which is what we intended.

Overview of 2 Methods to Utterance Concatenation

1

Single-Intent Datasets

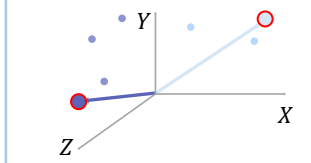
Banking77 | ATIS | CLINC150 | SNIPS

Preprocessing

- Lowercase
- Remove punctuation

2 For Manual Approach

Utterance Selection



Randomly select 2~3 utterances

Manual Concatenation

AND variants
Various Conjunctions
Inherent Ambiguity
Gerund Phrase

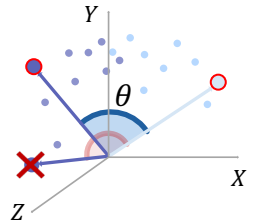
Naïve Approach
- Approach employed by MixX

Manual Approach
- Rule-based concatenation technique

2 For Generative Approach

3

Utterance Selection



Select 2~3 similar utterances based on cosine similarity

Generative Concatenation

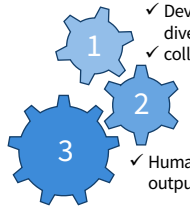
AND variants
Various Conjunctions
Inherent Ambiguity
Gerund Phrase
Omission
Coreference

ChatGPT Bard Llama

Generative Approach
- Concatenate utterances by using generative language model, ChatGPT

Filtering Concatenated Utterances

Enhance the quality of outcomes from generative LLMs by implementing defined criteria and incorporating reviews from human experts to filter the results.



✓ Develop metrics to evaluate integration diversity and complexity
✓ collect results above a set threshold.

✓ Collect only the incorrect predictions made by the best model in multi-intent detection.

✓ Human-expert review of the generated output filtered by the above two steps.

4 BlendX

BlendBanking77 | BlendATIS
BlendCLINC150 | BlendSNIPS

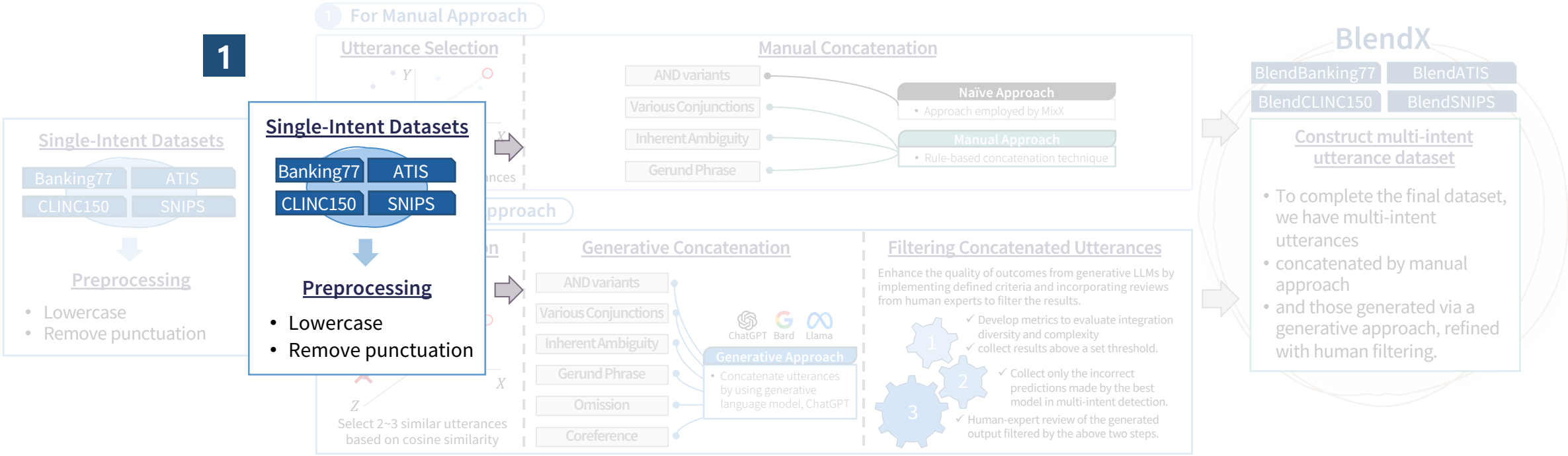
Construct multi-intent utterance dataset

- To complete the final dataset, we have multi-intent utterances
- concatenated by manual approach
- and those generated via a generative approach, refined with human filtering.

Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

1. **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.
2. **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

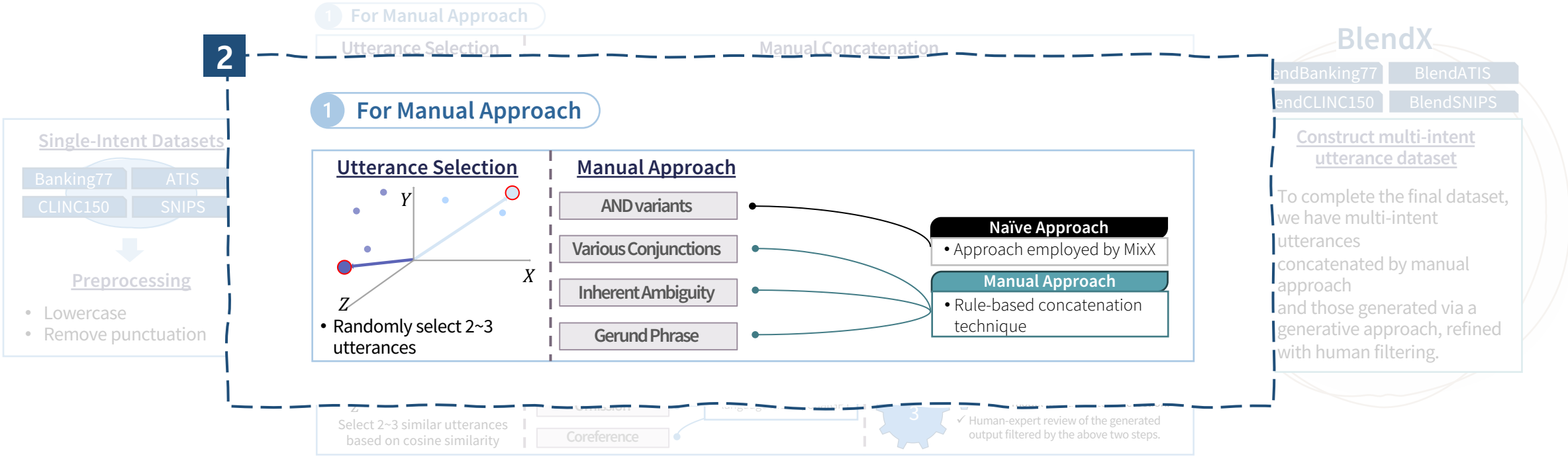
Overview of 2 Methods to Utterance Concatenation



Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

1. **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.
2. **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

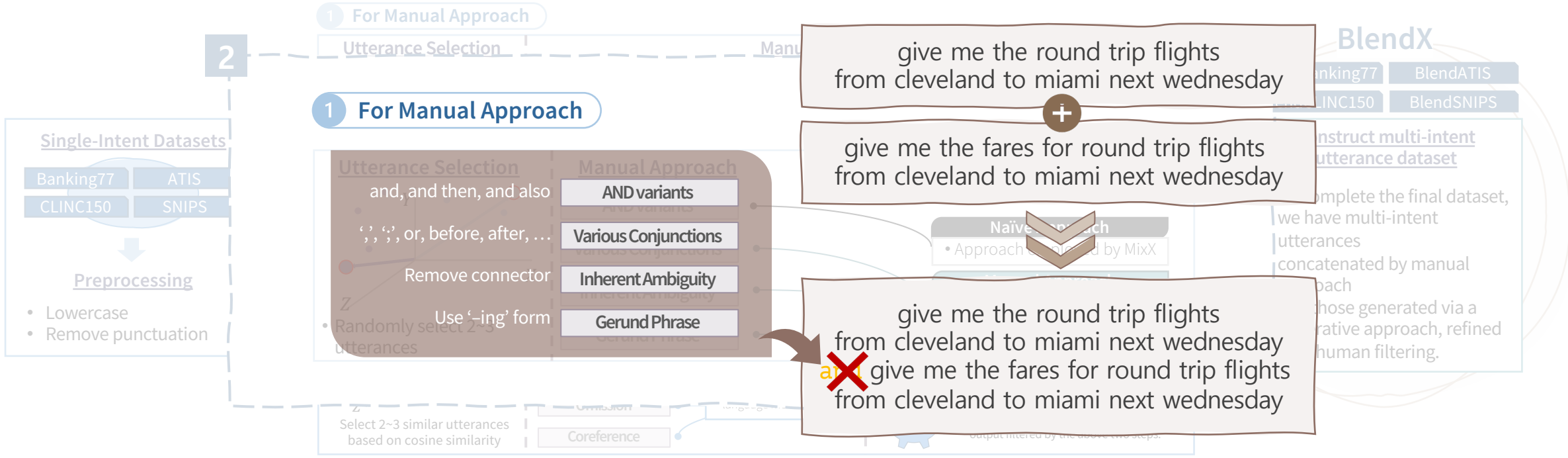
Overview of 2 Methods to Utterance Concatenation



Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

1. **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.
2. **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

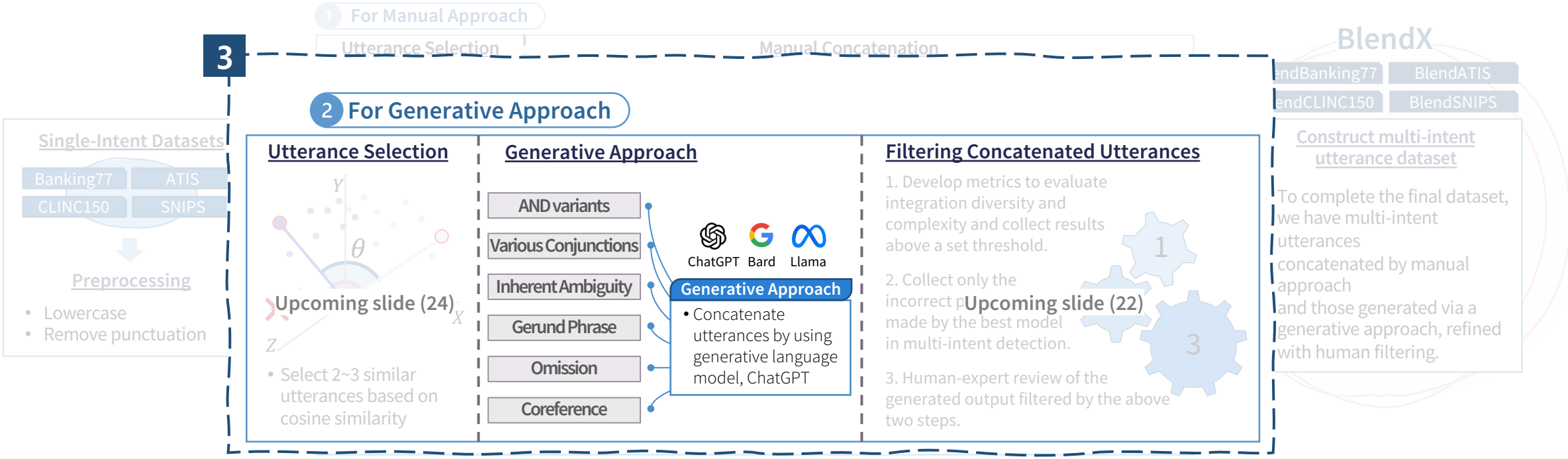
Overview of 2 Methods to Utterance Concatenation



Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

1. **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.
2. **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

