Explicit over Implict: Explicit Diversity Conditions for Effective Question Answer Generation

Vikas Yadav[1], Hyuk Joon Kwon[2], Vijay Srinivasan[2], Hongxia Jin[2]

ServiceNow, Samsung Research America

Contents

Synthetic Question Answer Generation using LLMs

- Introduction
    - Implicit diversity techniques and related issues
    - Our proposed explicit diversity conditions
- Improvements in Downstream QA model accuracy
- Impact of explicit diversity conditions on coverage and diversity of generated QA pairs from the input document.

Task: Question Answer Generation (QAG)

The tentacles of cydippid ctenophores are typically fringed with tentilla ("little tentacles"), although a few genera have simple tentacles without these sidebranches. The tentacles and tentilla are densely covered with microscopic colloblasts that capture prey by sticking to it. Colloblasts are specialized mushroom-shaped cells in the outer layer of the epidermis, and have three main components: a domed head with vesicles (chambers) that contain adhesive; a stalk that anchors the cell in the lower layer of the epidermis or in the mesoglea; and a spiral thread that coils round the stalk and is attached to the head and to the root of the stalk. The function of the spiral thread is uncertain, but it may absorb stress when prey tries to escape, and thus prevent the collobast from being torn apart. In addition to colloblasts, members of the genus Haeckelia, which feed mainly on jellyfish, incorporate their victims\' stinging nematocytes into their own tentacles – some cnidaria-eating nudibranchs similarly incorporate nematocytes into their bodies for defense. The tentilla of Euplokamis differ significantly from those of other cydippids: they contain striated muscle, a cell type otherwise unknown in the phylum Ctenophora; and they are coiled when relaxed, while the tentilla of all other known ctenophores elongate when relaxed. Euplokamis\' tentilla have three types of movement that are used in capturing prey: they may flick out very quickly (in 40 to 60 milliseconds); they can wriggle, which may lure prey by behaving like small planktonic worms; and they coil round prey. The unique flicking is an uncoiling movement powered by contraction of the striated muscle. The wriggling motion is produced by smooth muscles, but of a highly specialized type. Coiling around prey is accomplished largely by the return of the tentilla to their inactive state, but the coils may be tightened by smooth muscle.

What are the tentacles of cydippid ctenophores are usually fringed with?
tentilla

What are colloblasts? specialized mushroom-shaped cells in the outer layer of the epidermis

What makes the tentilla of euplokamis different from other cysippids? they contain striated muscle

How many types of movements do euplokamis tentilla have? three types of movement

What does the euplokamis use three types of movement for? capturing prey

Example Passage with Human Annotation

The tentacles of cydippid ctenophores are typically fringed with tentilla ("little tentacles"), although a few genera have simple tentacles without these sidebranches. The tentacles and tentilla are densely covered with microscopic colloblasts that capture prey by sticking to it. Colloblasts are specialized mushroom-shaped cells in the outer layer of the epidermis, and have three main components: a domed head with vesicles (chambers) that contain adhesive; a stalk that anchors the cell in the lower layer of the epidermis or in the mesoglea; and a spiral thread that coils round the stalk and is attached to the head and to the root of the stalk. The function of the spiral thread is uncertain, but it may absorb stress when prey tries to escape, and thus prevent the collobast from being torn apart. In addition to colloblasts, members of the genus Haeckelia, which feed mainly on jellyfish, incorporate their victims\' stinging nematocytes into their own tentacles – some cnidaria-eating nudibranchs similarly incorporate nematocytes into their bodies for defense. The tentilla of Euplokamis differ significantly from those of other cydippids: they contain striated muscle, a cell type otherwise unknown in the phylum Ctenophora; and they are coiled when relaxed, while the tentilla of all other known ctenophores elongate when relaxed. Euplokamis\' tentilla have three types of movement that are used in capturing prey: they may flick out very quickly (in 40 to 60 milliseconds); they can wriggle, which may lure prey by behaving like small planktonic worms; and they coil round prey. The unique flicking is an uncoiling movement powered by contraction of the striated muscle. The wriggling motion is produced by smooth muscles, but of a highly specialized type. Coiling around prey is accomplished largely by the return of the tentilla to their inactive state, but the coils may be tightened by smooth muscle.

