

Chatty-Gen Supplementary Materials

Contents

1	Chatty-Gen Flexible Configurations	2
2	Results of Microsoft Academic Graph	3
2.1	Diversity	3
2.2	Node types Coverage	3
2.3	Running with different LLMs	4
2.4	Examples	5
3	Prompts	6
3.1	Choosing Entity Predicate (I_{EL})	6
3.2	Generating Self-Contained questions (I_{IQ})	6
3.3	Generating Self-Contained Questions With Triples (I'_{IQ})	6
3.4	Generating SPARQL Queries (I_{SQ})	7
3.5	Generating Dialogue (I_{DG})	7
3.6	A Single Prompt Approach (I_S)	8
4	SPARQL Queries	9
4.1	Extract Types Distribution	9
4.2	Detect Parent type	9
4.3	Batch retrieve entities	9
4.4	Graph size	9
4.5	Unique Predicates	9
5	ConvQuestions Comparison and Analysis	10
5.1	Coverage of questions	10
5.2	Examples of Similar Questions	10
5.3	Use case	12
5.4	GPT-4o Examples	12
5.5	5.5 Multi-LLM-1 Examples	15
6	DBLP & YAGO Examples	17

1 Chatty-Gen Flexible Configurations

CHATTY-GEN is a flexible framework that allows users to generate benchmarks from knowledge graphs (KGs) tailored to their needs. This section outlines key parameters and their implications. General parameters include (1) the target KG and its SPARQL endpoint, (2) the required number of dialogues, (3) the number of questions per dialogue, and (4) the LLM. CHATTY-GEN supports both open-source and commercial LLMs, each with varying context lengths that limit input subgraph sizes. Subgraphs exceeding the LLM’s context length are discarded to simplify data provision.

Regarding entity selection, CHATTY-GEN offers two modes: automatic sampling and manual entity specification. Users can create benchmarks with specific entities by providing a text file, enabling testing with different entity types such as head, torso, and tail. Head entities typically have larger subgraphs, increasing LLM processing costs, so users must ensure they fit within the LLM’s context length. Tail entities require a sufficient number of surrounding facts to generate a dialogue. The thresholds for rare and shadowed types control which node types are included in the benchmark: a lower rare type threshold includes more node types, while a lower shadowed parent threshold excludes more parents.

Subgraph extraction parameters include (1) the number of hops from the entity. Increasing the number of hops enlarges the subgraph and LLM processing costs, resulting in questions that focus on multiple entities rather than a single center entity; (2) predicate direction, where users can select the direction of predicates (incoming, outgoing, or both); and (3) the minimum number of questions per valid dialogue, which translates to the number of unique predicates surrounding the entity.

2 Results of Microsoft Academic Graph

2.1 Diversity

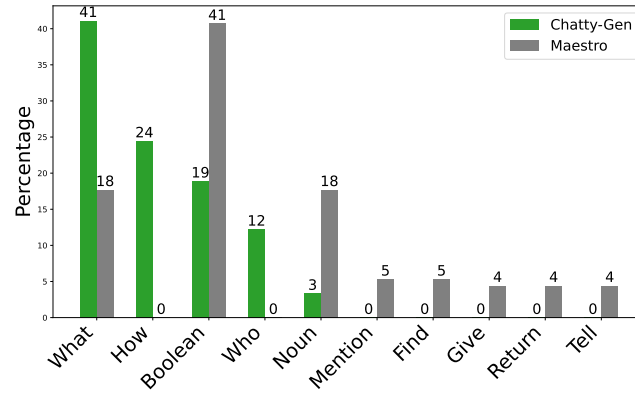


Figure 1: Comparison of the diversity of question types generated by Maestro and Chatty-Gen for MAG graph

2.2 Node types Coverage

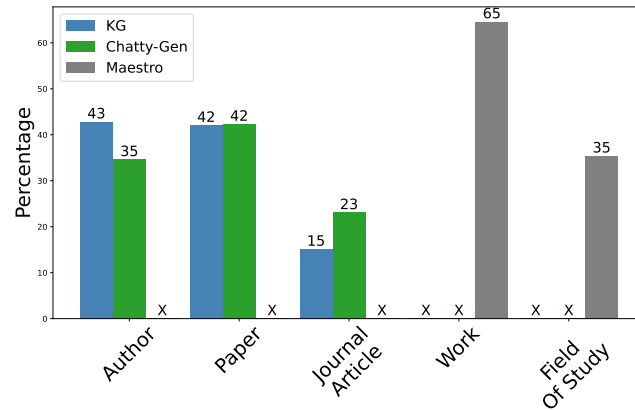


Figure 2: Comparison of the node-type distribution for the selected seed entities. by Maestro and Chatty-Gen for MAG graph

2.3 Running with different LLMs

Table 1 shows the ability of CHATTY-GEN to successfully run using different types of LLM on the large MAG graph. The MAG results align with YAGO and DBLP results, where GPT models produce the best performance. Multi-LLM-1 produces comparable performance to GPT models saving the generation cost.