Overview of 2 Methods to Utterance Concatenation

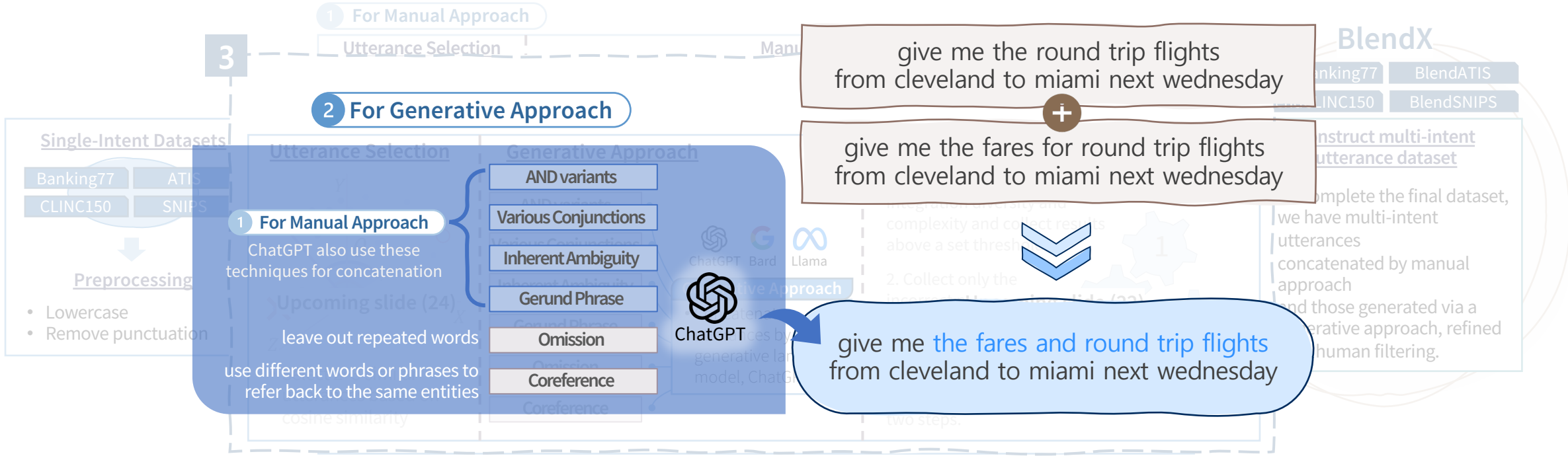


Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

1 **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.

2 **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

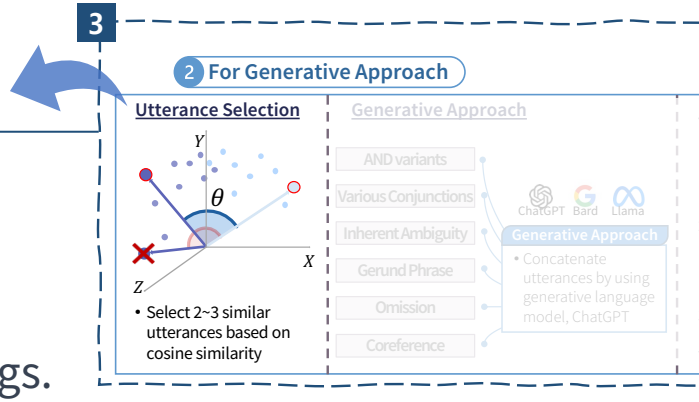
Overview of 2 Methods to Utterance Concatenation



Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

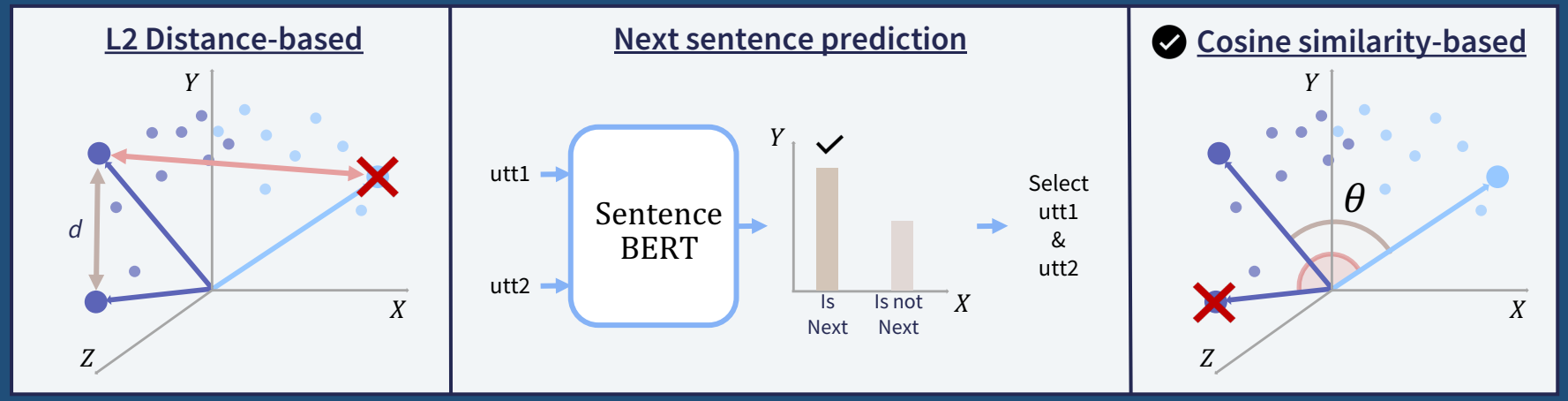
1. **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.
2. **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

