

Data Augmentation for Conversational AI

The Web Conference 2024



Tutorial website

Presenters



Heydar Soudani

PhD Candidate
Radboud University
heydar.soudani@ru.nl



Evangelos Kanoulas

Full Professor
University of Amsterdam
e.kanoulas@uva.nl



Roxana Petcu

PhD Candidate
University of Amsterdam
r.m.petcu@uva.nl



Faegheh Hasibi

Assistant Professor
Radboud University
f.hasibi@cs.ru.nl

Part 4: Conversation Generation - Information Seeking

Duration: 45 min

Presenter: Heydar Soudani

Conversational Information Seeking (CIS)

- The main goal of CIS is fulfill users' information needs
- Allow users to search information using natural language dialogue, instead of traditional search queries

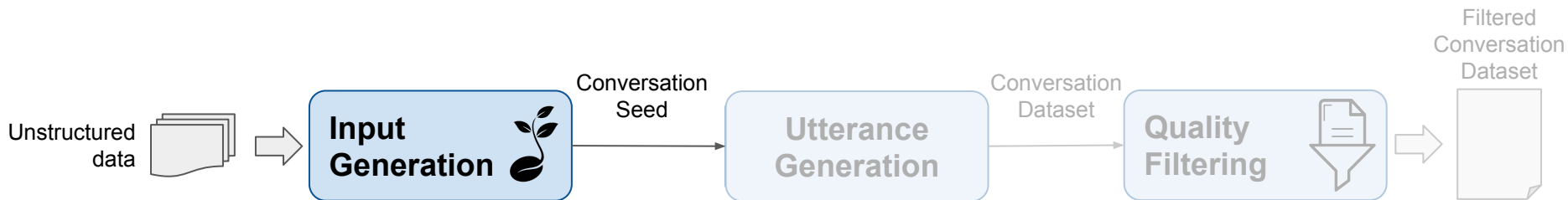
(Zamani et al., 2023)

Key Features of CIS

- Generation control, Topic shifting
- Multi-evidence answer generation
- Query ambiguity, asking clarification questions

(Wu et al., 2022), (Deng et al., 2023)

CIS Data Generation



Example

(Askari et al., 2024)



MSDialog
Intents:

CQ
FD
GG
PA
IR
OQ



Conversation Seed

Entity type: Person

Entity type attribute: Occupation

Entity name: Albert Einstein

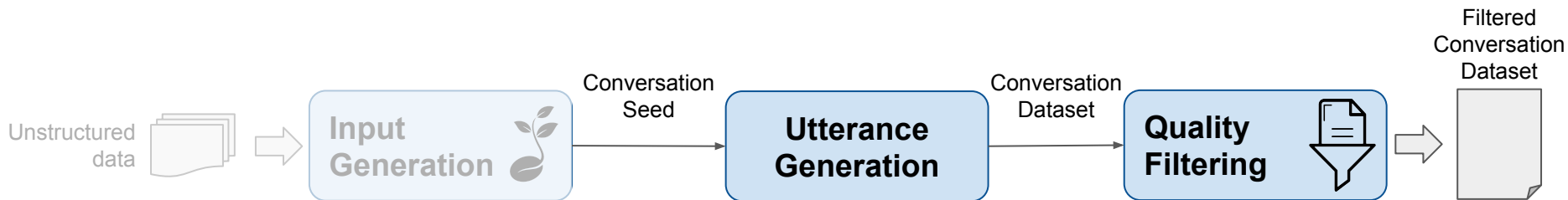
Entity background document: Albert Einstein was a German-born theoretical physicist who is ...

Conversation starter: Can you delve into the efforts and contributions of Albert Einstein in the field of physics?

Dialogue flow: [original question, clarifying question and information request, further details]

(Qu et al., 2018)

CIS Data Generation



Example

(Askari et al., 2024)



MSDialog
Intents:
CQ
FD
GG
PA
IR
OQ

Conversation Seed

Entity type:

Entity type attribute:

Entity name:

Entity background document: ...

Conversation starter: ...

Dialogue Flow: [original question, clarifying question and information request, further details]

Conversation Sample

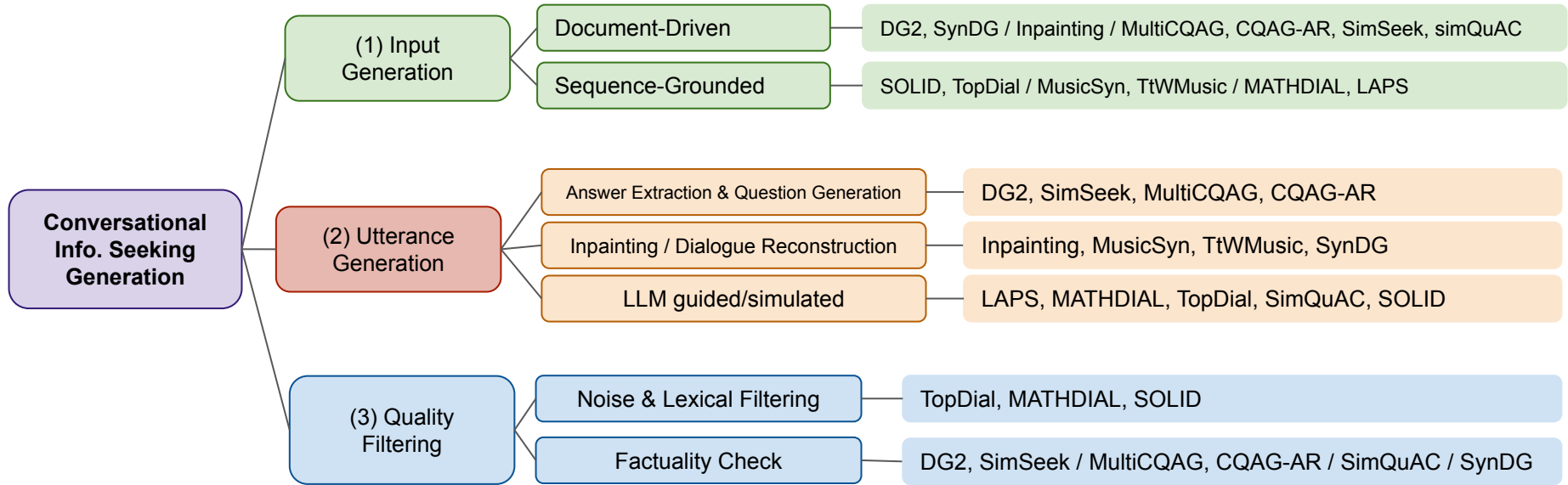
User: Can you delve into the specific efforts and contributions made by Albert Einstein in the field of physics? (*intent: original question*)

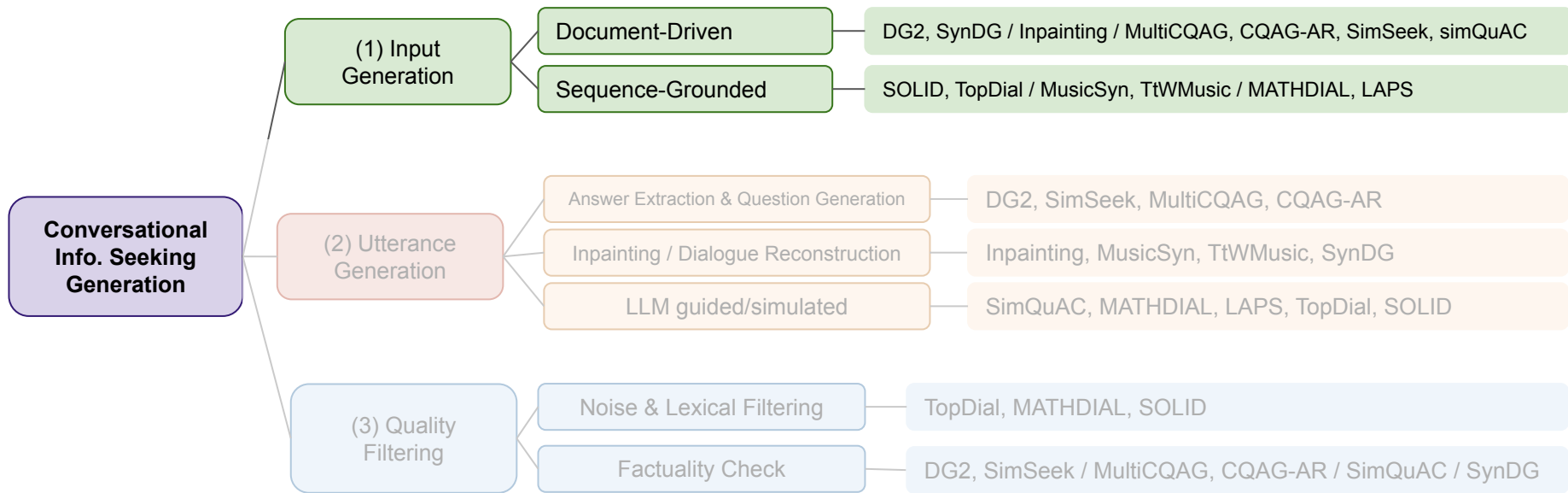
Agent: Sure! Albert Einstein made groundbreaking contributions to physics, especially with his theory of relativity. What aspect would you like me to focus on or any specific topic you're interested in? (*intent: clarifying question*)

User: Could you provide more details about his theory of relativity and how it revolutionized our understanding of space and time? (*intent: further details*)

...

(Qu et al., 2018)





Input Generation

What does “Conversation Seed” contain?