Passage

What are the tentacles of cydippid ctenophores are usually fringed with? tentilla

What are colloblasts? specialized mushroom-shaped cells in the outer layer of the epidermis

What makes the tentilla of euplokamis different from other cysippids? they contain striated muscle

How many types of movements do euplokamis tentilla have? three types of movement

What does the euplokamis use three types of movement for? capturing prey

Human

SQuAD2.0, Rajpurkar et al., 2018

Examples of human annotated QA pairs from input document

QA pairs are generated only from 1st and 4th position of the document

Only “what” and “how” types of questions are annotated.

Lack of Diversity in state-of-the-art QAG model with diverse beam search

The tentacles of cydippid ctenophores are typically fringed with tentilla ("little tentacles"), although a few genera have simple tentacles without these sidebranches. The tentacles and tentilla are densely covered with microscopic colloblasts that capture prey by sticking to it. Colloblasts are specialized mushroom-shaped cells in the outer layer of the epidermis, and have three main components: a domed head with vesicles (chambers) that contain adhesive; a stalk that anchors the cell in the lower layer of the epidermis or in the mesoglea; and a spiral thread that coils round the stalk and is attached to the head and to the root of the stalk. The function of the spiral thread is uncertain, but it may absorb stress when prey tries to escape, and thus prevent the collobast from being torn apart. In addition to colloblasts, members of the genus Haecklia, which feed mainly on jellyfish, incorporate their victims\' stinging nematocytes into their own tentacles – some cnidaria-eating nudibranchs similarly incorporate nematocytes into their bodies for defense. The tentilla of Euplokamis differ significantly from those of other cydippids: they contain striated muscle, a cell type otherwise unknown in the phylum Ctenophora; and they are coiled when relaxed, while the tentilla of all other known ctenophores elongate when relaxed. Euplokamis\' tentilla have three types of movement that are used in capturing prey: they may flick out very quickly (in 40 to 60 milliseconds); they can wriggle, which may lure prey by behaving like small planktonic worms; and they coil round prey. The unique flicking is an uncoiling movement powered by contraction of the striated muscle. The wriggling motion is produced by smooth muscles, but of a highly specialized type. Coiling around prey is accomplished largely by the return of the tentilla to their inactive state, but the coils may be tightened by smooth muscle.

Passage

What are the tentacles of cydippid ctenophores are usually fringed with? tentilla
What are colloblasts? specialized mushroom-shaped cells in the outer layer of the epidermis

What makes the tentilla of euplokamis different from other cysippids? they contain striated muscle
How many types of movements do euplokamis tentilla have? three types of movement
What does the euplokamis use three types of movement for? capturing prey

Human

SQuAD2.0, Rajpurkar et al., 2018

What are the tentacles of cydippid ctenophores typically fringed with? Tentilla
What are the tentacles of cydippid ctenophores called? little tentacles
What are the tentacles of cydippid ctenophores called? tentilla
What are the tentacles of cydippid ctenophores? fringed with tentilla
What do the tentacles of cydippid ctenophores consist of? tentilla

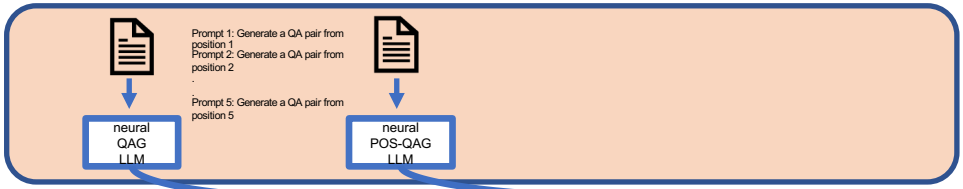
Beam Search

QA pairs generated from State of Art BART model using diverse decoding Beam search

