Multi-Intent Utterance Dataset Construction for Spoken Language Understanding

August 2024

Yejin Yoon

Dept. of Artificial Intelligence Application

HYU 한양대학교 HANYANG UNIVERSITY

Contents

1 **PRE-REQUISITE**
Key Terms

2 **INTRODUCTION**
Problem States / Background

3 **DISCUSSION 1 : Dataset Construction**
Utterance Concatenation / Utterance Selection

4 **DISCUSSION 2 : Experiments and Analysis**
3 Custom Metrics / SOTA Models Evaluation / Visualization

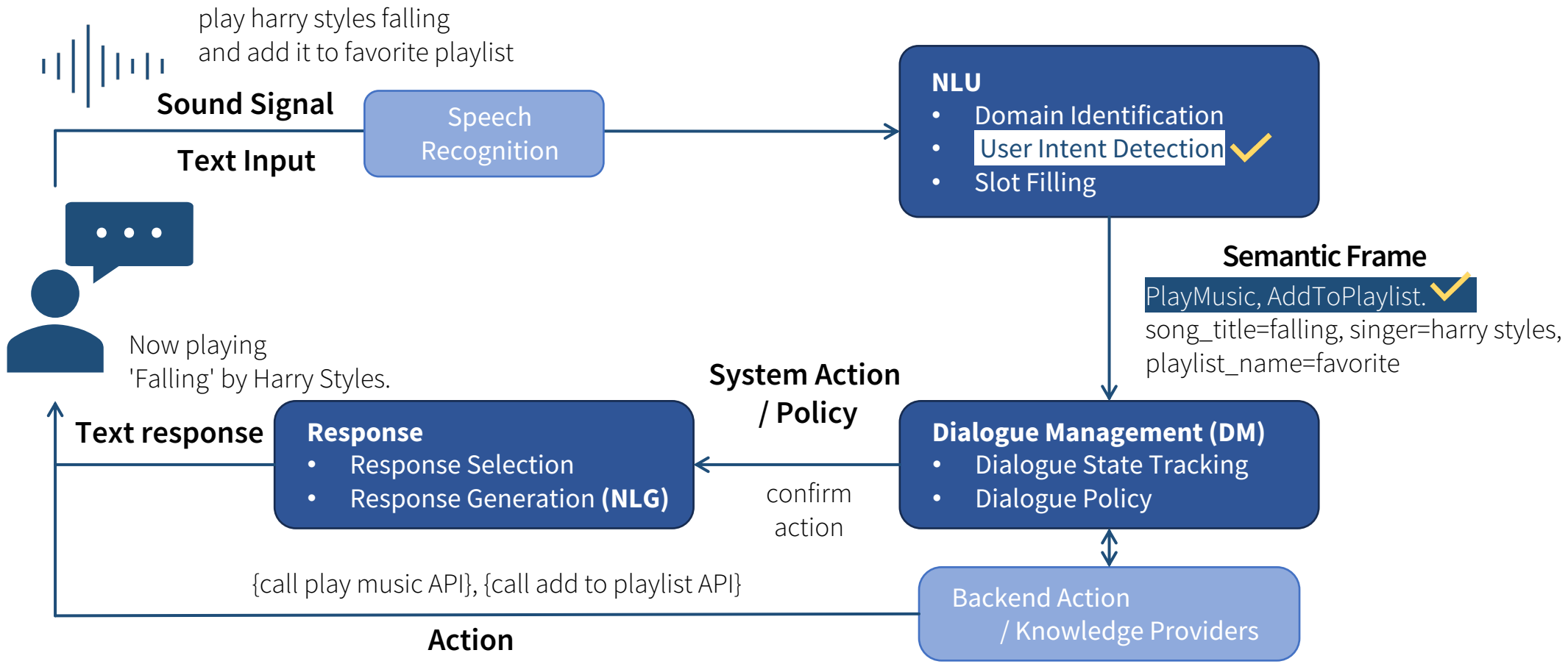
5 **CONCLUSION**
Contribution

PRE-REQUISITE

Task-oriented Dialogue System

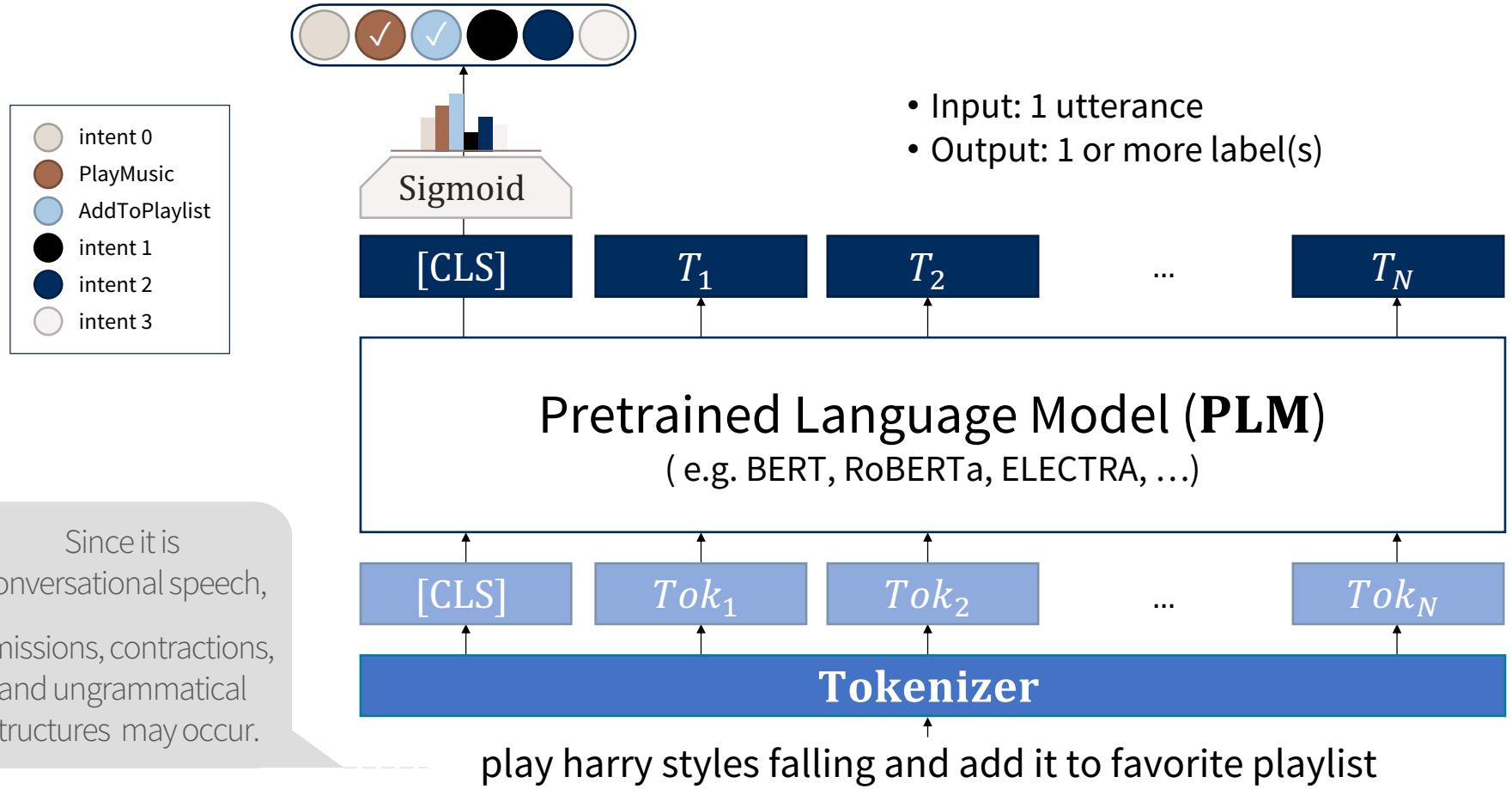
Multi-intent Detection

Task-oriented Dialogue System



- Help users achieve their specific goals
- Focus on understanding users, tracking states, and generating next actions