- Information containing a main topic, subtopics, and key details about the topic
- **Dialogue Flow**: a comprehensive perspective of the conversation

Example from Doc2Dial dataset

(Feng, et al., 2020)

Conversation Seed

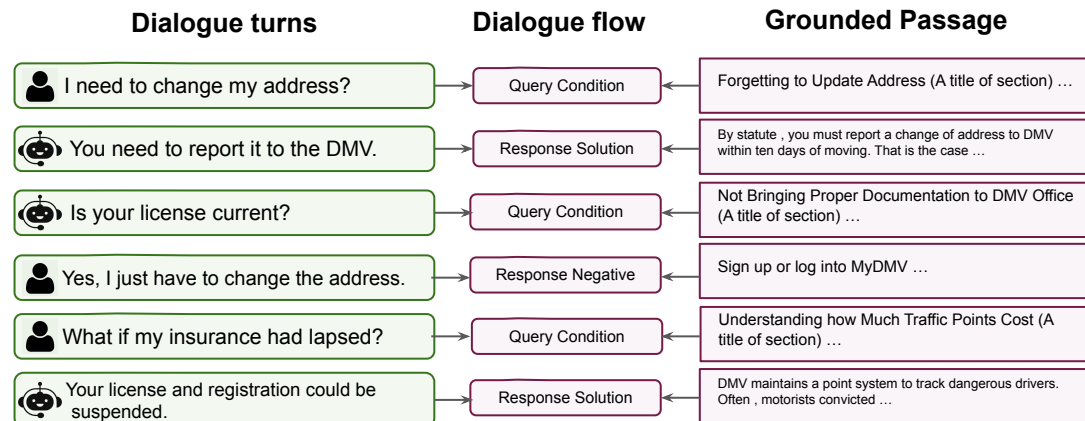
Title: Top 5 DMV Mistakes and How to Avoid Them

Document: Many DMV customers make easily avoidable mistakes that cause them significant problems, including ...

Dialogue flow: [Query_condition, Respond_solution, Query_condition, Response_negative, query_condition, ...]



Conversation Sample



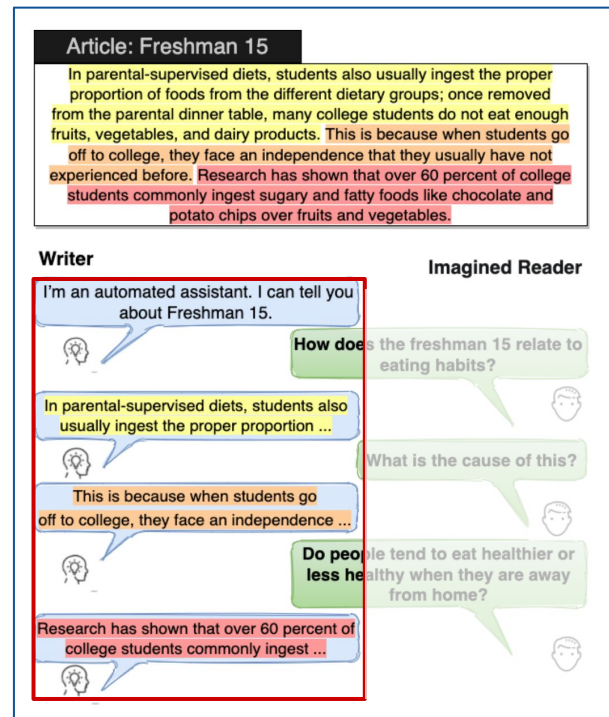
Input Generation - Document-Driven

- Why are documents used for CIS data generation?

Inpainting

- **Idea:** Documents are conceptualized as dialogues between the writer and an imaginary reader
- The dialogue flow consists directly of the document's sentences

(Dai et al., 2022)



Input Generation - Document-Driven

Document Segmentation

- A document is segmented into multiple passages
- Passage Ranker

$$p(\underbrace{c_t}_{\text{Selected passage in turn } t} \mid \underbrace{\{u_i, a_i\}_{i < t}}_{\text{Conversation History}}, \underbrace{C}_{\text{Document}})$$

- Not fixed and pre-defined
- Dialogue flow: a sequence of passages
- May not consist of sequential passages from a document

(Wu et al., 2022)

Top 5 DMV Mistakes and How to Avoid Them

<Passage 1> Many DMV customers make easily avoidable mistakes that cause them significant problems, ...

<Passage 2>

<Passage 3> Not Bringing Proper Documentation to DMV Office. About ten percent of customers visiting a DMV office do not bring what they need to complete their transaction and see if your transaction can be ...

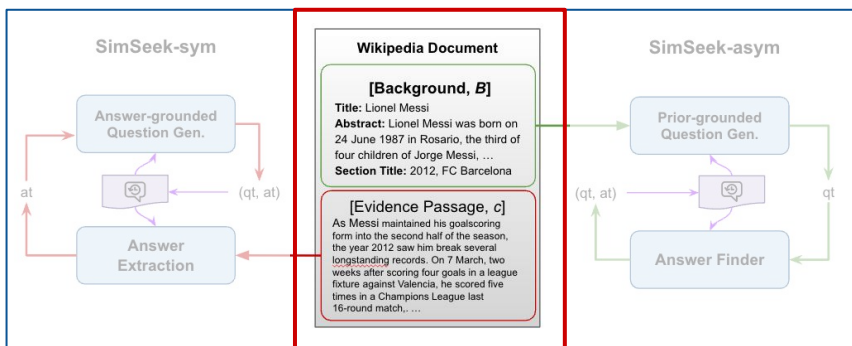
<Passage 4>

<Passage 5> We send 500,000 inquiry letters a year. If the inquiry letter does not resolve the problem, we must suspend the vehicle registration and, if it persists, your driver license!
We suspend 300,000 registrations a year for failure to maintain insurance. ...

Input Generation - Document-Driven

Whole Document

- Input a document or provide background information, and leave it to the "utterance generation" component to decide which part of the document to use



(Kim et al., 2022)

Title : Esports Section Title : History Early history (1972-1989)

Document c

The earliest known video game competition took place on 19 October 1972 at Stanford University for the game "Spacewar". Stanford students were invited to an "Intergalactic spacewar olympics" whose grand prize was a year's subscription for "Rolling Stone", with Bruce Baumgart winning the five-man-free-for-all tournament and Tovar and Robert E. Maas winning the Team Competition. The Space Invaders Championship held by Atari in 1980 was the earliest large scale video game competition, attracting more than 10,000 participants across the United States, establishing competitive gaming as a mainstream hobby. ...

Background B

Esports (also known as electronic sports, e-sports, or eSports) is a form of competition using video games. Most commonly, esports takes the form of organized, multiplayer video game competitions, particularly between professional players, individually or as teams. Although organized online and offline competitions have long been a part of video game culture, these were largely between amateurs until the late 2000s, when participation by professional gamers and spectatorship in these events through live streaming saw a large surge in popularity. By the 2010s, esports was a significant factor in the video game industry, with many game developers actively designing toward a professional esports subculture.

Conversation in WIKI-SIMSEEK

q₁ : What is the history of esports?

a₁ : The earliest known video game competition took place on 19 October 1972 at Stanford University for the game "Spacewar".

q₂ : What was the result of this competition?

a₂ : Bruce Baumgart winning the five-man-free-for-all tournament and Tovar and Robert E. Maas winning the Team Competition.

q₃ : Did esports grow from there?

a₃ : The Space Invaders Championship held by Atari in 1980 was the earliest large scale video game competition, attracting more than 10,000 participants across the United States.

q₄ : What happened after the Space Invaders Championship?

...

Input Generation - Sequence-Grounded

Fixed Sequence

1) Topics with their Background knowledges

- Select / Generate a topic

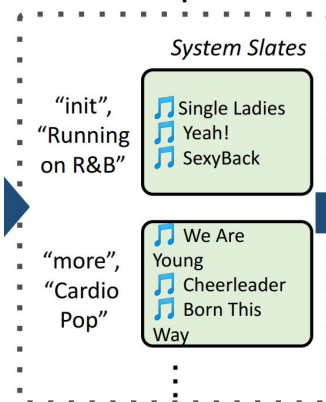
2) A sequence of dialogue acts

- Sampling a valid sequence

Item Collections



Slate Sequences



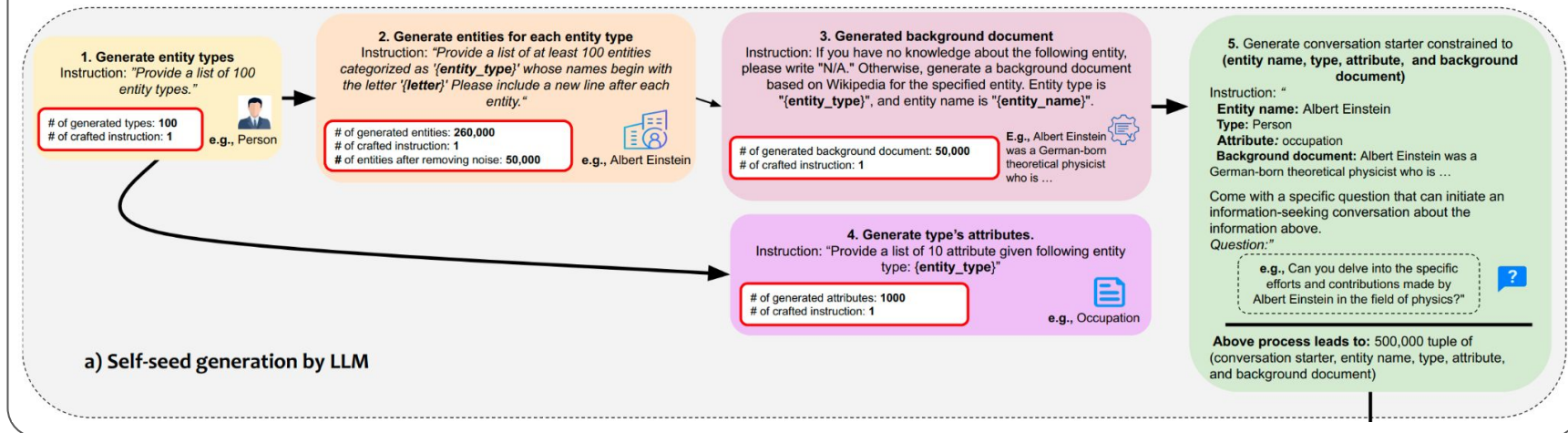
(Leszczynski et al., 2023)