Utterance Selection for ② Generative Approach

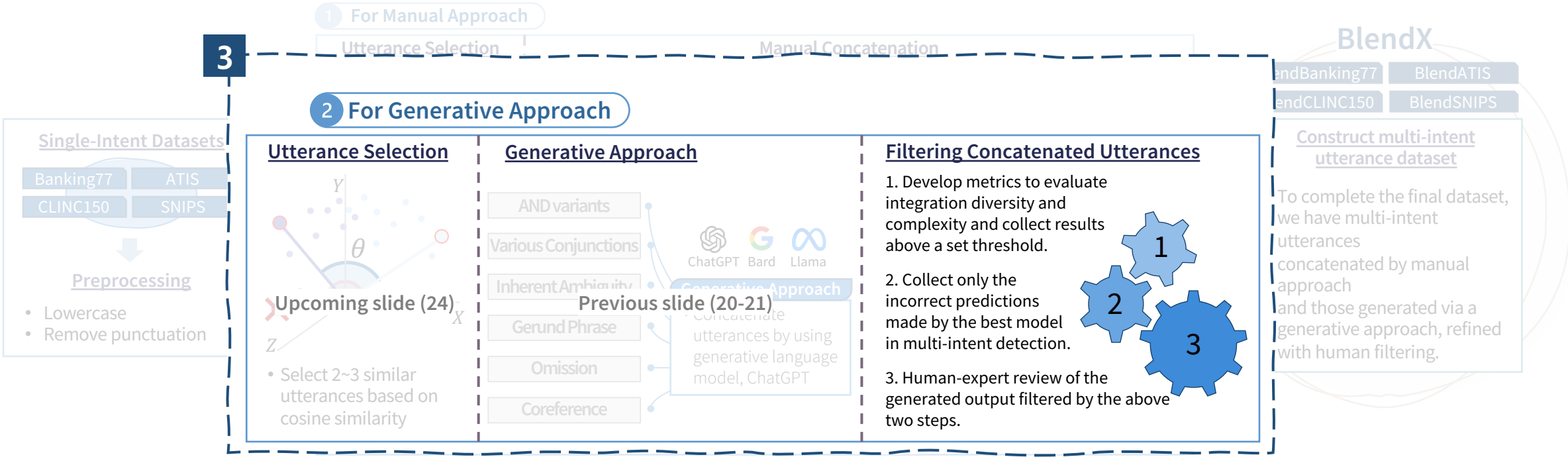


- **Process**
  1. Generate embeddings for each single-intent utterance using SentenceBERT.
  2. Select utterances for concatenation based on high similarity between embeddings.
     * Chosen utterances will have different intents.

- **Selection approach**
  - L2 Distance-based: Select utterances with close proximity in embedding space.
  - Next sentence prediction: Binary classification of whether a given pair of utterances are sequential.
  - ✓ Cosine similarity-based: Choose utterances with high cosine similarity between embeddings.



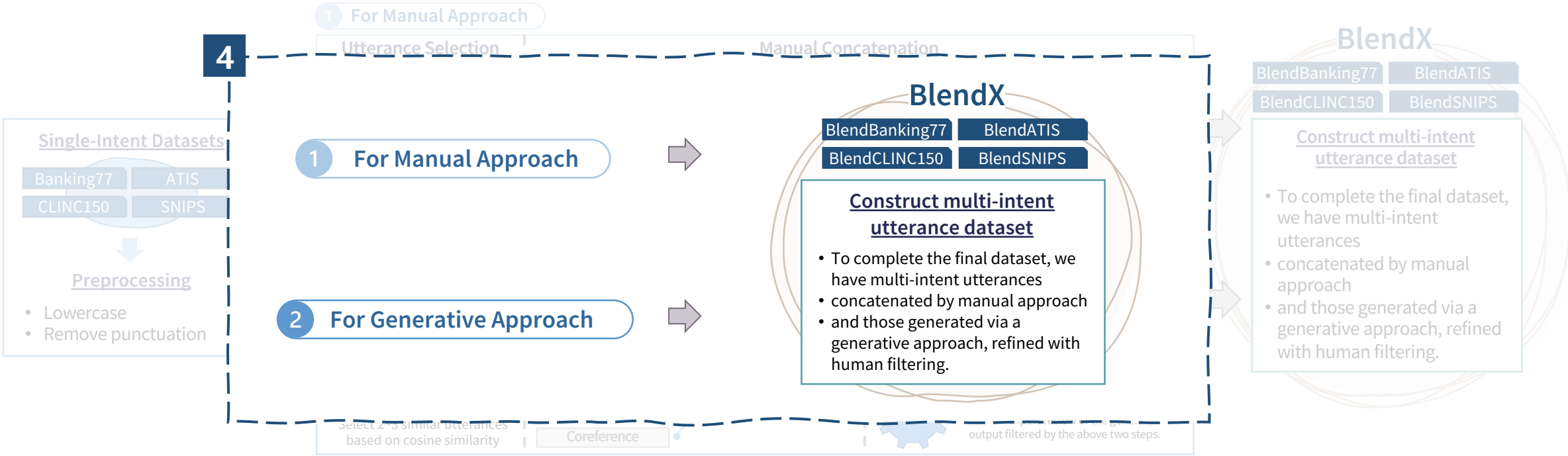
Overview of 2 Methods to Utterance Concatenation