Table 1: Evaluating the performance of our multi-stage approach using commercial and open-source LLMs on MAG

Approach	LLM	MAG							
		Success rate %	Dialogue-S	Question-E	SPARQL-E	Dialogue-E	Parsing-E	Time(mins)	# Tokens M (\$)
Our Multi-stages (3 prompts)	GPT-4o	87	20	0	2	0	1	14	0.08(0.40)
	GPT-4	95	20	0	1	0	0	10	0.05(0.48)
	GPT-3.5	80	20	2	3	0	0	7	0.06(0.03)
	Gemini-1-pro	69	20	1	1	0	7	29	0.07(0.01)
	Gemini-1.5-pro	39	20	0	3	1	30	128	0.12(0.23)
	LLAMA-3-8b	11	20	2	147	11	0	786	0.47
	LLAMA-3-8b-inst	2	11	2	424	0	60	332	1.25
	LLAMA-2-13b	3	15	10	443	3	19	1298	1.22
	CodeLLAMA-7b	5	20	44	222	39	59	550	1.25
	CodeLLAMA-13b	74	20	0	6	1	0	33	0.08
	Mistral-7b-v0.1	8	20	3	211	2	2	331	0.55
	Multi-LLM-1	90	20	0	2	0	0	21	0.06
	Multi-LLM-2	51	20	0	3	0	16	42	0.09(0.91)

2.4 Examples

Figure 3 shows examples of the generated dialogues by Chatty-Gen on two entity types (Author, and publication) from the MAG KG.

Entity: <https://makg.org/entity/245814409>

1. What is the total number of citations that James Tuot has?
2. How many papers has he created?
3. What is his rank?
4. Who is the creator associated with him?

Entity: <https://makg.org/entity/6185136>

1. What is the publication date of Theories of the Firm Through Duality?
2. Who is the creator of it?
3. Does it have a DOI?
4. How many keywords are associated with it?
5. What is the name of the publisher of it?

Figure 3: Examples of generated questions by Chatty-Gen for Microsoft Academic graph.

Figure 4 shows examples of the generated questions by Maestro on MAG KG. It follows a similar pattern to the examples of the other KGs.

1. What is the own resource of the paper "ECCV Workshops 1"?
2. Give me the own resource of the paper "ECCV Workshops 1"?
3. Return the has parent of the field of study "Pathosystem"?
4. Is the work "114296846r3" the own resource of the paper "ECCV Workshops 1"?
5. Is the field of study "Pathogen" the has parent of the field of study "Pathosystem"?

Figure 4: Examples of generated questions by Maestro for Microsoft Academic graph.

3 Prompts

3.1 Choosing Entity Predicate (I_{EL})

```
1 # Instruction:
2 Given the specified node type and its associated
3 predicates, choose a suitable predicate to be used to
4 extract labels for this type.
5
6 # Input:
7 node type:
8 predicates:
9
10 # Response:
```

Listing 1: Choosing representative predicate for type prompt

3.2 Generating Self-Contained questions (I_{IQ})

```
1 # Instruction:
2 Generate a list of questions based on the given entity
3 and its subgraph. The subgraph is represented as a
4 list of triples. Each question must ask about a fact
5 from the triples in the subgraph and must fall into
6 one of the following categories: list, count, boolean,
7 wh (open-ended), or date-related questions. Each
8 question must include the entity. Each question must
9 be answerable solely from the information in the
10 provided subgraph without explicitly mentioning it.
11
12 # Input:
13 Entity:
14 Subgraph:
15 number of questions:
16
17 # Response:
```

Listing 2: Generating independent questions prompt

3.3 Generating Self-Contained Questions With Triples (I'_{IQ})

```
1 # Instruction:
2 {Question Generation Prompt}
3 For each question, choose the triples from the input
4 subgraph which was used to generate the question.
```

```

5 Return both the question and the exact triple from the
6 subgraph that it was based on.
7
8 # Input:
9 Entity:
10 Subgraph:
11 number of questions:
12
13 # Response:

```

Listing 3: Generating Independent Questions Prompt with Triples

3.4 Generating SPARQL Queries (I_{SQ})

```

1 # Instruction:
2 Given a question and set of triples used to generate
3 this question. Create the SPARQL query representing
4 the question. Do not include the answer in the query.
5
6 # Input:
7 Question:
8 Triples:
9
10 # Response:

```

Listing 4: Generating SPARQL query from Question and Triples

3.5 Generating Dialogue (I_{DG})

```

1 # Instruction:
2 Given an entity and a set of questions focused on this
3 entity, choose the appropriate pronoun that refers to
4 it. Replace the entity with its pronoun in the
5 questions and return the modified questions. Ensure
6 that the modified questions do not contain the
7 original entity and that the pronoun used in the
8 modified questions is contextually appropriate and
9 grammatically correct.
10
11 # Input:
12 Questions:
13 Seed Entity:
14
15 # Response:

```

Listing 5: Generating SPARQL query from Question and Triples

3.6 A Single Prompt Approach (I_S)

```
1 # Instruction:
2 Generate a set of questions, a dialogue and sparqls
3 based on the provided entity and its subgraph. The
4 subgraph is represented as a varied list of triples.
5 Each question should be a fact from the triples in the
6 subgraph and fall into one of the following categories
7 : list, count, boolean, wh (open-ended), or
8 date-related questions. Each question should have the
9 entity and be answerable solely from the information
10 in the provided subgraph without explicitly mentioning
11 it. For the generated questions, generate a
12 corresponding dialogue where the first is standalone
13 and subsequent questions with replaced entity with its
14 pronoun. And a list of SPARQL queries that retrieves
15 answers. Return the following: questions, dialogue,
16 and SPARQL queries.
17
18 # Input:
19 Entity:
20 Subgraph:
21 number of questions:
22
23 # Response:
```

Listing 6: Generating Benchmark (standalone questions-SPARQL queries-Dialogue questions) using a single prompt

4 SPARQL Queries

The list of queries used for CHATTY-GEN cost-effective retrieval from KGs.

4.1 Extract Types Distribution

```
1 SELECT ?type (COUNT(?entity) AS ?count)
2 WHERE { ?entity rdf:type ?type. }
3 GROUP BY ?type
4 ORDER BY DESC(?count)
```

Listing 7: Getting KG’s node types distribution

4.2 Detect Parent type

```
1 SELECT ?parent
2 WHERE {
3 <t> <http://www.w3.org/2000/01/rdf-schema#subClassOf>
4 ?parent}
```

Listing 8: Detecting the parent type of type t

4.3 Batch retrieve entities

```
1 SELECT ?entity ?label
2 WHERE { ?entity rdf:type <t> . ?entity <l> ?label . }
3 LIMIT b
4 OFFSET o
```

Listing 9: Retrieving Entities of types t using entity label l for processing b is the number of entities to retrieve and o is the offset to start from

4.4 Graph size

```
1 SELECT count(*) as ?count
2 WHERE { {?s ?p <e>}
3         UNION {<e> ?p ?o} }
```

Listing 10: Getting Graph size for entity e

4.5 Unique Predicates

```
1 SELECT count(DISTINCT ?p) as ?count
2 WHERE { {?s ?p <e>}
3         UNION {<e> ?p ?o} }
```

Listing 11: Getting Number of unique predicates surrounding entity e

5 ConvQuestions Comparison and Analysis

5.1 Coverage of questions

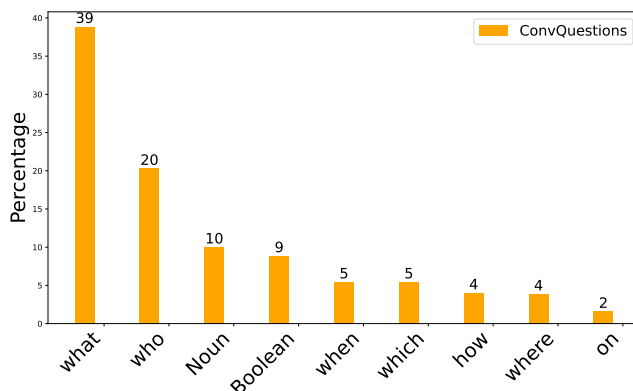


Figure 5: Distribution of question type in ConvQuestions dataset.

5.2 Examples of Similar Questions

5.2.1 Example 1 (Similarity Percentage: 50%)

ConvQuestions Dialogue:

- What is deadmau5’s real name?
- [Where was deadmau5 born?](#)
- What genre of music does deadmau5 perform?
- What label does deadmau5 publish his albums on?
- [When did deadmau5 start recording music?](#)

Chatty-Gen Dialogue:

- Does Deadmau5 have an image size specification listed?
- What genre is associated with Deadmau5?
- [When did Deadmau5 start their years of active music participation?](#)
- [Where was Deadmau5 born?](#)

5.2.2 Example 2 (Similarity Percentage: 60%)

ConvQuestions Dialogue:

- What is the release date of The Wonderful Wizard of Oz?
- Who was the author of The Wonderful Wizard of Oz?
- Who was the publisher of The Wonderful Wizard of Oz?
- Who illustrated The Wonderful Wizard of Oz?
- What was the name of the book that followed The Wonderful Wizard of Oz in the series?