Input Generation - Sequence-Grounded

(Askari et al., 2024)

Generate the background information

- Why Generation instead of Selection? Quality / Diversity
- Self-seeding approach, Prompt LLM to generate everything



Input Generation - Sequence-Grounded

Dialogue Acts - Fixed

- Main feature: validity
- Make conversation real, maintain the consistency
- How to ensure the validity? Using existing crowdsourcing dialogue datasets

SOLID: Full path
Used MSDialog-intent

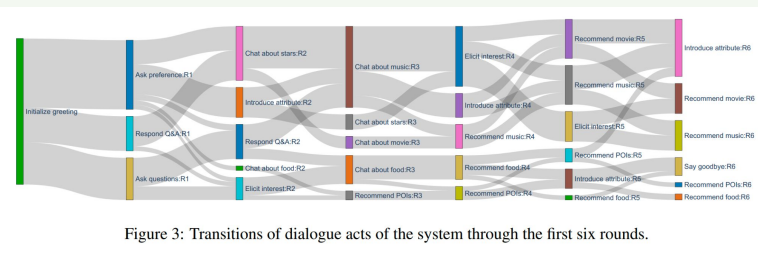
(Askari et al., 2024)

Table 7: Intent taxonomy and distribution in MSDialog

CodeLabel	Description	%
OQ	Original Question The first question from the user to initiate the dialog.	13
RQ	Repeat Question Other users repeat a previous question.	3
CQ	Clarifying Question User or agent asks for clarifications.	4
FD	Further Details User or agent provides more details.	14
FQ	Follow Up Question User asks for follow up questions about relevant issues.	5
IR	Information Request Agent asks for information from users.	6
PA	Potential Answer A potential answer or solution provided by agents.	22
PF	Positive Feedback User provides positive feedback for working solutions.	6
NF	Negative Feedback User provides negative feedback for useless solutions.	4
GG	Greetings/Gratitude Greetings or expressing gratitude.	22
JK	Junk No useful information in the utterance.	1
O	Others Utterances cannot be categorized using other classes.	1

TopDial: partial path
starting point and the target (act-topic)
Used DuRecDial 2.0

(Wang et al., 2023)



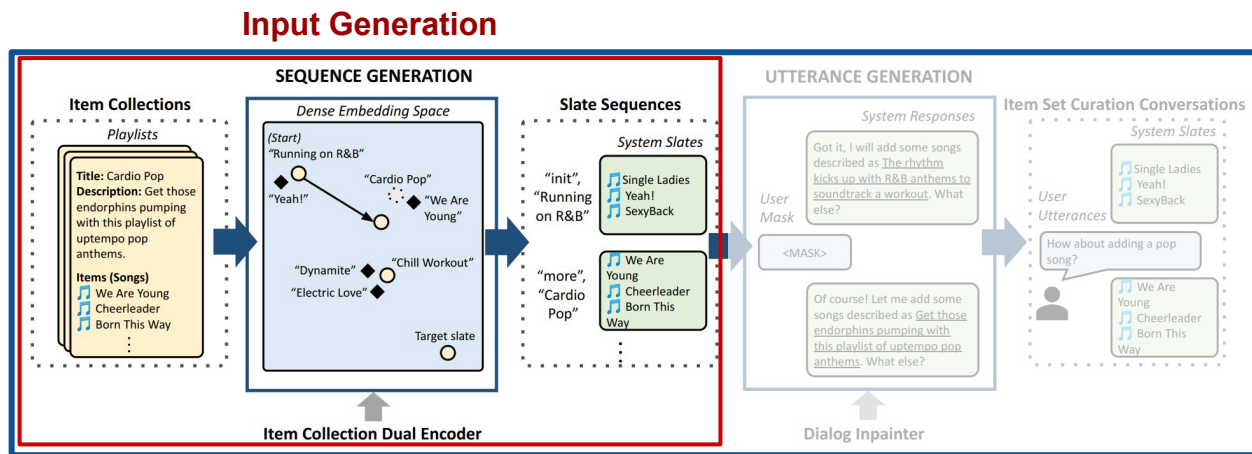
Input Generation - Sequence-Grounded

Dialogue Acts - Fixed

(Leszczynski et al., 2022)

(Leszczynski et al., 2023)

- How to ensure the validity? Closeness in embedding space
- Example: Walk the Talk

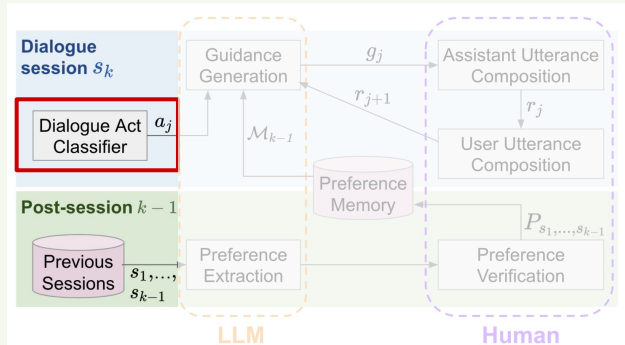


Input Generation - Sequence-Grounded

Dialogue Acts - Open

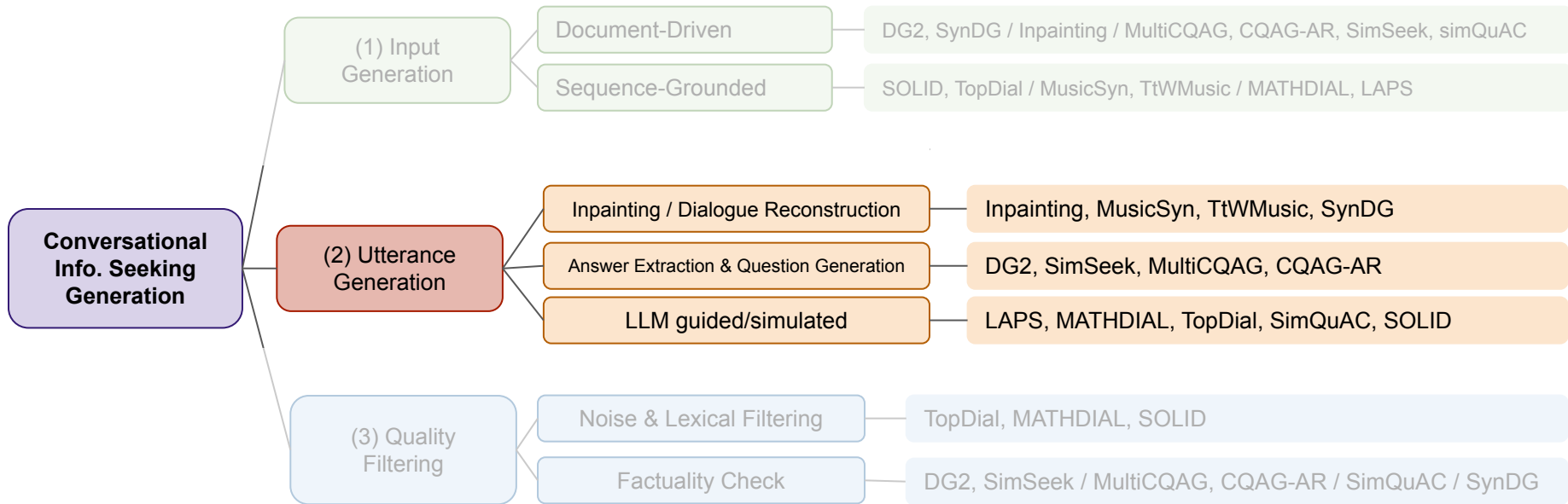
- Used in Human-AI collaboration based methods
- Dialogue act is predicted
 - Based on Dialogue history
 - Before the current turn is generated

LAPS: LLM classifier (Joko et al., 2024)



MathDial: Human selects (Macina et al., 2023)

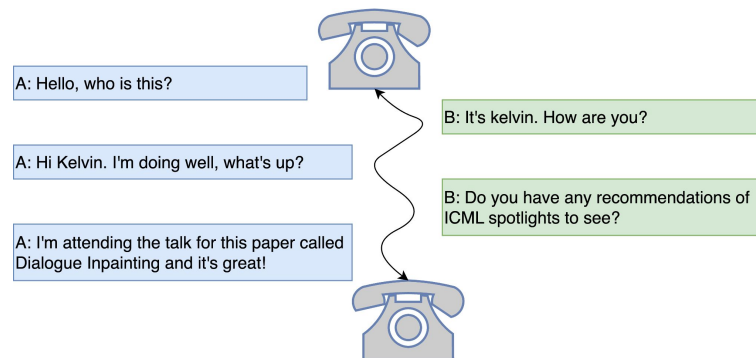
Category	Intent	Example
Focus	Seek Strategy	So what should you do next?
	Guiding Student Focus	Can you calculate ... ?
	Recall Relevant Information	Can you reread the question and tell me what is ... ?
Probing	Asking for Explanation	Why do you think you need to add these numbers?
	Seeking Self Correction	Are you sure you need to add here?
	Perturbing the Question	How would things change if they had ... items instead?
	Seeking World Knowledge	How do you calculate the perimeter of a square?
Telling	Revealing Strategy	You need to add ... to ... to get your answer.
	Revealing Answer	No, he had ... items.
Generic	Greeting/Fairwell	Hi ... , how are you doing with the word problem? Good Job! Is there anything else I can help with?
	General inquiry	Can you go walk me through your solution?