Multi-intent Detection



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

Introduction

Problem States

Background

Current Trends in MID

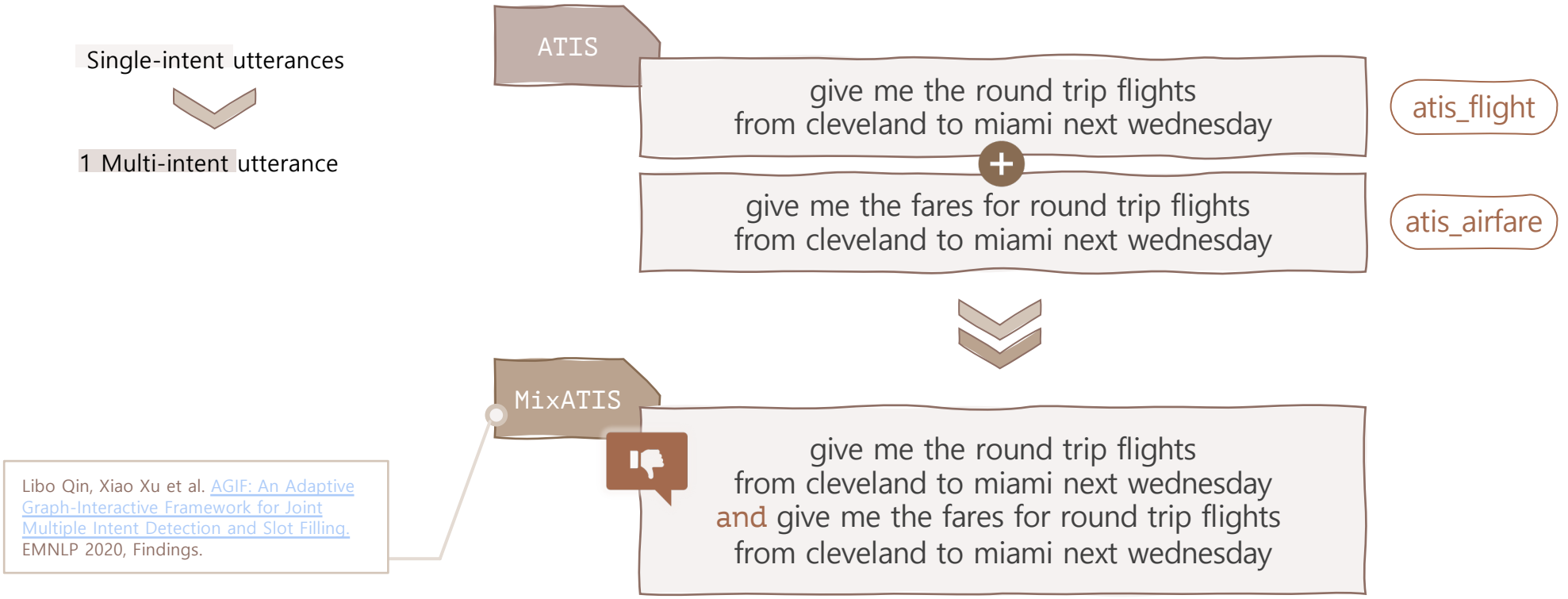
COLING2022 / TFMN/ Chen et al.

Model	MixATIS			MixSNIPS		
	Slot (F1)	Intent (Acc)	Overall (Acc)	Slot (F1)	Intent (Acc)	Overall (Acc)
SF-ID (*concat*) (2019)	87.4	66.2	34.9	90.6	95.0	59.9
Stack-Propagation ($thresh = 0.5$) (2019)	87.8	72.1	40.1	94.2	96.0	72.9
Joint Multiple ID-SF ($thresh = 0.5$) (2019)	84.6	73.4	36.1	90.6	95.1	62.9
AGIF ($thresh = 0.5$)(2020)	86.7	74.4	40.8	94.2	95.1	74.2
GL-GIN ($thresh = 0.5$)(2021)	**88.3**	76.3	43.5	94.9	95.6	75.4
SDJN ($thresh = 0.5$)(2022a)	88.2	77.1	44.6	94.4	96.5	75.7
SDJN+BERT ($thresh = 0.5$)(2022a)	87.5	78.0	46.3	95.4	96.7	79.3
Bert-baseline ($thresh = 0.3$)	83.1	74.8	42.6	95.5	95.7	80.2
Bert-baseline ($thresh = 0.5$)	86.3	74.5	44.8	95.5	95.6	80.1
Bert-baseline ($thresh = 0.8$)	85.6	75.8	43.5	95.2	96.7	80.6
TFMN (Bert-base)	88.0	**79.8**	**50.2**	**96.4**	**97.7**	**84.7**

- **Transformer-Based** Models in MID: Achieving Unprecedented High Performance.
- Competitive Edge: Latest Models Contending in Subtle Decimal Point Differences.
- MID and Slot Filling: **Jointly-learning** (multi-task learning) in Advanced Research Fields.

Benchmark Datasets Analysis (1/2)

EMNLP2022 Findings / AGIF/ Qin et al.

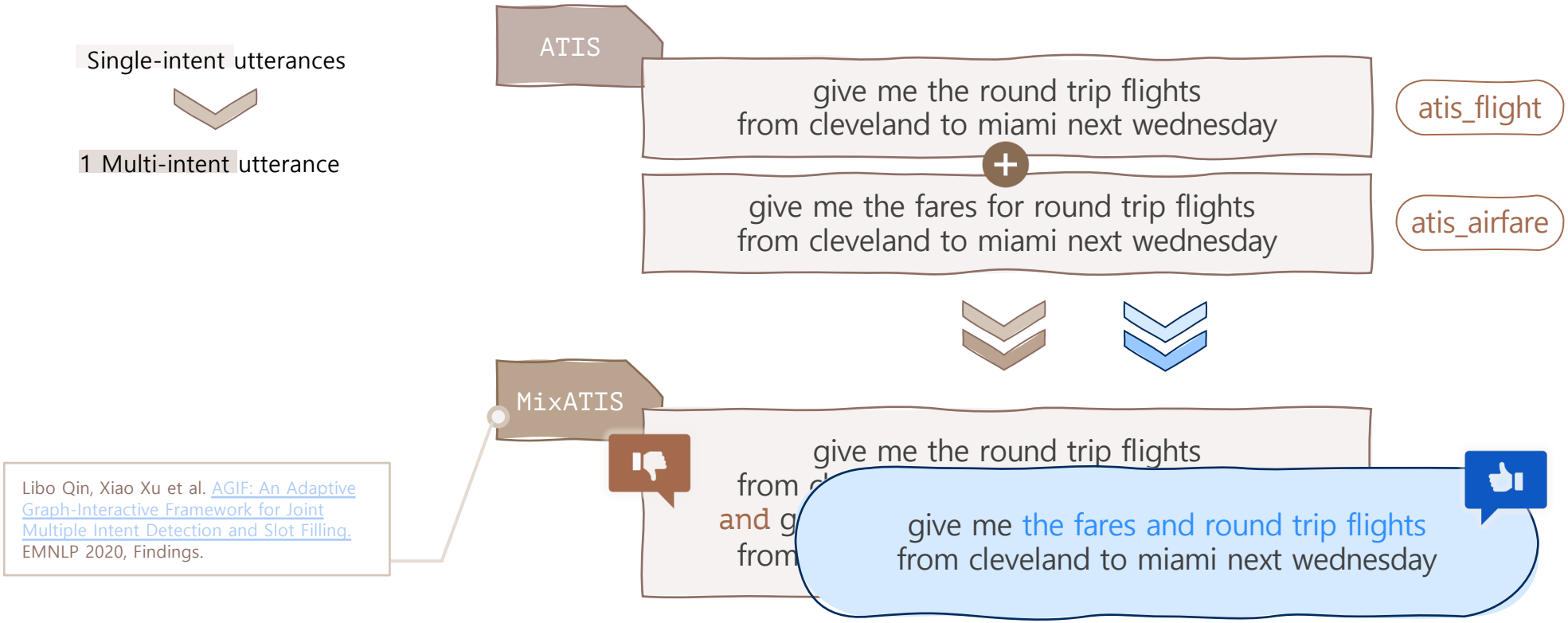


The dataset relies on only a few specific connectors ("and", "and then", "and also") when concatenating 2 or more single-intent utterances.

→ 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.