Chatty-Gen Dialogue:

- Who is the author of The Wonderful Wizard of Oz?
- In which genre is the book The Wonderful Wizard of Oz classified?
- Is there any stated illustrations or illustrators for The Wonderful Wizard of Oz?
- What language is The Wonderful Wizard of Oz written in?
- Which publishing house released The Wonderful Wizard of Oz?

5.2.3 Example 3 (Similarity Percentage: 60%)

ConvQuestions Dialogue:

- Who wrote it?
- Who was the publisher of the novel Carrie?
- What was the name of the main character in the novel Carrie?
- Did the novel Carrie get made into a movie?
- When was the novel Carrie published?

Chatty-Gen Dialogue:

- Who is the author of Carrie (novel)?
- How many pages does Carrie (novel) have?
- Is there information regarding the release date for Carrie (novel)?
- Who published Carrie (novel)?
- What literary genre does Carrie (novel) belong to?

5.3 Use case

Evaluation Prompt:

```
1 # Instruction:
2 You are given two sets of questions : Set A and Set B.
3 Both sets ask questions about the same topic. Consider
4 factors such as fluency , clarity , and variety in the
5 questions . Based on these criteria , determine which
6 set has better overall quality . Return only Set A or
7 Set B.
8
9 # Input:
10 Set A:
11 Set B:
12
13 # Response:
```

Listing 12: Comparing the quality of questions generated by Chatty-gen to questions in ConvQuestions for the same entity

5.4 GPT-4o Examples

5.4.1 Example 1 (Tie)

Entity: http://dbpedia.org/resource/The_Wonderful_Wizard_of_Oz

ConvQuestions Dialogue:

- When was The Wonderful Wizard of Oz published?
- Who was the author of The Wonderful Wizard of Oz?
- Who was the publisher of The Wonderful Wizard of Oz?
- Who illustrated The Wonderful Wizard of Oz?
- What was the name of the book that followed The Wonderful Wizard of Oz in the series?

Chatty-Gen Dialogue:

- Who is the author of The Wonderful Wizard of Oz?
- In which genre is the book The Wonderful Wizard of Oz classified?
- Is there any stated illustrations or illustrators for The Wonderful Wizard of Oz?
- What language is The Wonderful Wizard of Oz written in?
- Which publishing house released The Wonderful Wizard of Oz?

5.4.2 Example 2 (ConvQuestions)

Entity: http://dbpedia.org/resource/Avril_Lavigne

ConvQuestions Dialogue:

- Where was Avril Lavigne born?
- In which country is the city of Belleville where Avril Lavigne was born?
- What is the height of Avril Lavigne?
- What is the title of the first album released by Avril Lavigne?
- What is the release date of the Avril Lavigne album Let Go?

Chatty-Gen Dialogue:

- What is Avril Lavigne's birthplace?
- How would you describe the occupations held by Avril Lavigne?
- Is Avril Lavigne currently under any known citizenship?
- When did Avril Lavigne start her active years as per the records?

5.4.3 Example 3 (Chatty-Gen)

Entity: http://dbpedia.org/resource/Major_League_Soccer

ConvQuestions Dialogue:

- When did Major League Soccer form?
- In which country was the MLS played out of?
- Who was the organizer of the MLS?
- What is the league directly before the MLS?
- When does the MLS season end?

Chatty-Gen Dialogue:

- What conference structure exists within Major League Soccer?
- How many teams are currently in Major League Soccer?
- Is Major League Soccer part of a relegation system?
- When was Major League Soccer founded?
- Which countries represent other associations in Major League Soccer?

5.4.4 Example 4 (Tie)

Entity: http://dbpedia.org/resource/Camp_Rock

ConvQuestions Dialogue:

- What genre is Camp Rock?
- Who was cast in the lead role for Camp Rock?
- What is the birth date of Demi Lovato?
- On what date was Camp Rock first shown?
- What is the title of the sequel to Camp Rock?

Chatty-Gen Dialogue:

- When was Camp Rock released?
- Who directed Camp Rock?
- Who are the stars of Camp Rock?
- Is there a particular country associated with Camp Rock?
- Can you name a composer who worked on Camp Rock?