Utterance Generation - Inpainting

(Dai et al., 2022)

- **Reminder:** Dialogue Flow -> Directly the document's sentences
- **Idea:** Fine-tuning a model to reconstruct a dialogue
- **Real world Motivation:**
 - Overhearing someone else's phone call
 - Hear on side, try to guess another side



- **Task:** Take a partial dialog \implies Generate a complete dialog

$$(u_1, u_2, \diamond, u_4, \diamond) \implies d = (u_1, u_2, \dots, u_t, \dots, u_T)$$

Utterance Generation - Inpainting

(Dai et al., 2022)

Training: Dialog reconstruction

- Randomly mask one utterance (u_t)
- Train a generative model to predict the masked utterance
- Similar to the masked language modeling task used by BERT

$$d_{m(t)} = (u_1, \dots, u_{t-1}, \diamond, u_{t+1}, \dots, u_T)$$

$$p_{\theta}(u_t \mid d_{m(t)})$$

Inference: Transforming documents into dialogues

- Convert document into spans (e.g., sentences)
- Autoregressively generate utterances

$$(s_{\text{prompt}}, \diamond, s_1) \implies \hat{u}_1$$

$$(s_{\text{prompt}}, \hat{u}_1, s_1, \diamond, s_2) \implies \hat{u}_2$$

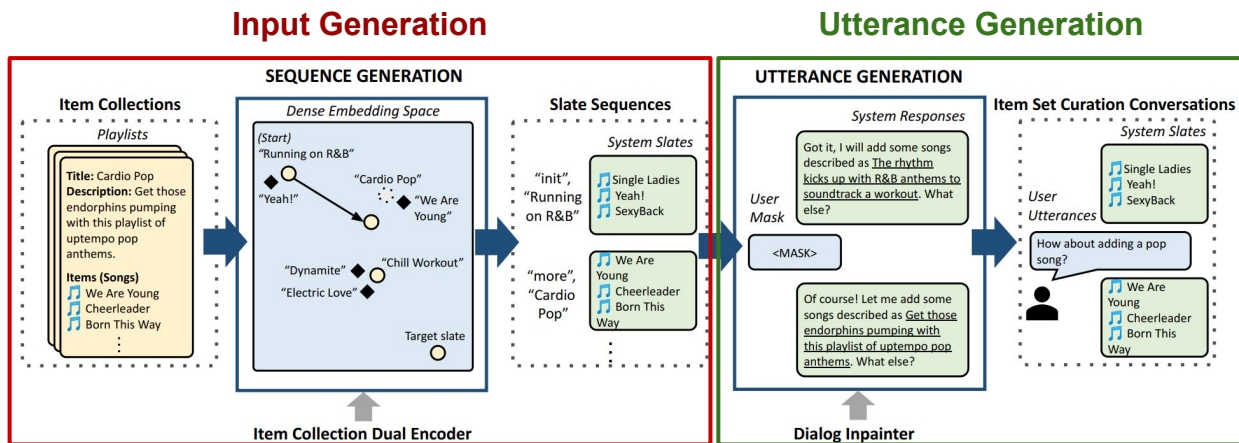
Utterance Generation - Inpainting

Another example of Inpainting

(Leszczynski et al., 2022)

(Leszczynski et al., 2023)

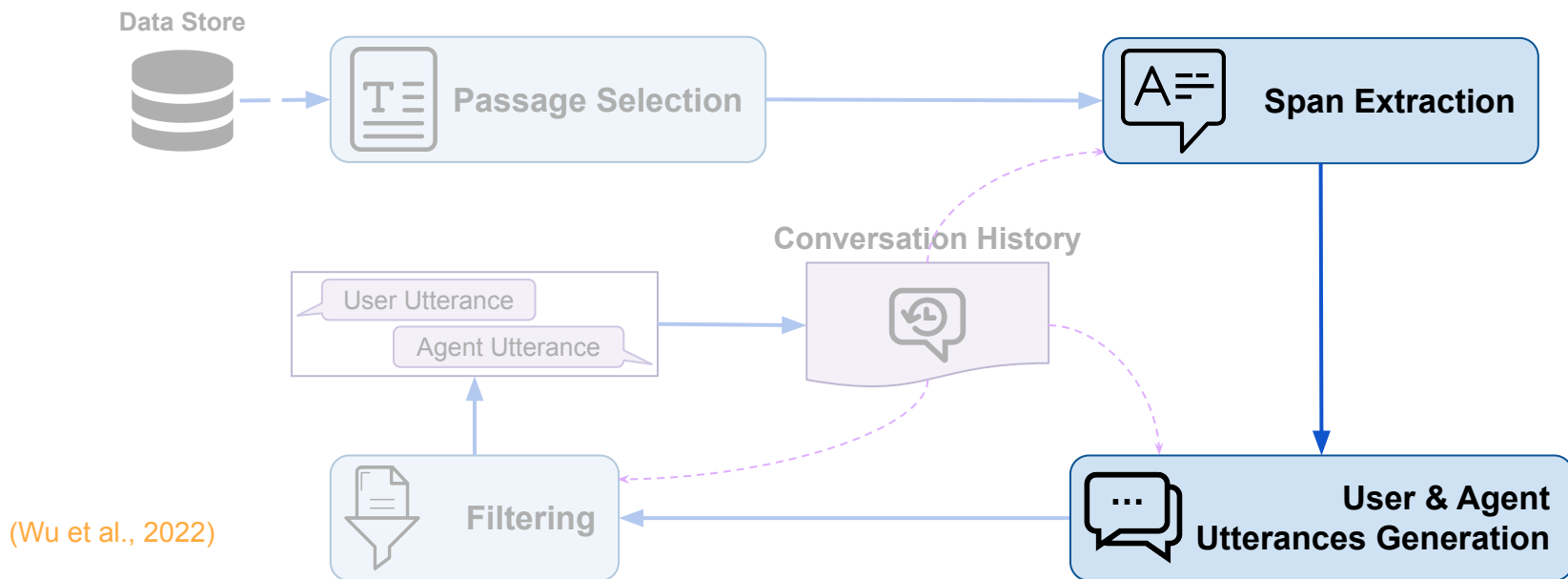
- **Reminder:** Dialogue Flow -> Slate (playlist) sequences



Utterance Generation - Answer Extraction & Question Generation

Reminder: Dialogue Flow -> not fixed, passages Passage Ranker

- The extended version of pipeline approach for “single-turn QA pair generation” (Alberti et al., 2019)



Utterance Generation - Answer Extraction & Question Generation

Answer/Span Extraction (Wu et al., 2022)

<Passage 3> Not Bringing Proper Documentation to DMV Office. About ten percent of customers visiting a DMV office do not bring what they need to complete their transaction and see if your transaction can be ...

Conversation History



Highlights the rationale span used to generate the dialogue turn

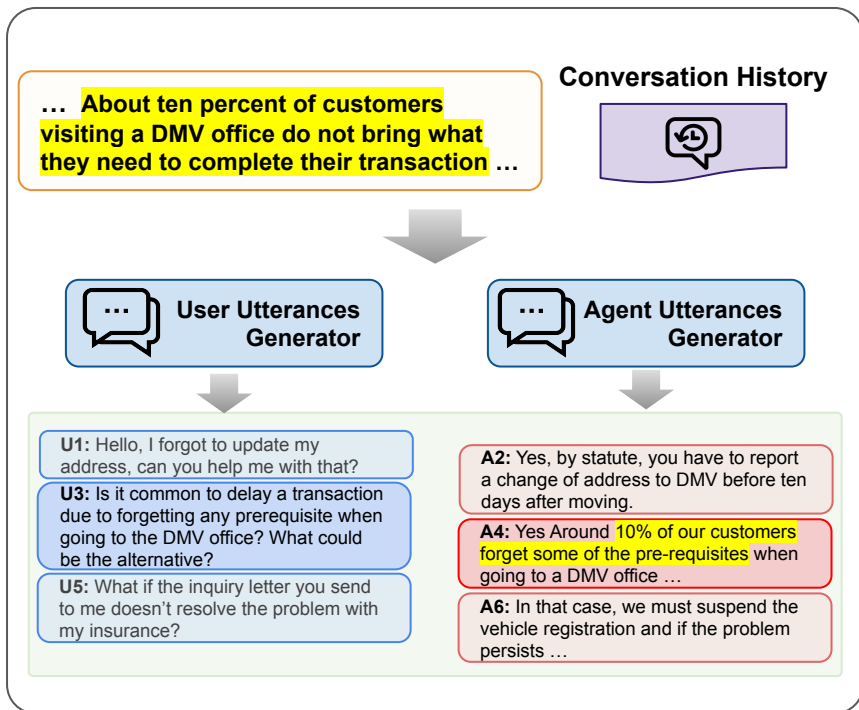
<Passage 3> Not Bringing Proper Documentation to DMV Office. **About ten percent of customers visiting a DMV office do not bring what they need to complete their transaction** and see if your transaction can be ...