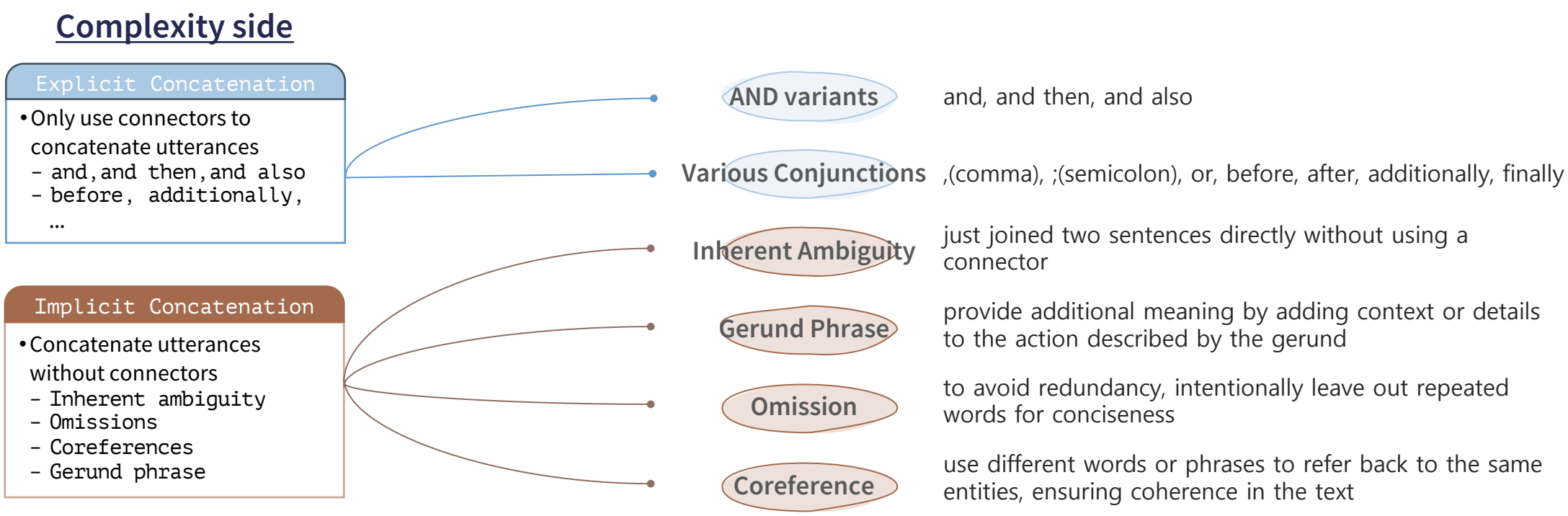
Discussion 1 : Dataset Construction

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

Success	Utterance 1	Utterance 2	Concatenation	Implication
X	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 (`RateBook`) removed
	show me the fares for a first class ticket from baltimore to dallas (`atis_airfare`)	i would like the time of all flights from san francisco to pittsburgh on sunday (`atis_flight_time`)	show me the fares for a first class ticket from baltimore to dallas and i would like to know all flights from san francisco to pittsburgh on sunday	Intent changed (from `atis_flight_time` to `atis_flight`)
	what is mci (`atis_abbreviation`)	list la (`atis_city`)	combine the sentences: "what is mci?" and "list la".	Failed to merge
O	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
	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 and successful results of the Generative Approach and their implications.

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

(Intuitive) ChatGPT in Concatenation (2/2)

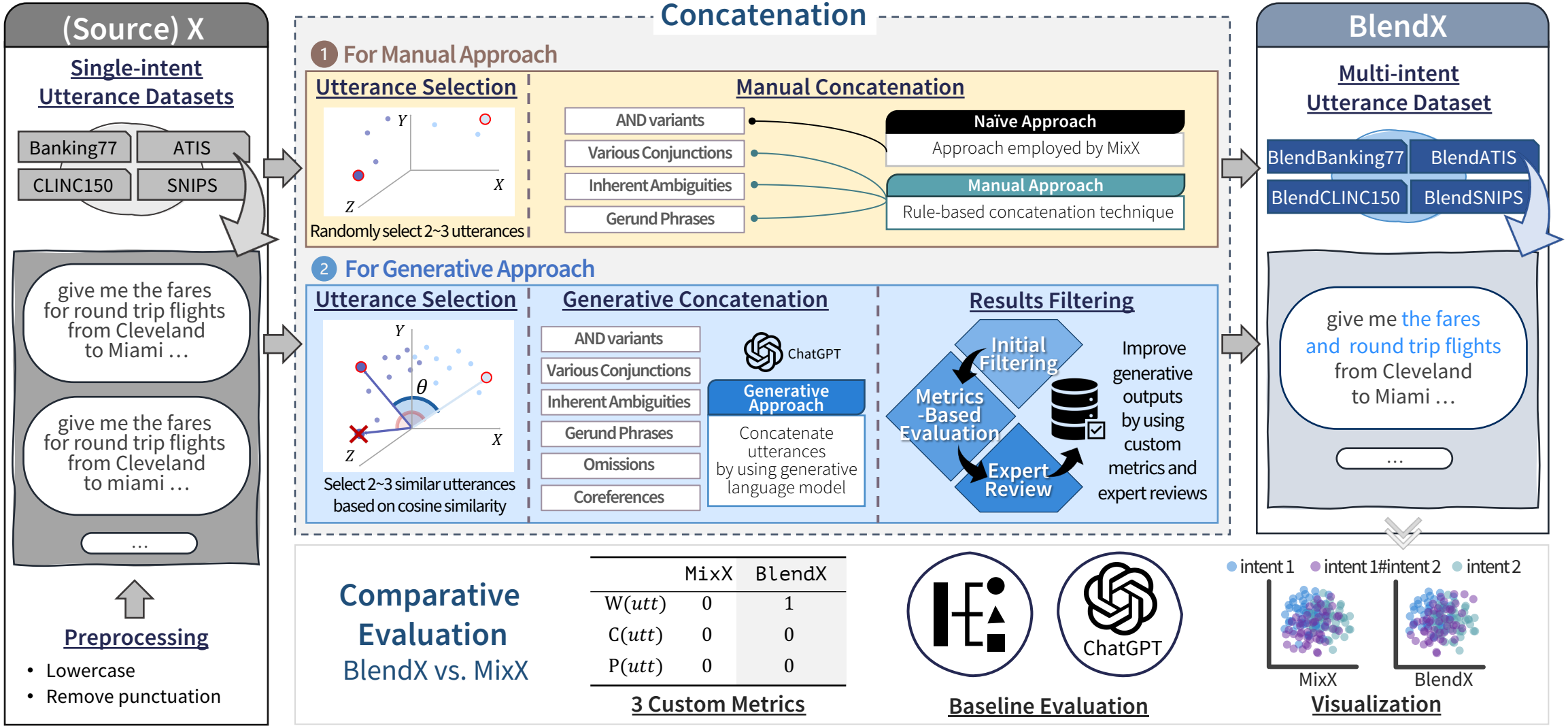
- **Failure of Using ChatGPT**
  - Over 90% of the 100 randomly sampled instances were connected by 'and'
  - Failed to maintain intent in at least 15 instances
    - Intent removed / changed
    - Fail to merge the utterances

Success	Utterance 1	Utterance 2	Concatenation	Implication
X	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 (`RateBook`) removed
	show me the fares for a first class ticket from baltimore to dallas (`atis_airfare`)	i would like the time of all flights from san francisco to pittsburgh on sunday (`atis_flight_time`)	show me the fares for a first class ticket from baltimore to dallas and i would like to know all flights from san francisco to pittsburgh on sunday	Intent changed (from `atis_flight_time` to `atis_flight`)
	what is mci (`atis_abbreviation`)	list la (`atis_city`)	combine the sentences: "what is mci?" and "list la".	Failed to merge
O	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
	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 and successful results of the Generative Approach and their implications.

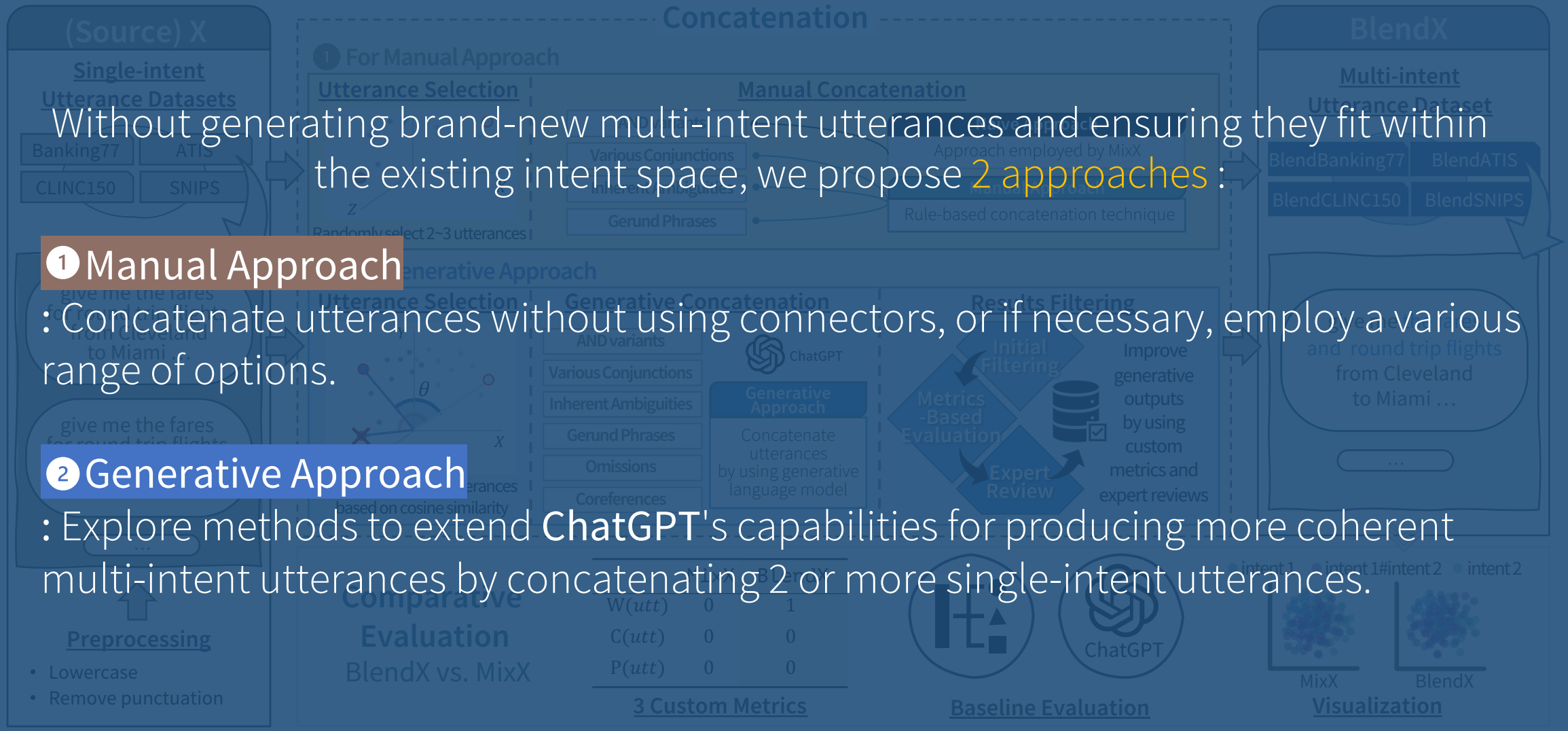
Only few cherry-picked examples were, brilliantly, implicitly concatenated, which is what we intended.