5.4.5 Example 5 (ConvQuestions)

Entity: http://dbpedia.org/resource/Son_Heung-min

ConvQuestions Dialogue:

- What country is Son Heung-min from originally?
- What is Son Heung-min's date of birth?
- Which club team is Son Heung-min a current member of?
- Which professional soccer league is Tottenham Hotspur F.C. a member of?
- Did Son Heung-min previously play in the German Bundesliga?

Chatty-Gen Dialogue:

- What is Son Heung-min's birth date?
- How many teams has Son Heung-min played for?
- What is the surname of Son Heung-min?
- Is Son Heung-min a male?
- What approximate height is Son Heung-min?

5.5 5.5 Multi-LLM-1 Examples

5.5.1 Example 1 (Tie)

Entity: http://dbpedia.org/resource/Camp_Rock

ConvQuestions Dialogue:

- What genre is Camp Rock?
- Who was cast in the lead role for Camp Rock?
- What is the birth date of Demi Lovato?
- On what date was Camp Rock first shown?
- What is the title of the sequel to Camp Rock?

Chatty-Gen Dialogue:

- Who is the director of the movie Camp Rock?
- What is the release date of the movie Camp Rock?
- What is the runtime of the movie Camp Rock?
- Who is starring in the movie Camp Rock?
- What is the subsequent work of the movie Camp Rock?

5.5.2 Example 2 (Tie)

Entity: http://dbpedia.org/resource/Janis_Joplin

ConvQuestions Dialogue:

- What year was she born?
- What is the name of the city where Janis Joplin was born?
- In what year was Janis Joplin awarded the Grammy Lifetime Achievement Award?
- What is Janis Joplin's nickname?
- Where was Janis Joplin buried?

Chatty-Gen Dialogue:

- What is the birth date of Janis Joplin?
- In which year did Janis Joplin start her active years?
- What is the genre of music that Janis Joplin is associated with?
- What is the name of the record label that Janis Joplin was associated with?
- What is the name of the sound recording that Janis Joplin was associated with?

5.5.3 Example 3 (ConvQuestions)

Entity: <http://dbpedia.org/resource/Gloryhammer>

ConvQuestions Dialogue:

- Who is the lead singer of Gloryhammer?
- In what year was Gloryhammer formed?
- Who is the keyboardist of the band Gloryhammer?
- Other than Gloryhammer which band is the keyboardist Christopher Bowes part of?
- What year was the band Alestorm formed?

Chatty-Gen Dialogue:

- What is the genre of Gloryhammer?
- Who is a band member of Gloryhammer?
- What is the record label of Gloryhammer?
- What is the associated musical artist of Gloryhammer?

5.5.4 Example 4 (ConvQuestions)

Entity: http://dbpedia.org/resource/Son_Heung-min

ConvQuestions Dialogue:

- What country is Son Heung-min from originally?
- What is Son Heung-min's date of birth?
- Which club team is Son Heung-min a current member of?
- Which professional soccer league is Tottenham Hotspur F.C. a member of?
- Did Son Heung-min previously play in the German Bundesliga?

Chatty-Gen Dialogue:

- What is the birth date of Son Heung-min?
- What is the position of Son Heung-min?
- What is the height of Son Heung-min?
- What is the team of Son Heung-min?
- What is the birth place of Son Heung-min?

5.5.5 Example 5 (Tie)

Entity: <http://dbpedia.org/resource/Moby-Dick>

ConvQuestions Dialogue:

- Who is the author that wrote the book Moby-Dick?
- What is Herman Melville's date of birth?
- What is Herman Melville's place of birth?
- In which year was the book Moby-Dick originally published?
- Was a film ever made based on the book Moby-Dick?

Chatty-Gen Dialogue:

- What is the title of Moby-Dick?
- Who is the author of Moby-Dick?
- What is the language of Moby-Dick?
- In what country was Moby-Dick published?
- When was Moby-Dick released?

6 DBLP & YAGO Examples

Example YAGO:

Entity: http://yago-knowledge.org/resource/A._S._M._Hutchinson

1. Where was Arthur Stuart-Menteth Hutchinson born?
2. What alternate names are known for him?
3. What was his occupation?
4. When did he die?
5. Where did he pass away?

Example DBLP:

Entity: <https://dblp.org/rec/phd/Turing38>

1. What type of work is 'Systems of Logic Based on Ordinals.' classified as in Bibtex?
2. Who authored it?
3. What is the DOI of it?",
4. In which year was it published?
5. How many creators were involved in it?

Figure 6: Examples of generated questions by Chatty-Gen for KGs of different domains.