Extract a rationale span from the selected passage

$$p(r_t | \underbrace{\{u_i, a_i\}_{i < t}}_{\text{Conversation History}}, \overbrace{c_t}_{\text{Selected passage in turn } t})$$

Utterance Generation - Answer Extraction & Question Generation

User & Agent Utterance Generation (Wu et al., 2022)



User utterance generator

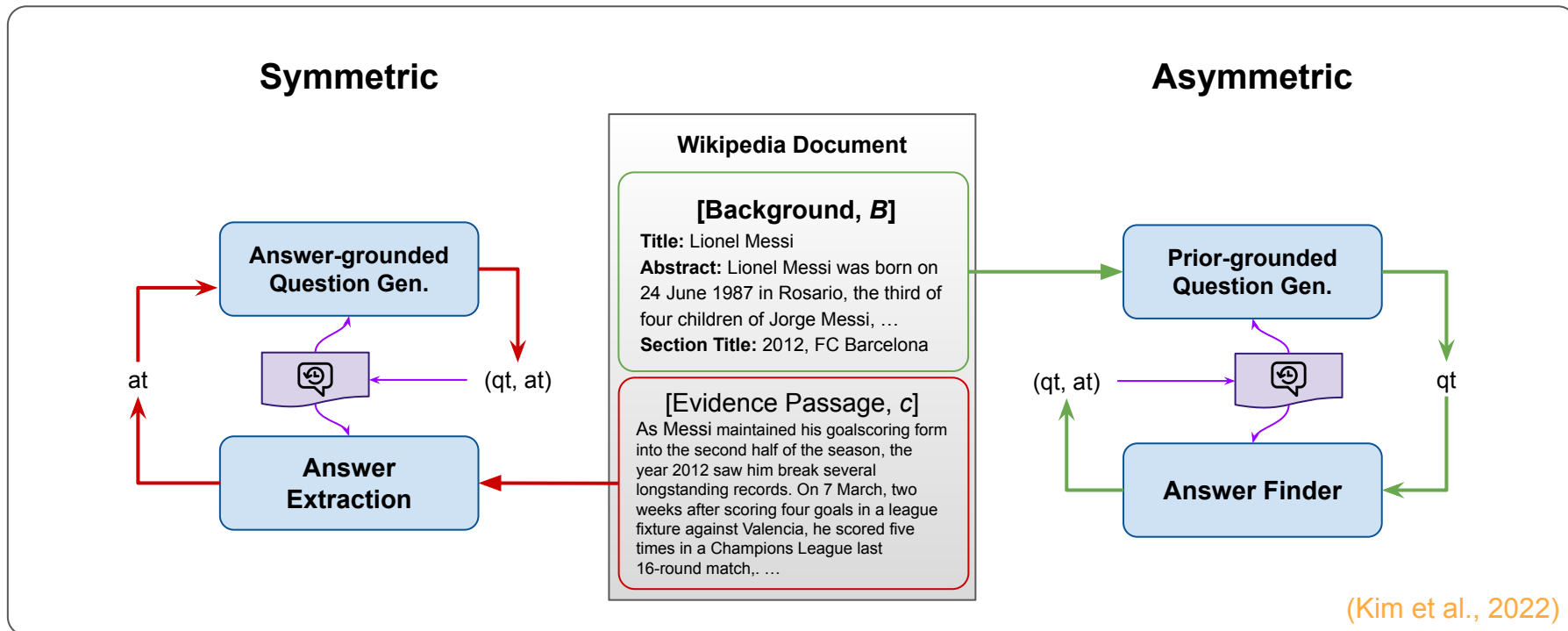
- Generates a question with the answer span
- Highlight the rationale span by wrapping its text

$$p(u_t) = p(u_t | \underbrace{\{u_i, a_i\}_{i < t}}_{\text{Conversation History}}, \underbrace{c'_t}_{\text{Selected passage in turn } t})$$

Agent utterance generator

- Generates the response with the answer span
- The dialogue history now includes the previous generated user utterance

Utterance Generation - Answer Extraction & Question Generation



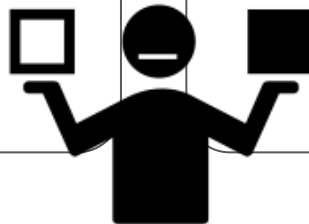
Utterance Generation - Answer Extraction & Question Generation

Symmetric

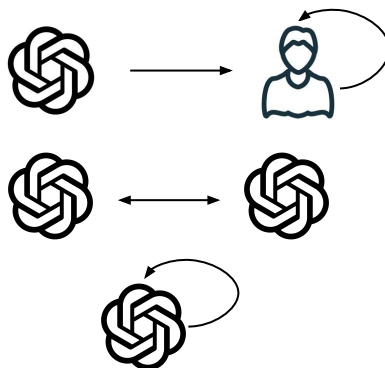
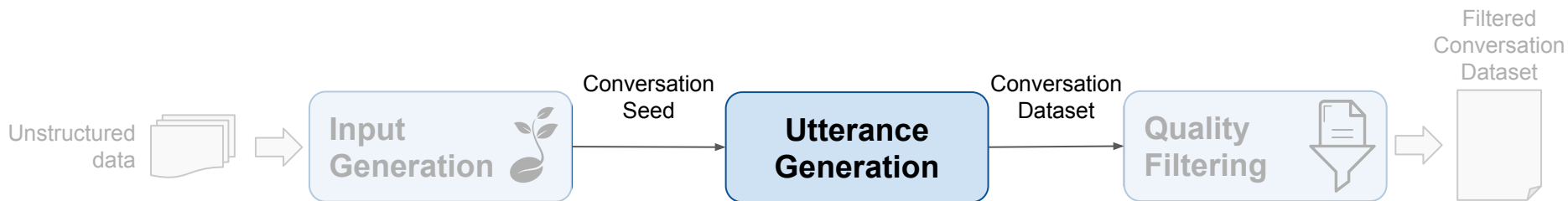
- First extracts an answer candidate from the passage
- Questioner can access all answer-relevant information
 - **Pro:** Coherency with answer
 - **Con:** Constraint to the predetermined answer

Asymmetric

- First asks a question without accessing an answer or passage
- Questioner asks any questions relevant to the topic, guessing inaccessible passage
 - **Pro:** encouraging information-seeking behaviour



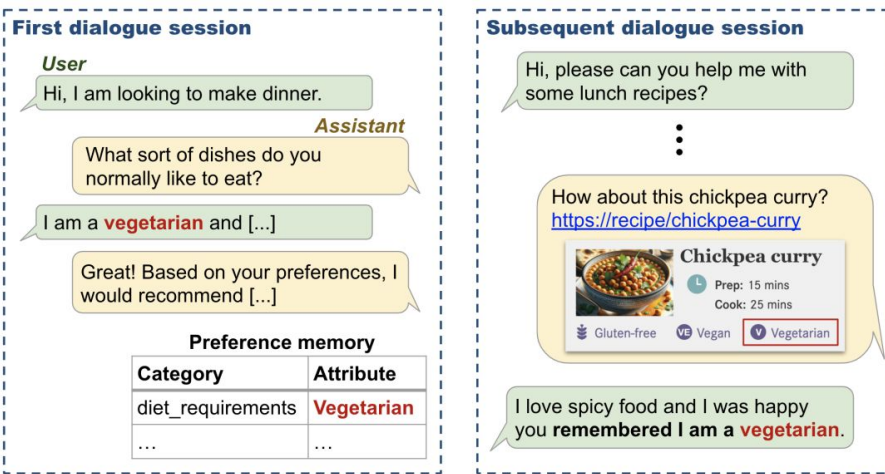
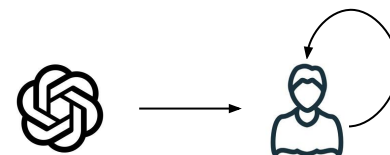
Utterance Generation - LLM guided/simulated



Utterance Generation - LLM guided/simulated

LLM-Guided Human Generation (LAPS) (Joko et al., 2024)

- Task: personalized multi-session dialogue



Utterance Generation - LLM guided/simulated

LLM-Guided Human Generation (LAPS) (Joko et al., 2024)

Main components:

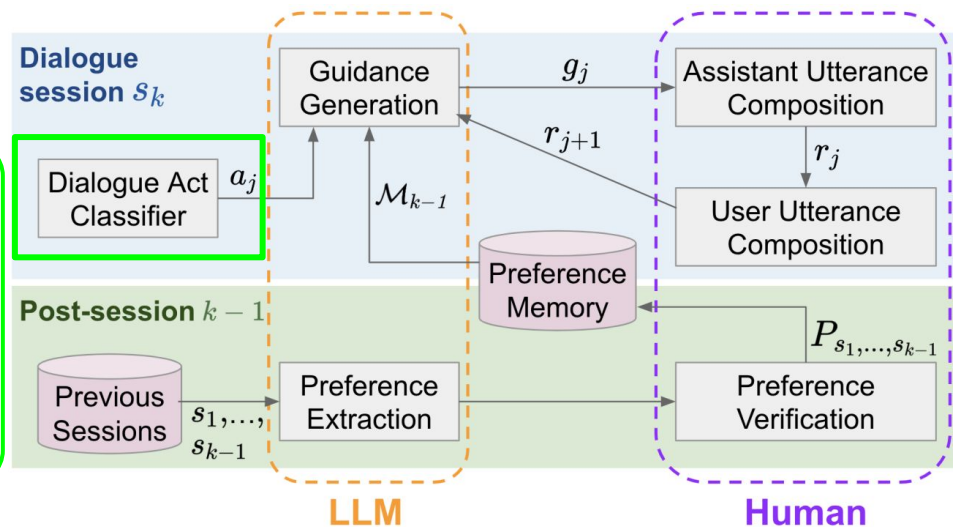
1) Dialogue act classification

2) Guidance generation

3) Utterance composition

4) Preference elicitation

- (1) Greeting
- (2) Preference elicitation
- (3) Recommendation
- (4) Follow-up questions
- (5) Goodbye

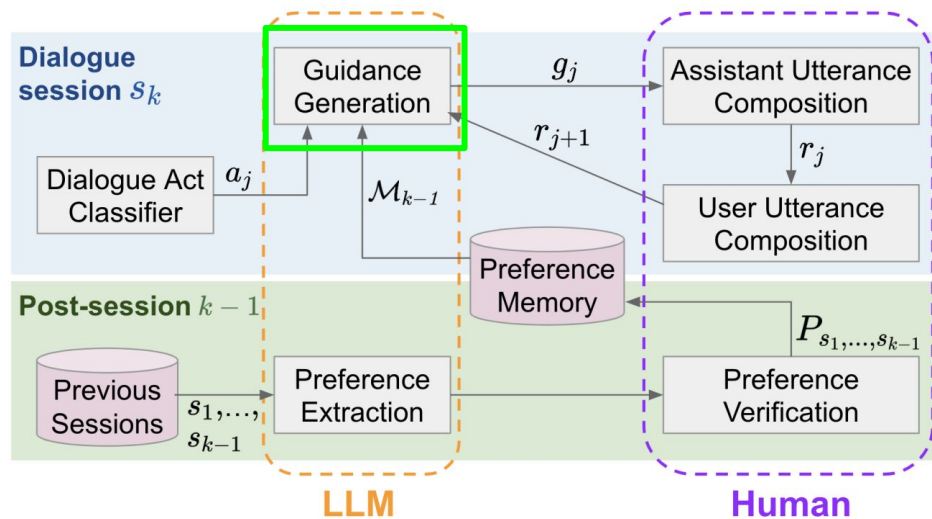


Utterance Generation - LLM guided/simulated

LLM-Guided Human Generation (LAPS) (Joko et al., 2024)