Overview of 2 Methods to Utterance Concatenation



	MixX	BlendX
W(utt)	0	1
C(utt)	0	0
P(utt)	0	0

Overview of 2 Methods to Utterance Concatenation



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

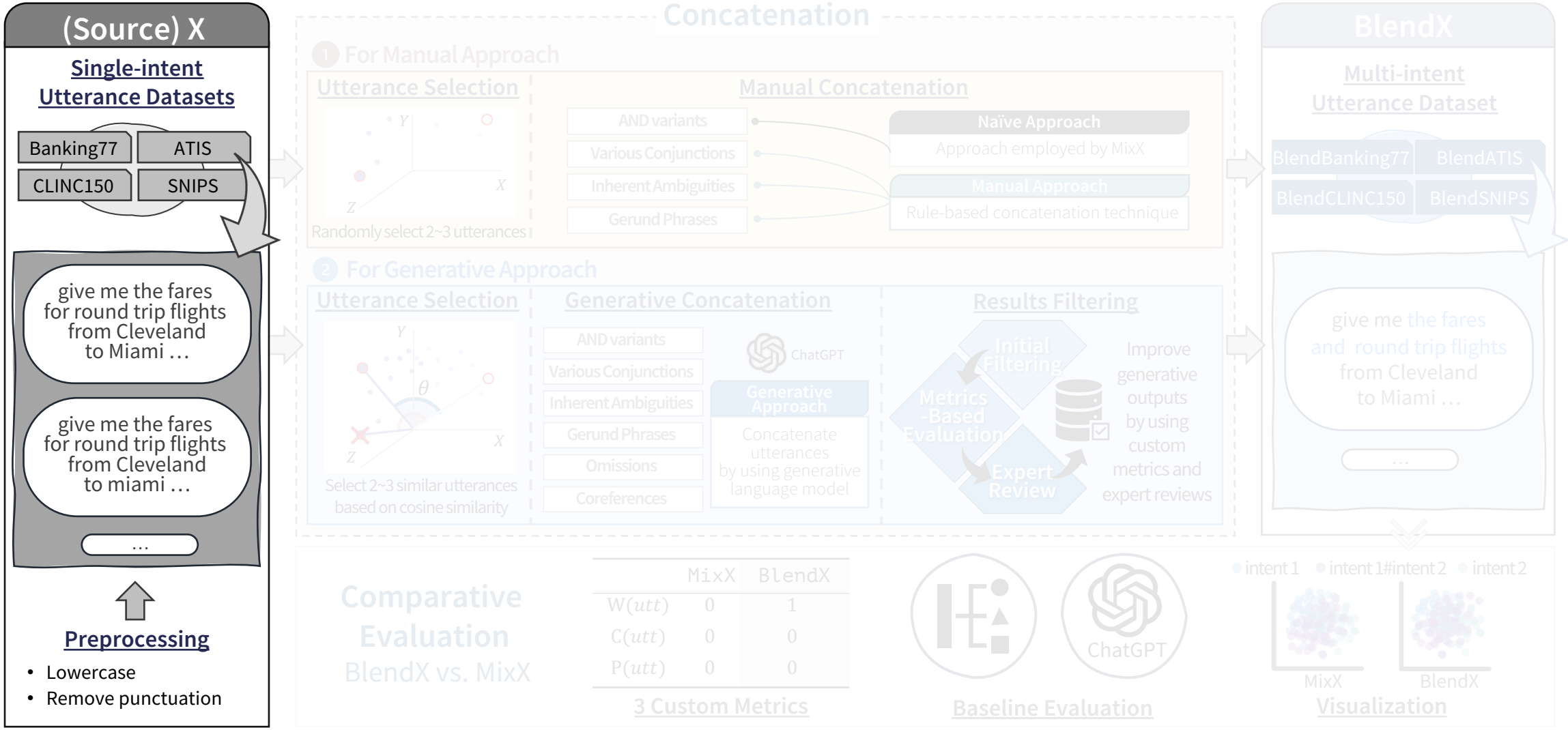
❶ Manual Approach

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

❷ Generative Approach

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

1) Preprocessing

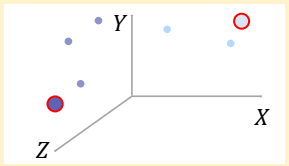


2) Concatenation

Concatenation

1 For Manual Approach

Utterance Selection



Randomly select 2~3 utterances

Manual Concatenation

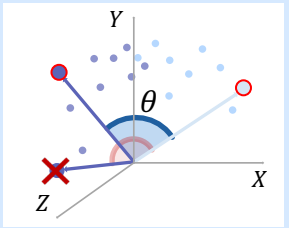
- AND variants
- Various Conjunctions
- Inherent Ambiguities
- Gerund Phrases

Naïve Approach
Approach employed by MixX

Manual Approach
Rule-based concatenation technique

2 For Generative Approach

Utterance Selection



Select 2~3 similar utterances based on cosine similarity

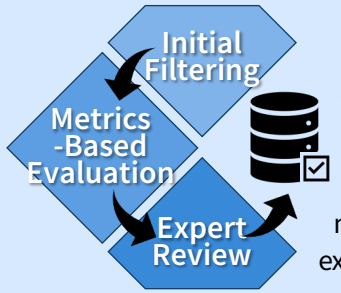
Generative Concatenation

- AND variants
- Various Conjunctions
- Inherent Ambiguities
- Gerund Phrases
- Omissions
- Coreferences

ChatGPT

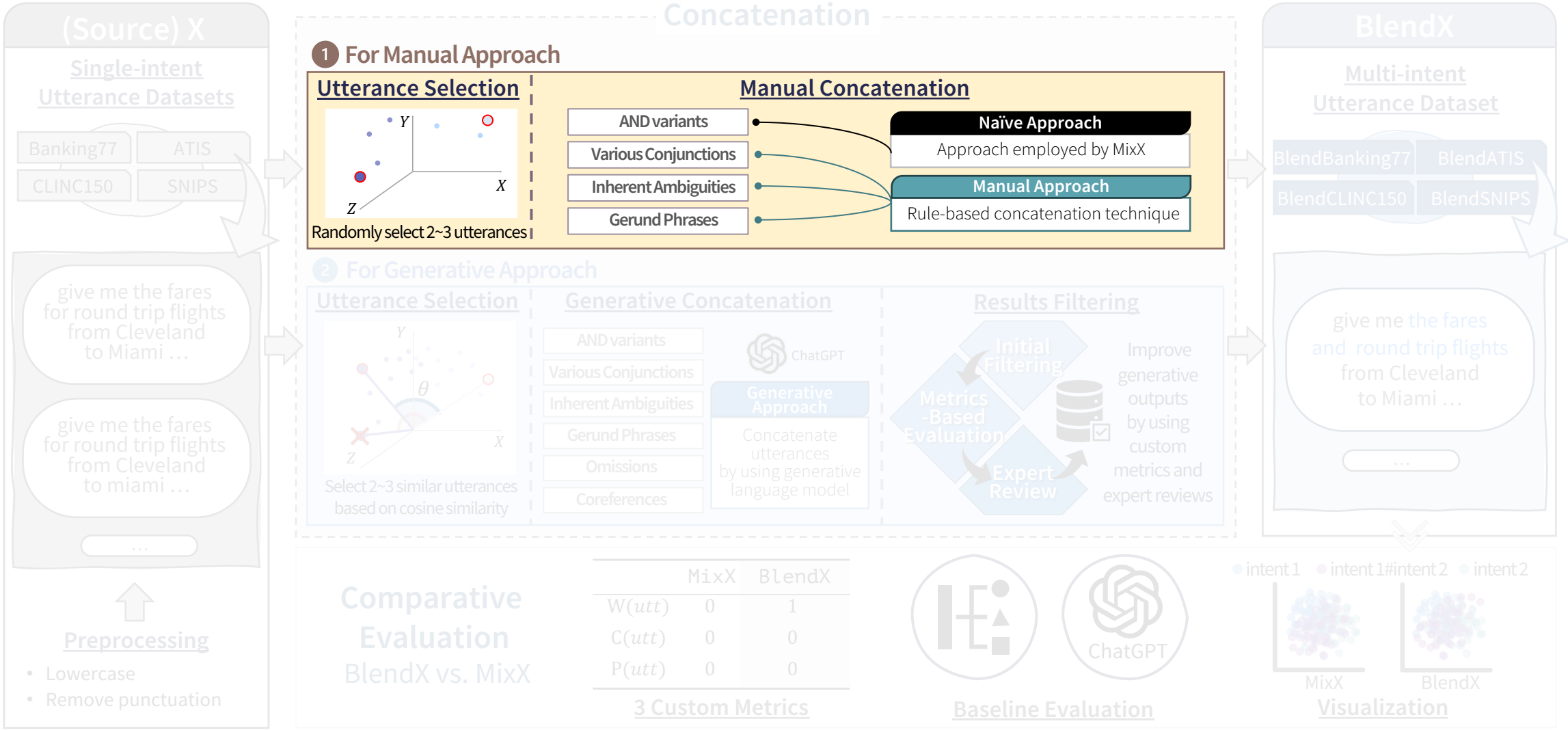
Generative Approach
Concatenate utterances by using generative language model