Without generating new multi-intent utterances and ensuring they fit within the existing intent space, we propose 2 approaches :

1. **Manual Approach:** Concatenate utterances without using connectors, or if necessary, employ a various range of options.
2. **Generative Approach:** Explore methods to extend **ChatGPT**'s capabilities for producing more coherent multi-intent utterances by concatenating 2 or more single-intent utterances.

Overview of 2 Methods to Utterance Concatenation



BlendX : Complex multi-intent detection with blended patterns

Dataset	# of intents	Training	Dev	Test	Total
BlendSNIPS	7	50,625	2,613	2,615	55,853
BlendATIS	18	20,250	1,125	1,125	22,500
BlendBanking77	77	36,390	2,009	2,021	40,420
BlendCLINC150	147	54,899	2,889	2,977	60,765

∑ (total) = 179,538

- Source Dataset : SNIPS, ATIS, Banking77, CLINC150
- Random selection for **Manual** Concatenation Approach
- Cosine Similarity-based selection for **Generative** Concatenation Approach

Discussion 2

3 Metrics

Experiment and Analysis

Evaluation – 3 Metrics (1/3)

- **3 Metrics**
  - utt: concatenated utterance with 2 or more intents
  - n: Number of single-intent utterances used for concatenation

$\mathrm{W}(utt, n)$

Word count

$$\mathrm{W}(utt,n) \stackrel{\text{def}}{=} \mathbf{1}_{\mathbb{Z}-\mathbb{N}}\left(|utt|_{word} - \sum_{i=1}^{n} |utt_i|_{word}\right).$$

Check if the **word count** difference before and after an utterance concatenation is zero or negative

(to ascertain a decrease in word count)

$\mathrm{C}(utt, n)$

Conjunction

$$\mathrm{C}(utt,n) \stackrel{\text{def}}{=} \mathbf{1}_{\mathbb{Z}-\mathbb{N}}\left(|utt|_{conj} - \sum_{i=1}^{n} |utt_i|_{conj}\right).$$

Verify if the number of **conjunctions** before and after an utterance changes to zero or less

(to determine the elimination or reduction of conjunctions)

* **conjunctions** such as: and, or, before, after, additionally, finally, ‘,’, ‘;’

$\mathrm{P}(utt, n)$

Pronoun

$$\mathrm{P}(utt,n) \stackrel{\text{def}}{=} \mathbf{1}_{\mathbb{N}}\left(|utt|_{pron} - \sum_{i=1}^{n} |utt_i|_{pron}\right).$$

Assess if the difference in **pronoun count** before and after an utterance is one or more

(to identify the usage of pronouns)

* **pronoun** such as : it, them, their, theirs, this, that, those, these

An implicitly concatenated utterance is likely to receive 1 in the metrics evaluation.

Evaluation – 3 Metrics (2/3)

- **Example of applicating 3 metrics**

	Concatenation	utt1	utt2	Difference	Metric
example #1	add another song to my 88 keys playlist playing it	play my 88 keys playlist	add another song to my 88 keys playlist		
Words	10	5	8	10 - (5 + 8) = -3	W(• , 2) = 1
Conjunctions	0	0	0	0 - (0 + 0) = 0	C(• , 2) = 1
Pronouns	1	0	0	1 - (0 + 0) = 1	P(• , 2) = 1
example #2	i need to clear my to-do list and then repeat it	clear my to do list	repeat my to do list		
Words	11	5	5	11 - (5 + 5) = 1	W(• , 2) = 0
Conjunctions	1	0	0	1 - (0 + 0) = 1	C(• , 2) = 0
Pronouns	1	0	0	1 - (0 + 0) = 1	P(• , 2) = 1

Utterance 1	play my 88 keys playlist (`PlayMusic`)			
Utterance 2	add another song to my 88 keys playlist (`AddToPlaylist`)			
Strategies	**Concatenation Results**	$\mathrm{W}(utt, 2)$	$\mathrm{C}(utt, 2)$	$\mathrm{P}(utt, 2)$
Explicit Concatenation	play my 88 keys playlist **and** also add another song to my 88 keys playlist	0	0	0
Implicit Concatenation				
Inherent Ambiguity	play my 88 keys playlist add another song to my 88 keys playlist	1	1	0
Omissions	play my 88 keys playlist and add another song	1	0	0
Coreferences	play my 88 keys playlist and add another song to it	1	0	1
Gerund Phrase	add another song to my 88 keys playlist playing it	1	1	1

Table 3: Various concatenation classes, accompanied by their examples and respective metric values.