Main components:

- 1) Dialogue act classification
- 2) Guidance generation
- 3) Utterance composition
- 4) Preference extraction

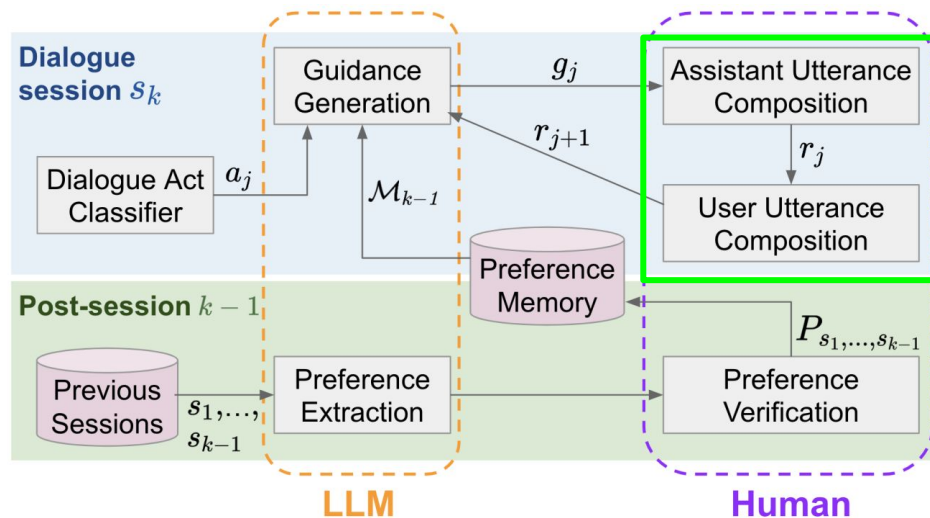


Utterance Generation - LLM guided/simulated

LLM-Guided Human Generation (LAPS) (Joko et al., 2024)

Main components:

- 1) Dialogue act classification
- 2) Guidance generation
- 3) Utterance composition
- 4) Preference extraction



Utterance Generation - LLM guided/simulated

LLM-Human Collaboration (MathDial)

(Macina et al., 2023)

- Task: Dialogue tutors
- Main components:
 - LLM as a student
 - Human as a teacher

Category	Intent	Example
Focus	Seek Strategy	So what should you do next?
	Guiding Student Focus	Can you calculate ...?
	Recall Relevant Information	Can you reread the question and tell me what is ...?
Probing	Asking for Explanation	Why do you think you need to add these numbers?
	Seeking Self Correction	Are you sure you need to add here?
	Perturbing the Question	How would things change if they had ... items instead?
	Seeking World Knowledge	How do you calculate the perimeter of a square?
Telling	Revealing Strategy	You need to add ... to ... to get your answer.
	Revealing Answer	No, he had ... items.
Generic	Greeting/Fairwell	Hi ..., how are you doing with the word problem? Good Job! Is there anything else I can help with?
	General inquiry	Can you go walk me through your solution?

Solve step-by-step:
James writes a 3-page letter to 2 different friends twice a week. How many pages does he write a year?



Human Teacher



LLM Student

*Your solution: James writes 3 x 2 ...
 Persona: Winnie thinks her answer is correct*

Hi Winnie, could you please walk me through your solution? **(generic)**

Sure! I first calculated the number of letters written in a week, which is 3 pages x 2 letters = 6 pages per week.

There is also one important keyword there: twice. What does it refer to? **(focus)**

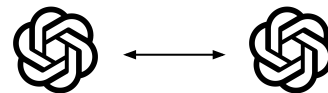
Correctly solved by student?

Correct

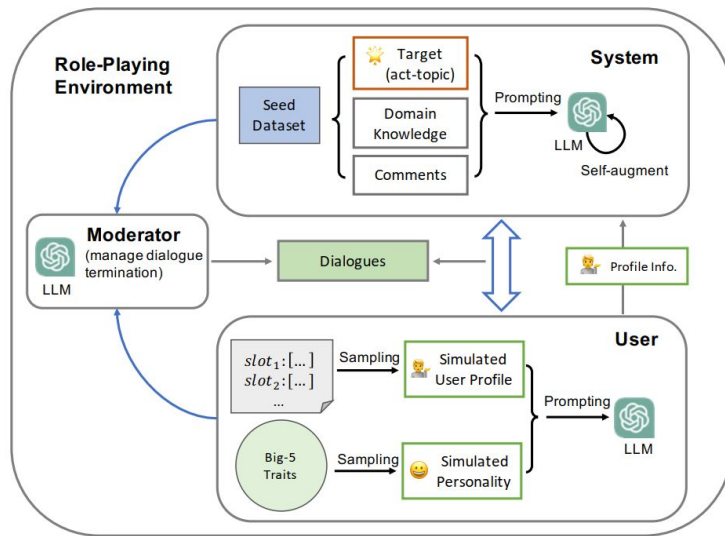
Incorrect

Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (TopDial) (Wang et al., 2023)



- **Task:** Target-oriented Recommendation System
- 3 LLMs collaboration
 - LLM as a User
 - LLM as a System
 - LLM as a Moderator



Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (SimQuAC) (Abbasiantaeb et al., 2024)

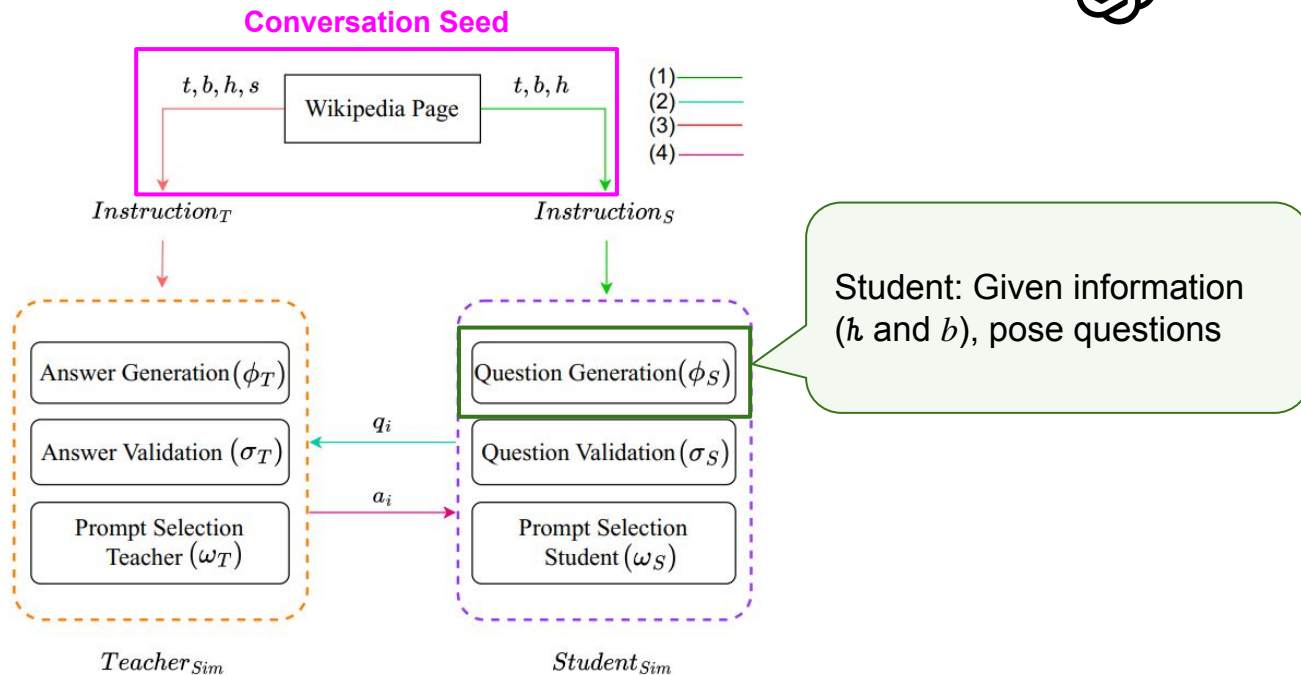
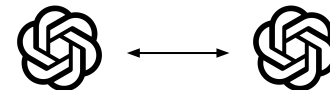
- QuAC: information-seeking QA dialogs dataset (Choi et al., 2018)
 - Continuation dialogue act:
 - it allows teachers to guide the student's questioning towards aspects of the article that are especially important or interesting
- replacing both human parties with LLMs
- Implement both the student and teacher by zero-shot prompting GPT-4

Section: 🦆 Daffy Duck, Origin & History

STUDENT: **What is the origin of Daffy Duck?**
TEACHER: ↔ first appeared in Porky's Duck Hunt
STUDENT: **What was he like in that episode?**
TEACHER: ↔ assertive, unrestrained, combative
STUDENT: **Was he the star?**
TEACHER: ↔ No, barely more than an unnamed bit player in this short
STUDENT: **Who was the star?**
TEACHER: ↗ No answer
STUDENT: **Did he change a lot from that first episode in future episodes?**
TEACHER: ↔ Yes, the only aspects of the character that have remained consistent (...) are his voice characterization by Mel Blanc
STUDENT: **How has he changed?**
TEACHER: ↔ Daffy was less anthropomorphic
STUDENT: **In what other ways did he change?**
TEACHER: ↔ Daffy's slobbery, exaggerated lisp (...) is barely noticeable in the early cartoons.
STUDENT: **Why did they add the lisp?**
TEACHER: ↔ One often-repeated "official" story is that it was modeled after producer Leon Schlesinger's tendency to lisp.
STUDENT: **Is there an "unofficial" story?**
TEACHER: ↔ Yes, Mel Blanc (...) contradicts that conventional belief
...