Results Filtering

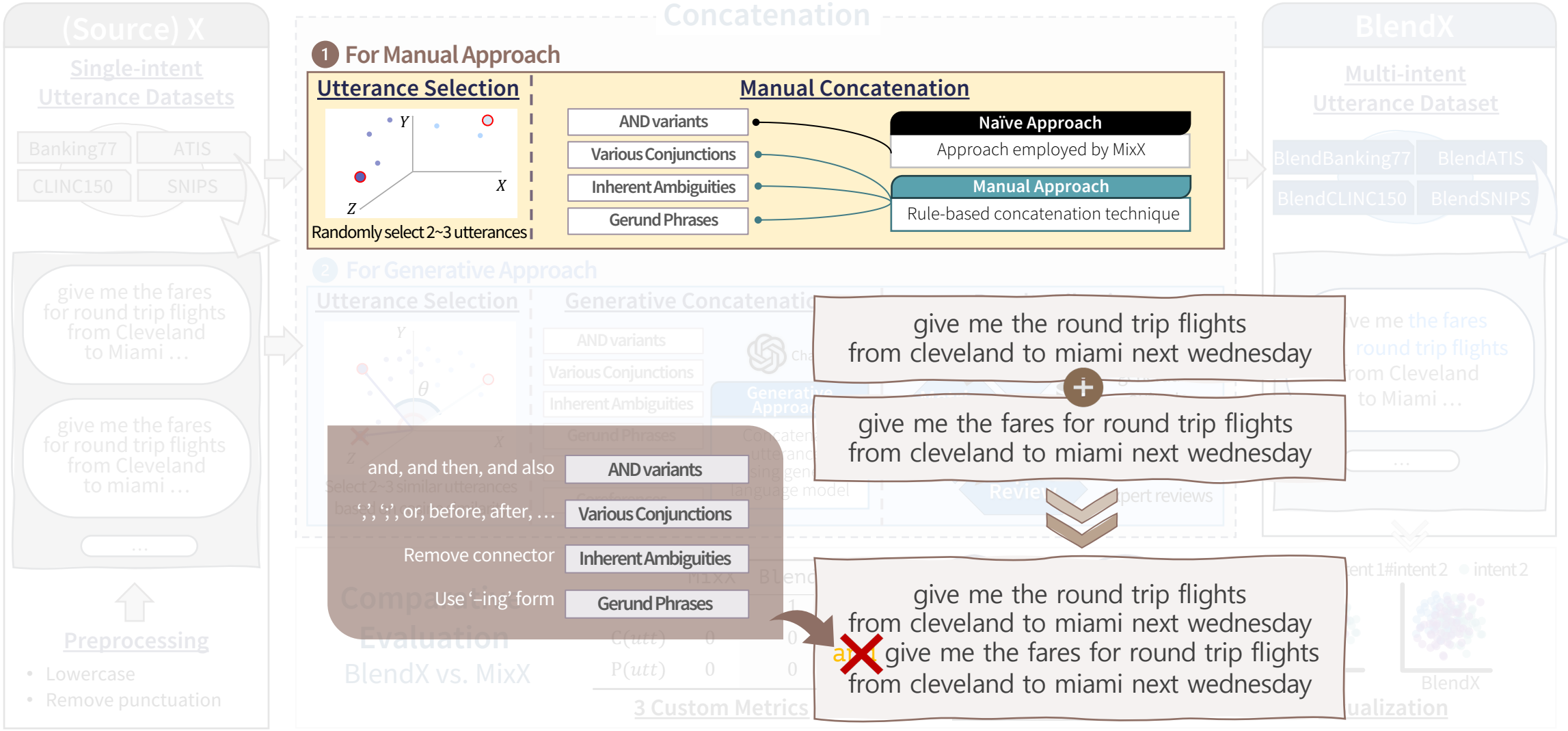


Improve generative outputs by using custom metrics and expert reviews

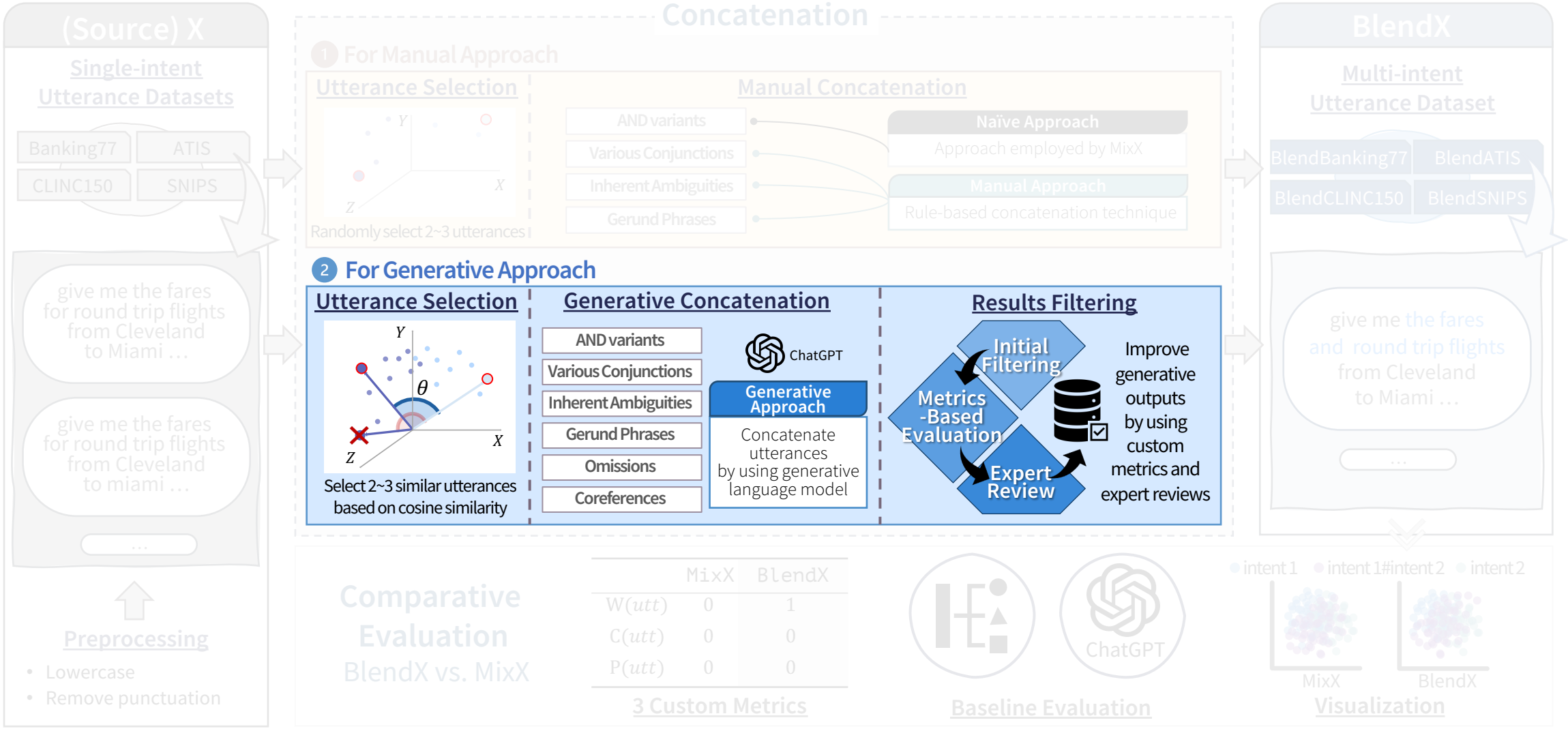
2-1) Manual Approach



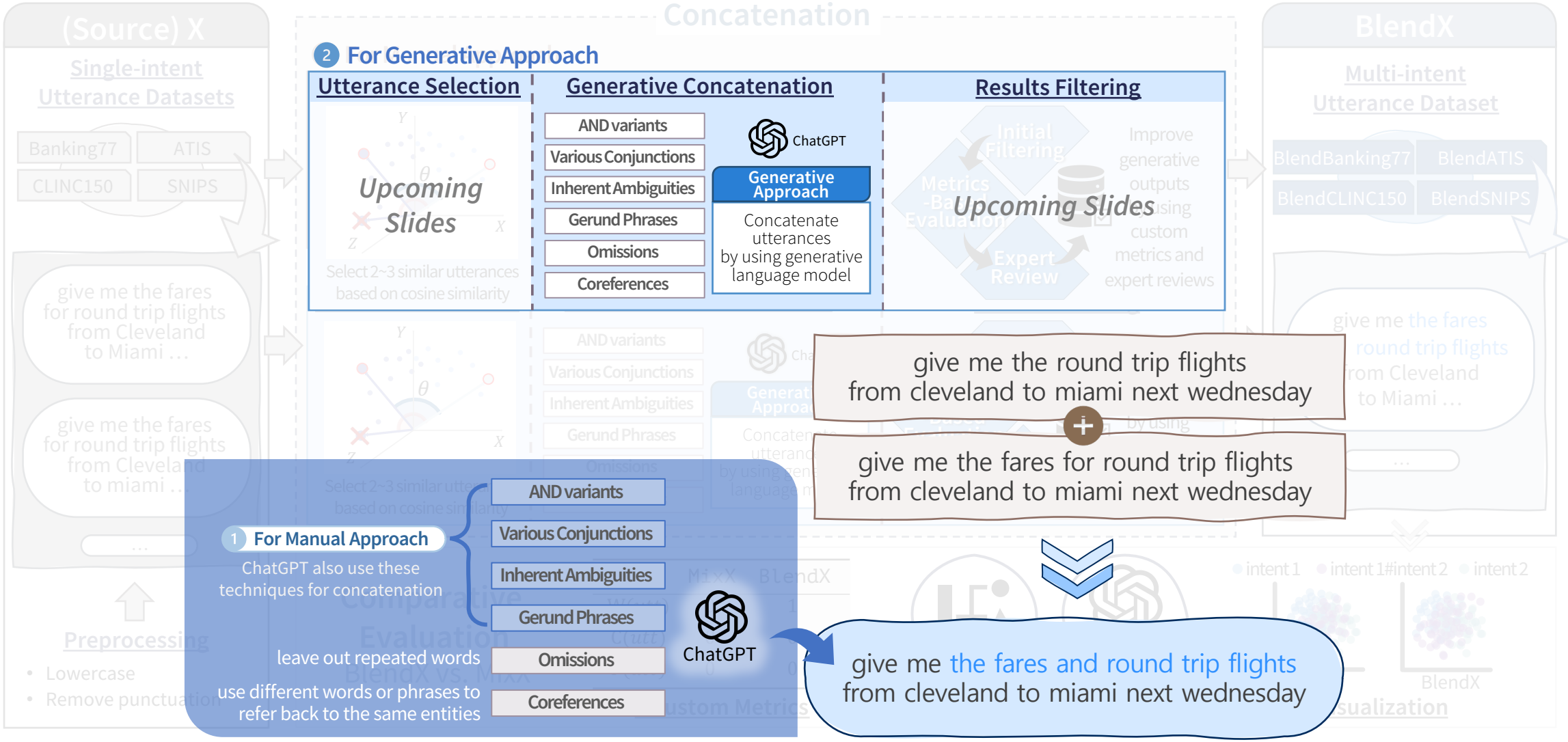
2-1) Manual Approach



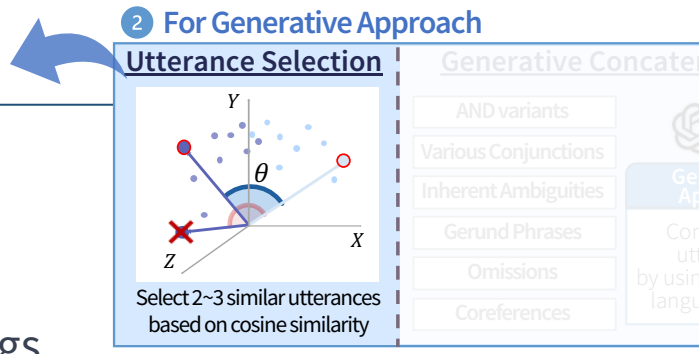
2-2) Generative Approach



2-2) Generative Approach

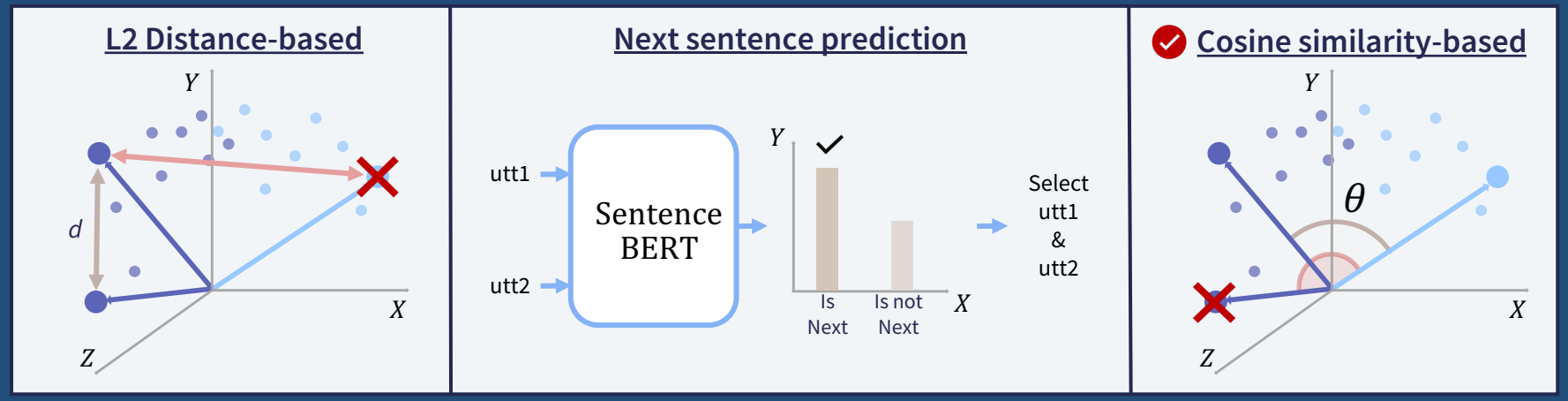


Utterance Selection for **2-2) Generative Approach**



- **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.

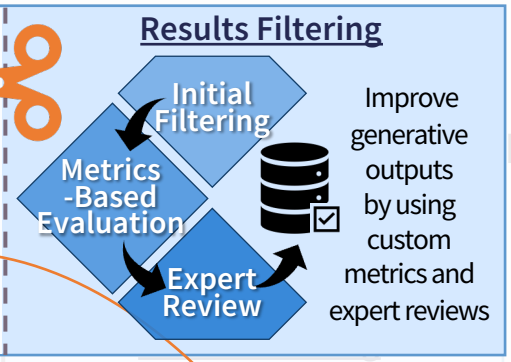
- **Selection approach**
  - L2 Distance-based: Select utterances with close proximity in embedding space.
  - Next sentence prediction: Binary classification of whether a given pair of utterances are sequential.
  - ✓ - **Cosine similarity-based**: Choose utterances with high cosine similarity between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