Evaluation – 3 Metrics (3/3)

- **Results using 3 metrics for each approach**

Metric	SNIPS			ATIS			Banking77			CLINC150		
	Naïve	Manual	Generative	Naïve	Manual	Generative	Naïve	Manual	Generative	Naïve	Manual	Generative
$W(utt, 2)(\uparrow)$	0%	**37%**	29%	0%	**36%**	18%	0%	**46%**	37%	0%	**48%**	28%
$C(utt, 2)(\uparrow)$	0%	**56%**	10%	0%	**52%**	15%	0%	**50%**	27%	0%	**56%**	32%
$P(utt, 2)(\uparrow)$	0%	0%	**7%**	0%	0%	**8%**	0%	0%	**13%**	0%	0%	**6%**

Table 4: Comparative analysis of the three concatenation approaches: Naïve, Manual, and Generative. Notably, the Manual method demonstrates pronounced efficiency in reducing utterance length.

Our approach, incorporating both **manual** and **generative** methods, achieves a more diverse range of explicit and implicit concatenation compared to existing techniques.

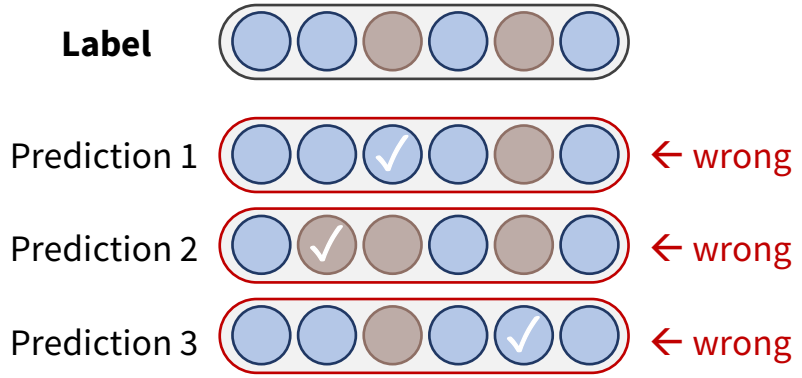
- Notably, **MixX** did not involve **implicit** concatenation.
  "**Naïve**" refers to the original construction method of **MixX**, meaning concatenation using only `and`, `and then`, and `and also`.
- Particularly, **manual concatenation** often resulted in shorter utterance lengths.
- Conversely, **generative concatenation** uniquely led to the use of pronouns.

Evaluation – SOTA Models with BlendX

- **Evaluate Baseline**

Model	Option		Dataset (Metric: accuracy)			
	Training	Test	SNIPS	ATIS	Banking77	CLINC150
TFMN	MixX	MixX	95.96*	76.80*	76.11	85.60
	MixX	BlendX	51.01	50.40	36.96	46.15
	BlendX	BlendX	92.96	76.00	62.69	78.06
SLIM	MixX	MixX	95.88*	91.48*	0.06	86.85
	MixX	BlendX	92.96	64.09	0.06	74.47
	BlendX	BlendX	95.72	77.33	0.10	84.44
gpt-3.5-turbo	-	MixX	77.56	33.60	23.72	45.55
	-	BlendX	73.23	29.96	22.76	40.98

- **Accuracy** in Multi-label Classification
  : We only considered it correct in cases of a **exact match**.



For (un)supervised SOTA models, we consistently observe a performance drop on our **BlendX** datasets with explicit as well as implicit concatenations.