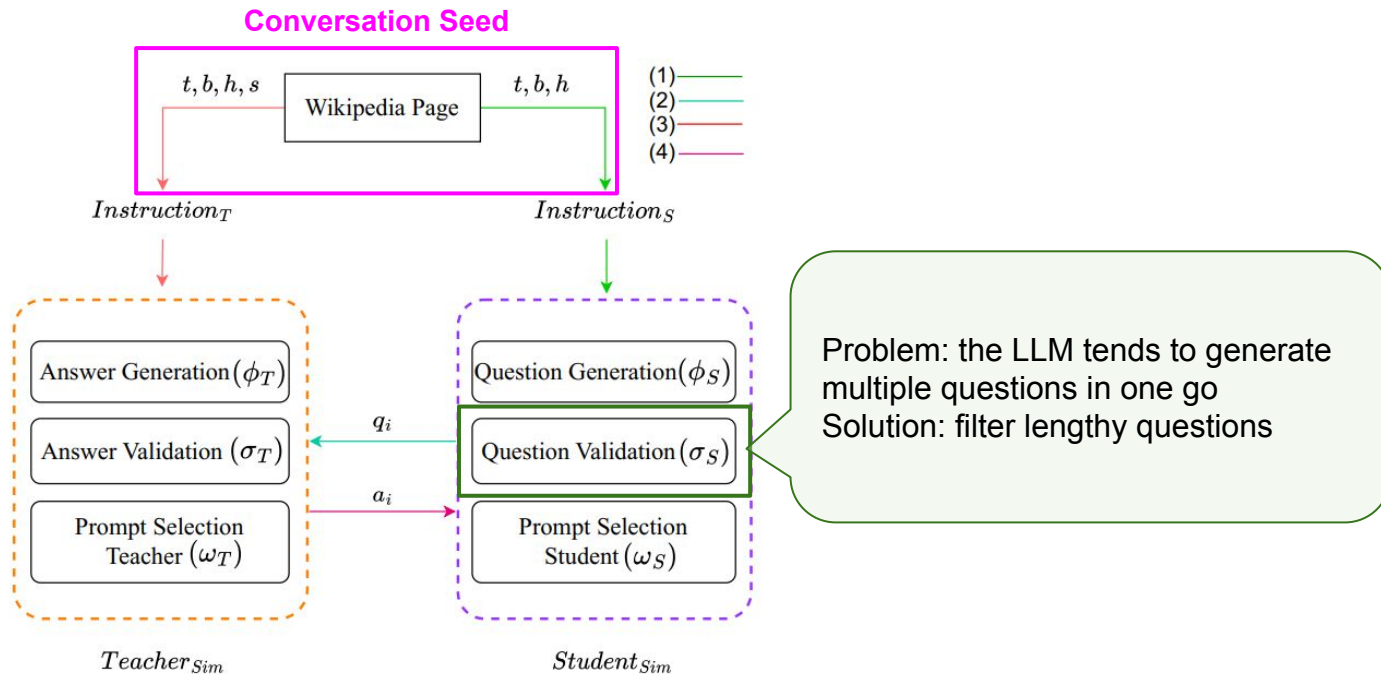
Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (SimQuAC) (Abbasiantaeb et al., 2024)



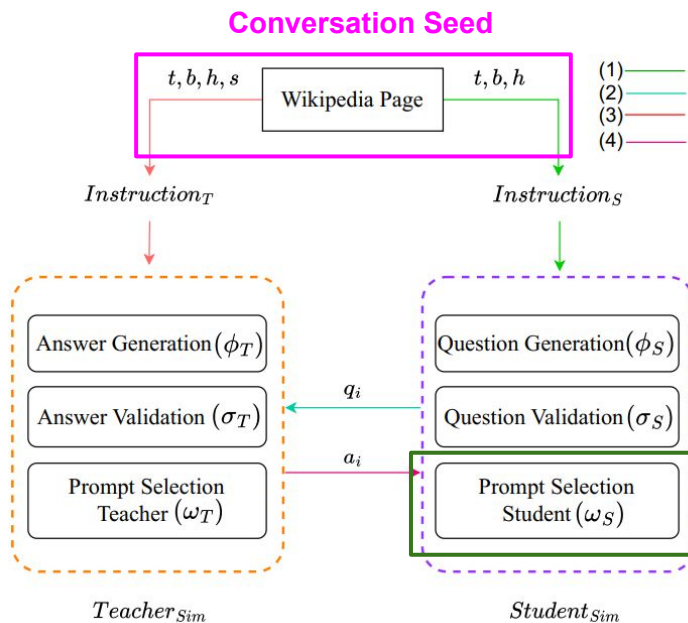
Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (SimQuAC) (Abbasiantaeb et al., 2024)



Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (SimQuAC) (Abbasiantaeb et al., 2024)



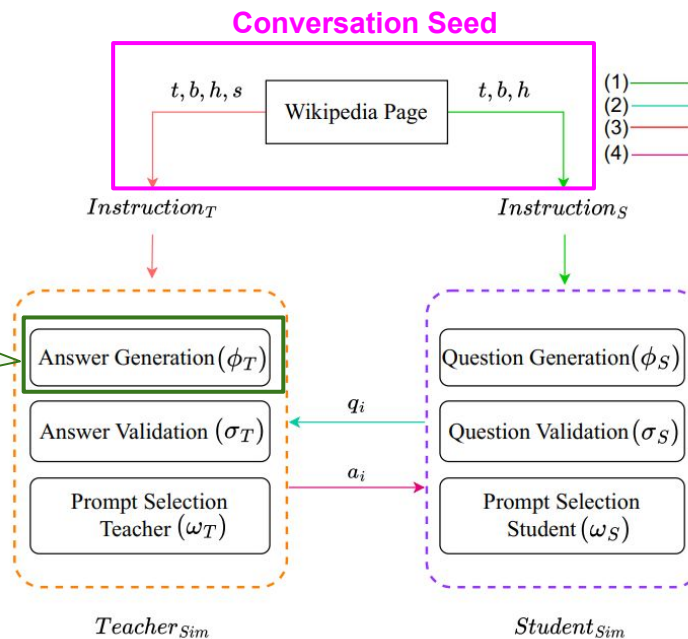
Problem: unanswered from the given text, higher chance that the subsequent question might be overly specific and cannot be answered

Solution: Randomly selects one of the following guiding prompts

- (i) Ask a general question
- (ii) Ask a question starting with where, when, or who
- (iii) Ask a question about what is interesting in this article
- (iv) Ask a question about another aspect of the topic

Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (SimQuAC) (Abbasiantaeb et al., 2024)



copy the exact spans from the passage to answer the given question

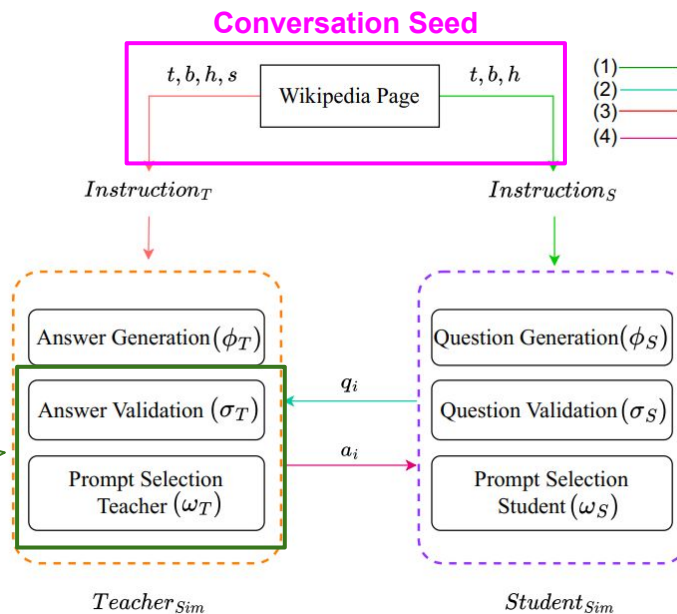
Utterance Generation - LLM guided/simulated

Fully LLM Generation & Supervision (SimQuAC) (Abbasiantaeb et al., 2024)

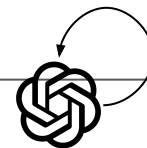
An iterative model to validate and refine the generated answers

- It checks whether the answer is copied from the text section or being "I cannot find the answer"

Solution: text search and multiple sequential prompts to generate other answers



Utterance Generation - LLM guided/simulated



One LLM plays all roles (SOLID) (Askari et al., 2024)

- **Reminder:** conversation seed: Generated background info + Sequence of intents
- How to apply intent in prompting?
 - Define Instruction

Table 10: The last part of the intent-based LLM-instruction. Actor type: Agent

Intent	Instruction
CQ	Reply with one follow-up response in conversation style.
FD	Reply with further details in conversation style.
GG	Continue the conversation by expressing gratitude for the user's questions.
PA	Provide a potential solution or answer in conversation style.
IR	Ask the user to provide relevant information needed for their previous question.
OQ	Formulate an original question posed by an agent.
FQ	Formulate a follow-up question from an agent, seeking further clarification or information.
RQ	Now you are talking from the point of view of a third participant in the conversation. Repeat Question:
PF	Express satisfaction and appreciation for the conversation.
NF	Convey dissatisfaction for the previous response.
JK	Reply with gibberish information. It can contain emojis.
O	Reply with a system error. Return N/A