Results Filtering for **2-2) Generative Approach**

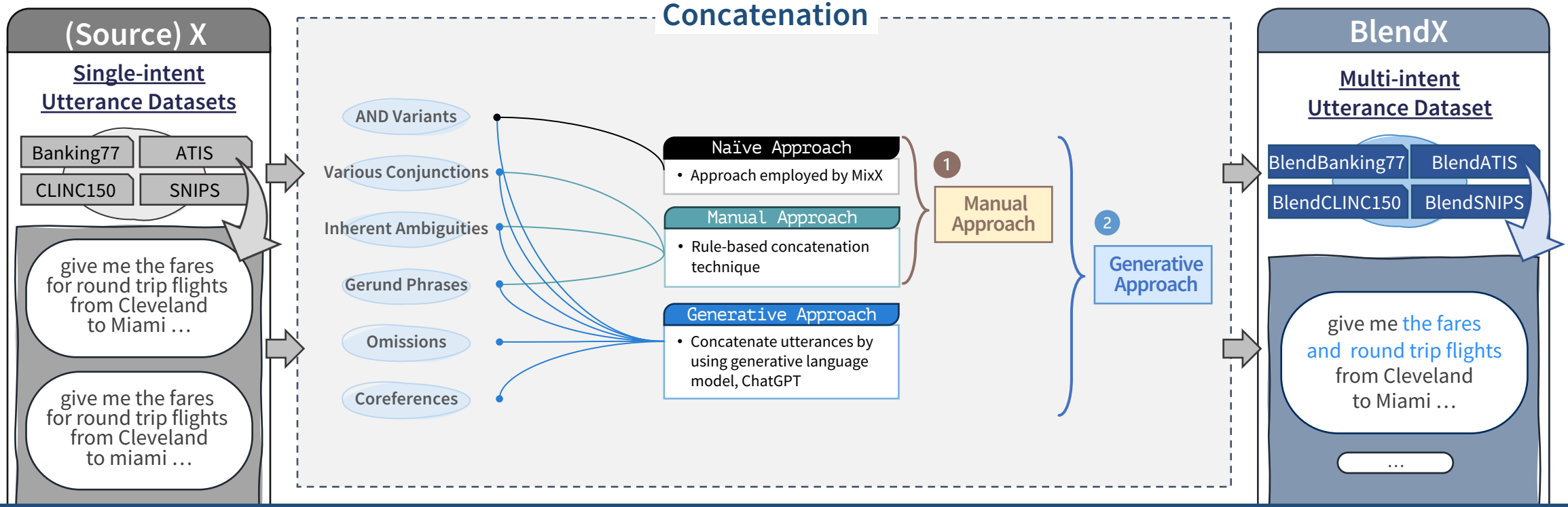
Success	Utterance 1	Utterance 2	Concatenation	Implication
X	weather for frankfort (GetWeather)	3 out of 6 for the last album (RateBook)	frankfort's weather gets a 3 out of 6 rating for the last album	Intent (RateBook) removed
	show me the fares for a first class ticket from baltimore to dallas (atis_airfare)	i would like the time of all flights from san francisco to pittsburgh on sunday (atis_flight_time)	show me the fares for a first class ticket from baltimore to dallas and i would like to know all flights from san francisco to pittsburgh on sunday	Intent changed (from atis_flight_time to atis_flight)
	what is mci (atis_abbreviation)	list la (atis_city)	combine the sentences: "what is mci?" and "list la".	Failed to merge
O	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
	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 and successful results of the Generative Approach and their implications.



1. Remove clear failures generated by ChatGPT
   : explicit mentions of an intent label, unnecessary punctuation, . . .
2. Use metrics to evaluate integration diversity and complexity and collect results above a set threshold.
3. Human-expert review of the generated output filtered by the above two steps
   and remove if a generated utterance deviated significantly from its original intents.

3) BlendX – Multi-intent utterance datasets



BlendX *: Complex Multi-intent Detection with Blended Patterns*

Dataset	Intents #	Training	Dev	Test	Total
SNIPS	7	50,625	2,613	2,615	55,853
ATIS	18	20,250	1,125	1,125	22,500
Banking77	77	36,390	2,009	2,021	40,420
CLINC150	147	54,896	2,889	2,977	60,762

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

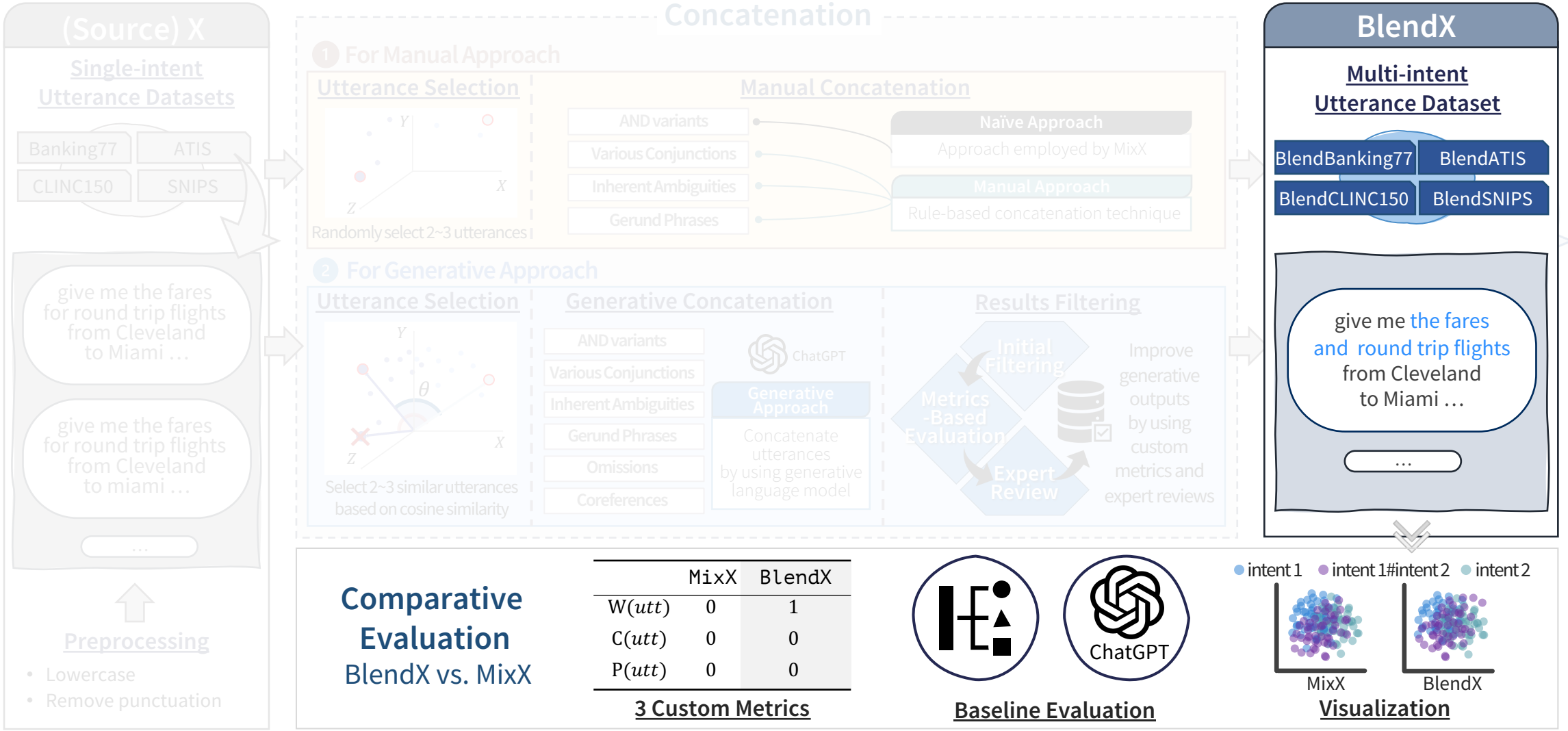
Discussion 2: Experiment and Analysis

3 Custom Metrics Evaluation

SOTA Models Evaluation

Visualization of BlendX

4) Comparative Evaluation – BlendX vs. MixX



Comparative Evaluation
BlendX vs. MixX

	MixX	BlendX
W(utt)	0	1
C(utt)	0	0
P(utt)	0	0

3 Custom Metrics

ChatGPT

Baseline Evaluation

Visualization

4-1) 3 Custom Metrics (1/3)

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

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

Word count

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

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

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

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

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

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.

4-1) 3 Custom Metrics (2/3)

- **Example of applicating 3 metrics**

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.

	Concatenated Utterance	Utterance 1	Utterance 2	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

4-1) 3 Custom 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.

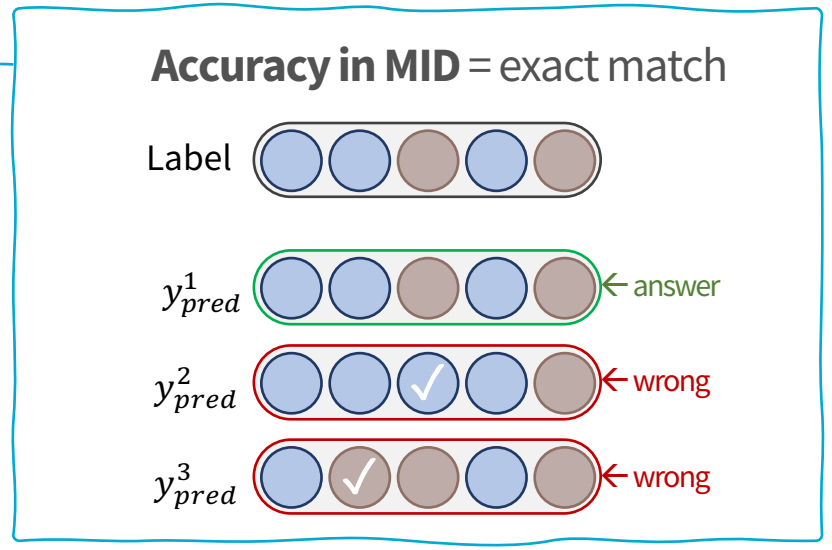
Our approach, incorporating both **manual** and **generative** methods, achieves a more diverse range of explicit and implicit concatenation compared to existing techniques.