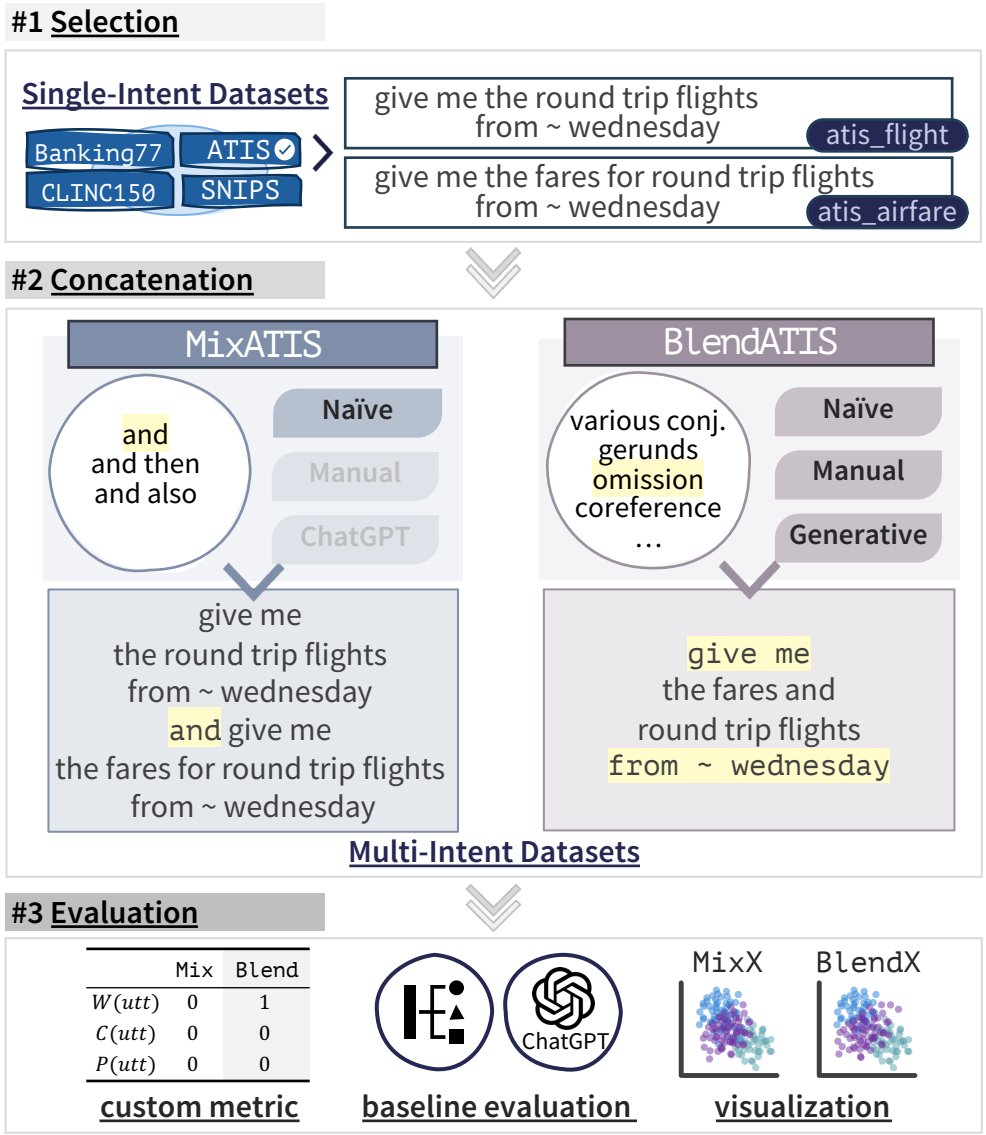
- 3-Baseline: implemented without slot-filling part
  - ✓ TFMN : predict # of intents k, and then top-k intents over the probability distribution
  - ✓ SLIM : threshold-based classification model using sigmoid function
  - ✓ ChatGPT : OpenAI's generative model ($\mathbf{gpt{-}3.5{-}turbo{-}0613}$)

Conclusion

Contribution

Main Findings

- **Identified limitations in existing multi-intent datasets**
  - MixX: Reliance on explicit concatenation through the `'and'` connector.
- **BlendX:** Constructing a more complex and realistic multi-intent dataset
  - Proposed 3 novel concatenation approaches : Naïve, Manual, Generative
  - Beyond random sentence selection, applied a similarity-based strategy in the **generative** concatenation approach.
  - Designed 3 statistical metrics for comparing and validating **BlendX** against the existing **MixX**: W, C, P
  - Upcoming dataset release : Extensions of **MixX** (**CLINC150**/**Banking77**) and new publication of the **BlendX** dataset.



Thank You

Yejin Yoon

HYU NLP Lab.
Hanyang University, South Korea

stillwithyou@hanyang.ac.kr

HYU 한양대학교 HANYANG UNIVERSITY

GITW