Table 11: The last part of the intent-based LLM-instruction. Actor type: User

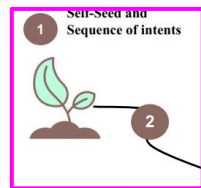
Intent	Instruction
CQ	Reply with one question asking for clarification in conversation style.
FD	Reply with more details in conversation style.
GG	Continue the conversation by expressing gratitude for the agent's help.
PA	Provide a potential solution or answer in conversation style.
IR	Reply with relevant information.
OQ	Formulate the first question posed by a user that initiates a QA dialog.
FQ	Formulate a follow-up question from a user, seeking further clarification or information.
RQ	Now you are talking from the point of view of a third participant in the conversation. Repeat Question:
PF	Express satisfaction and appreciation for a working solution.
NF	Convey dissatisfaction for the previous response.
JK	Reply with gibberish information. It can contain emojis.
O	Reply with a system error. Return N/A

Utterance Generation - LLM guided/simulated

One LLM plays all roles (SOLID) (Askari et al., 2024)

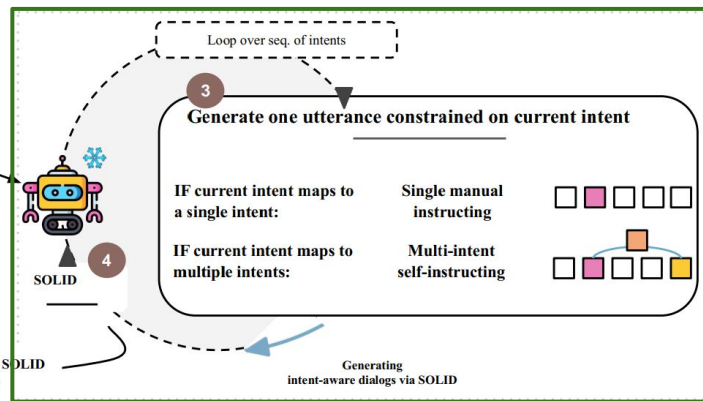
- Generates utterances guided by a specific intent or intents
- Each utterance generation fits under one of two cases
 - Single intent
 - Multiple intent
 - Prompt LLM to generate one merged instruction

Conversation Seed



5 Training on SOLID data

Utterance Generation

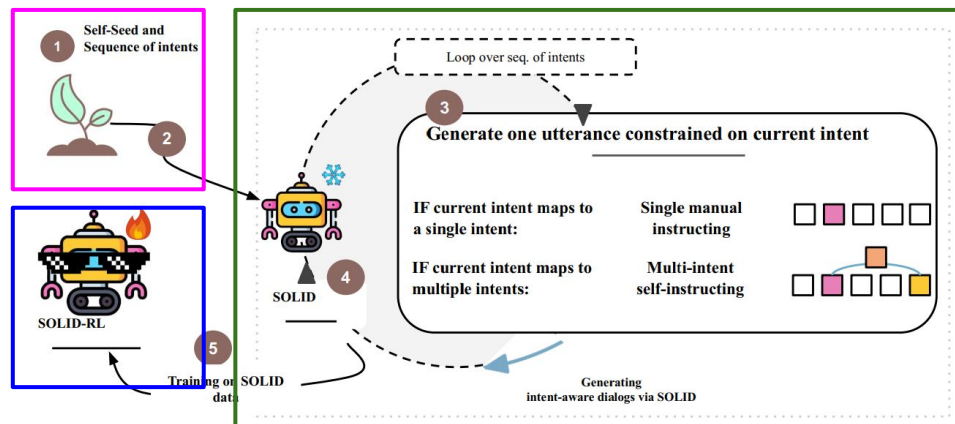


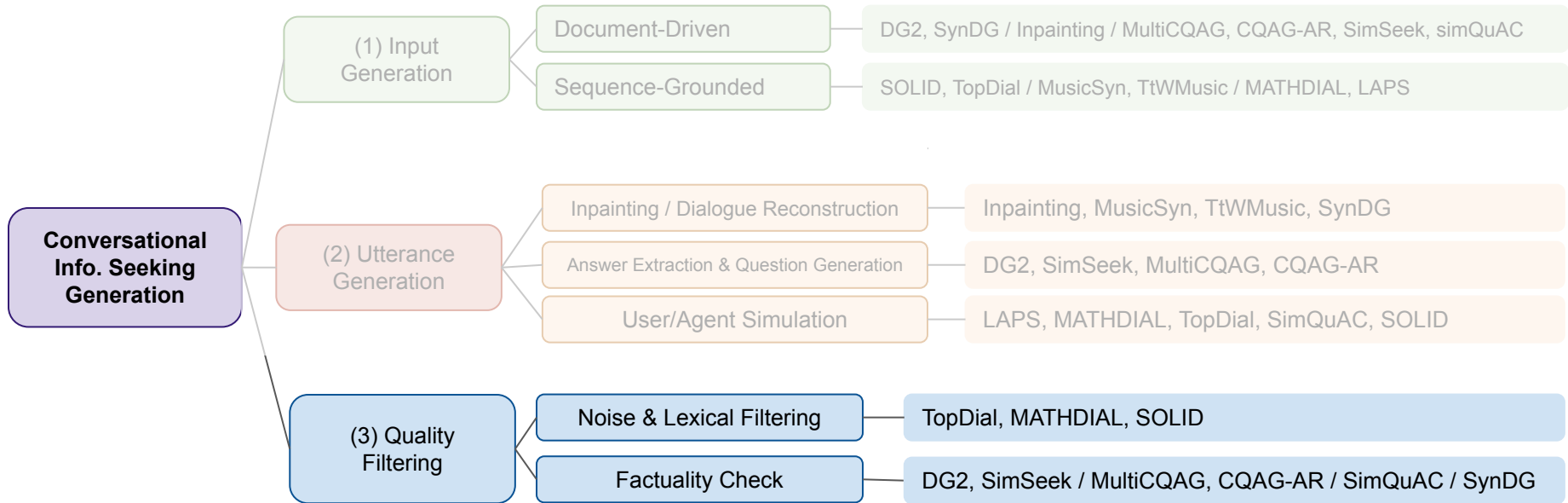
Utterance Generation - LLM guided/simulated

One LLM plays all roles - One Go Generation (SOLID-RL) (Askari et al., 2024)

- One Go generation advantages
 - Enhancing the naturalness
 - consistency of the conversation,
 - Increasing generation speed
- Approach
 - Fine-tuned on synthetic data

Fine-tuning for One-Go Generation

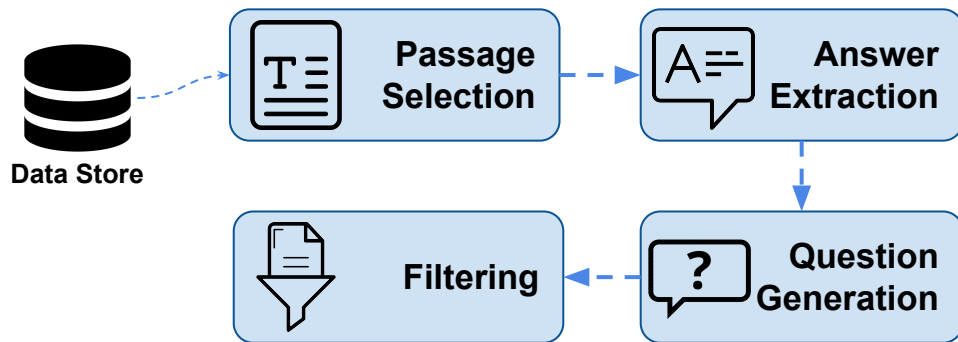




Quality Filtering - Factuality Check

Roundtrip Consistency

- For QA pair Generation



1) Passage Selection	C	... in 1903, boston participated in the first modern world series, going up against the pittsburgh pirates ...
2) Answer Extraction	C→A	1903
3) Question Generation	C, A→Q	when did the red sox first go to the world series
4) Filtering	C, Q→A' A±A'	1903 Yes

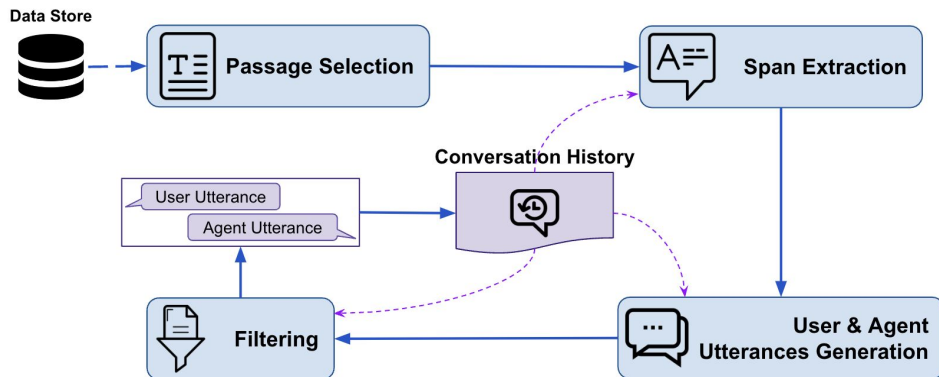
(Alberti et al., 2019)

Quality Filtering - Factuality Check

(Wu et al., 2022)

Roundtrip Consistency

- For Conversational Turn Generation



$$p(\hat{c}_t | \underbrace{\{u_i, a_i\}_{i < t}}_{\text{Conversation History}}, \underbrace{u_t}_{\text{User Utterance}}, \underbrace{C}_{\text{Document}})$$

$$p(\hat{r}_t | \underbrace{\{u_i, a_i\}_{i < t}}_{\text{Conversation History}}, \underbrace{u_t}_{\text{User Utterance}}, \underbrace{\hat{c}_t}_{\text{Document}})$$