- 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.

4-2) SOTA Models with BlendX

- **Evaluate Baseline**

Model	Option		Dataset (Metric: accuracy)			
	Training	Test	SNIPS	ATIS	Banking77	CLINC150
TFMN	MixX	MixX	95.68* ±0.57	77.98* ±0.57	76.61 ±1.17	85.88 ±1.03
	MixX	BlendX	52.51 ±1.86	42.51 ±1.48	37.31 ±0.81	42.45 ±2.40
	BlendX	BlendX	94.93 ±0.85	76.50 ±0.83	63.99 ±0.81	77.96 ±0.82
SLIM	MixX	MixX	95.97* ±0.23	77.10* ±0.28	83.71 ±0.88	88.67 ±0.56
	MixX	BlendX	93.51 ±0.18	72.80 ±1.48	69.89 ±0.46	73.39 ±2.46
	BlendX	BlendX	95.73 ±0.86	76.92 ±0.84	75.30 ±0.71	85.62 ±0.51
gpt-3.5-turbo	-	MixX	81.68	40.30	30.90	49.22
	-	BlendX	76.18	38.84	22.67	37.55

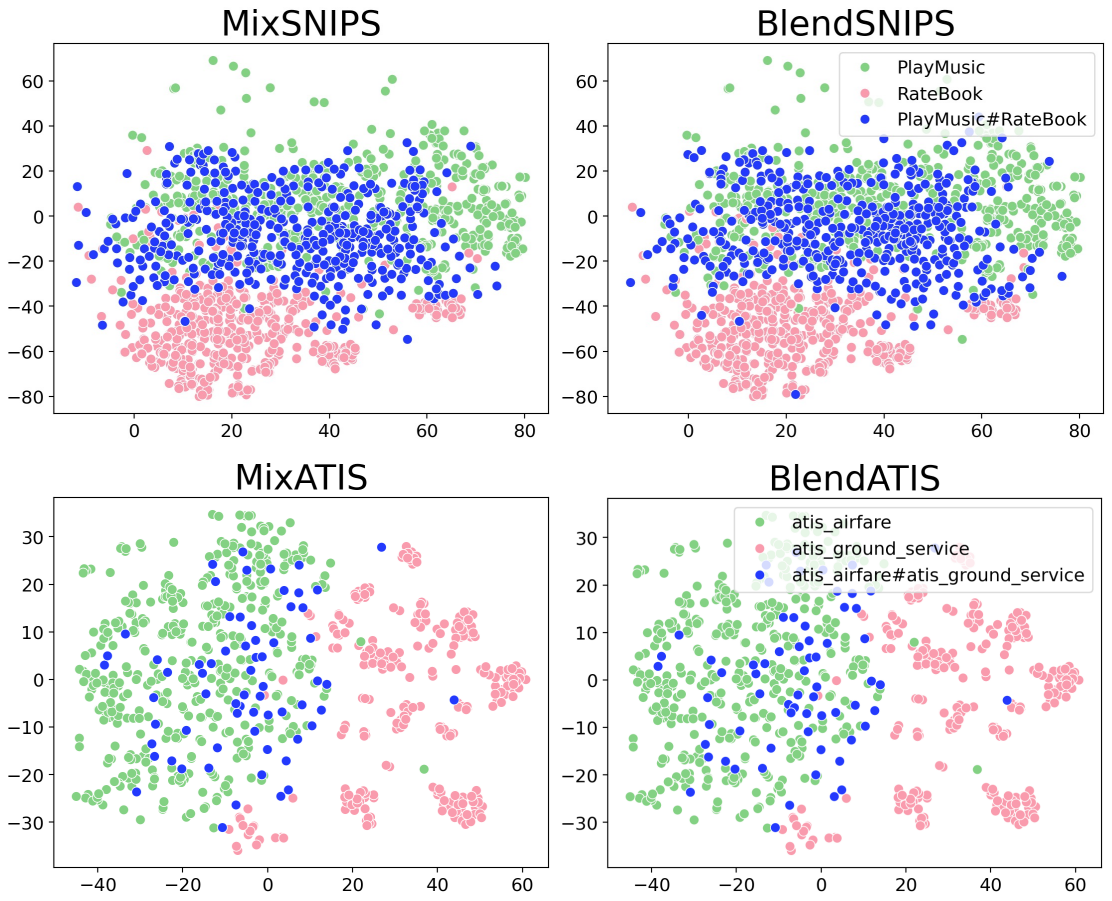


For (un)supervised SOTA models, we consistently observe a performance drop on our **BlendX** datasets with explicit as well as implicit concatenations.

- 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 ($\mathbf{gpt{-}3.5{-}turbo{-}0613}$)

4-3) Visualization of MixX and BlendX

- **Visualization of MixX and BlendX in the $\mathbb{R}^2$ space**
  - datasets
    - MixSNIPS & BlendSNIPS (Top)
    - MixATIS & BlendATIS (Bottom)
  - Techniques
    - Sentence embeddings via SBERT
    - 2-dim. Projection with t − SNE



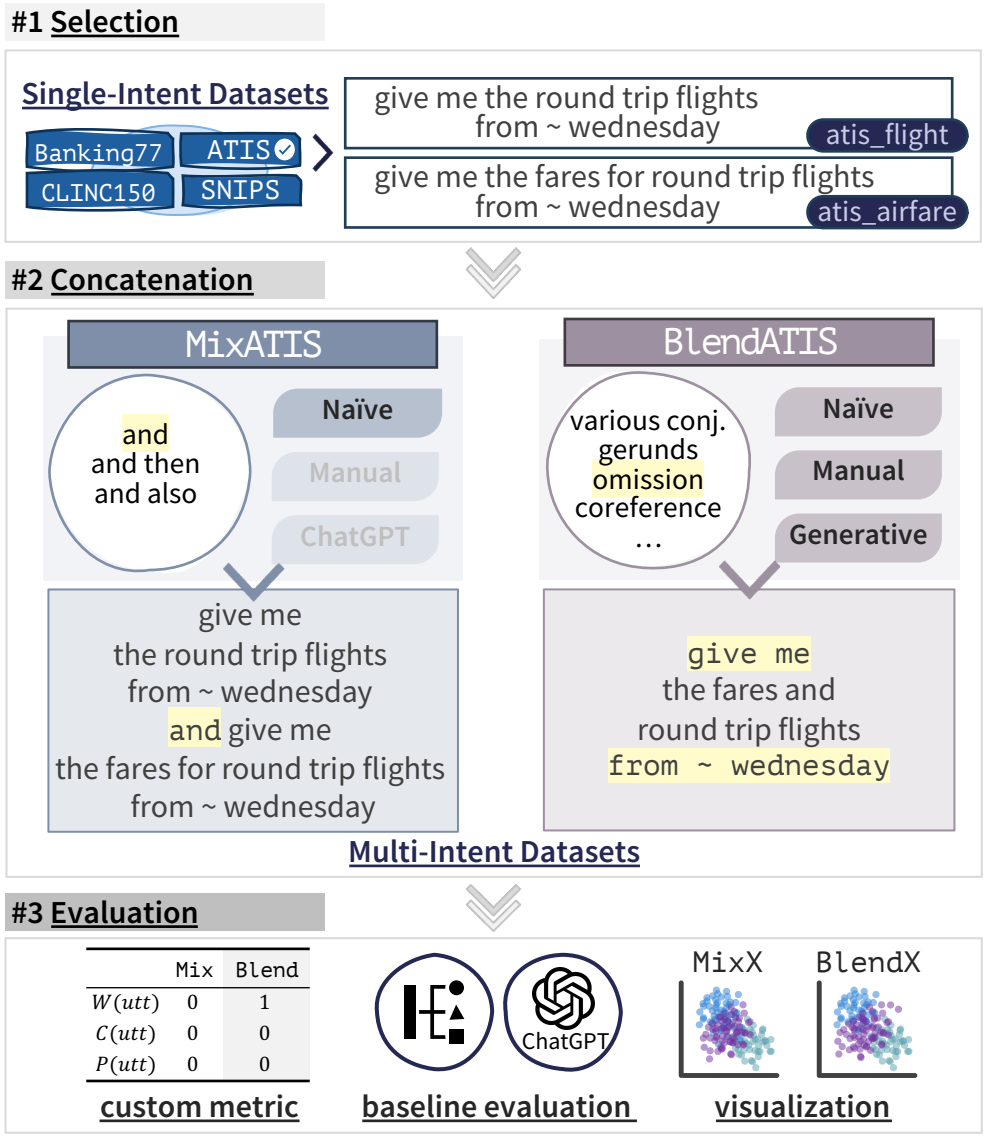
- Both MixX and BlendX visualizations show similar patterns.
- Despite the additional complexity in BlendX, the embeddings map the semantic space without causing significant distortions, capturing all intents effectively.

Conclusion

Contribution

Conclusion - 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.

Dept. of AI Application
Hanyang University, South Korea

stillwithyou@hanyang.ac.kr