All generated questions are from topmost part of the document

Severe lack of diversity in generation

Explicit Diverse QAG with POS condition



The tentacles of cydippid ctenophores are typically fringed with tentilla ("little tentacles"), although a few genera have simple tentacles without these sidebranches. The tentacles and tentilla are densely covered with microscopic colloblasts that capture prey by sticking to it. Colloblasts are specialized mushroom-shaped cells in the outer layer of the epidermis, and have three main components: a domed head with vesicles (chambers) that contain adhesive; a stalk that anchors the cell in the lower layer of the epidermis or in the mesoglea; and a spiral thread that coils round the stalk and is attached to the head and to the root of the stalk. The function of the spiral thread is uncertain, but it may absorb stress when prey tries to escape, and thus prevent the collobast from being torn apart. In addition to colloblasts, members of the genus Haeckelia, which feed mainly on jellyfish, incorporate their victims\' stinging nematocytes into their own tentacles – some cnidaria-eating nudibranchs similarly incorporate nematocytes into their bodies for defense. The tentilla of Euplokamis differ significantly from those of other cydippids: they contain striated muscle, a cell type otherwise unknown in the phylum Ctenophora; and they are coiled when relaxed, while the tentilla of all other known ctenophores elongate when relaxed. Euplokamis\' tentilla have three types of movement that are used in capturing prey: they may flick out very quickly (in 40 to 60 milliseconds); they can wriggle, which may lure prey by behaving like small planktonic worms; and they coil round prey. The unique flicking is an uncoiling movement powered by contraction of the striated muscle. The wriggling motion is produced by smooth muscles, but of a highly specialized type. Coiling around prey is accomplished largely by the return of the tentilla to their inactive state, but the coils may be tightened by smooth muscle.

Passage

What are the tentacles of cydippid ctenophores are usually fringed with? tentilla
What are colloblasts? specialized mushroom-shaped cells in the outer layer of the epidermis

What makes the tentilla of euplokamis different from other cysippids? they contain striated muscle
How many types of movements do euplokamis tentilla have? three types of movement
What does the euplokamis use three types of movement for? capturing prey

Human

What are the tentacles of cydippid ctenophores typically fringed with? Tentilla
What are the tentacles of cydippid ctenophores called? little tentacles
What are the tentacles of cydippid ctenophores called? tentilla
What are the tentacles of cydippid ctenophores? fringed with tentilla
What do the tentacles of cydippid ctenophores consist of? tentilla

Beam Search

1) What are the main components of a domed head with vesicles? a doming head with Vesicles (chambers) that contain adhesive; a stalk that anchors the cell in the lower layer of the epidermis or in the mesoglea; and a spiral thread

2) What is the domed head with vesicles called? chambers

3) What do the name of the group of jellyfish that feed mainly on jellyfish? Haeckelia

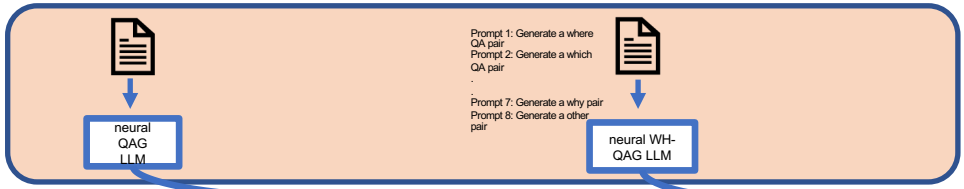
4) What is the cell type of Euplokamis' tentilla? striated muscle

5) How long do Euplokamis' tentacles become coiled? When relaxed?

POS prompt

QA pairs generated from 5 different positions with our POS condition

Explicit Diverse QAG with WH-type condition



The tentacles of cydippid ctenophores are typically fringed with tentilla ("little tentacles"), although a few genera have simple tentacles without these sidebranches. The tentacles and tentilla are densely covered with microscopic colloblasts that capture prey by sticking to it. Colloblasts are specialized mushroom-shaped cells in the outer layer of the epidermis, and have three main components: a domed head with vesicles (chambers) that contain adhesive; a stalk that anchors the cell in the lower layer of the epidermis or in the mesoglea; and a spiral thread that coils round the stalk and is attached to the head and to the root of the stalk. The function of the spiral thread is uncertain, but it may absorb stress when prey tries to escape, and thus prevent the collobast from being torn apart. In addition to colloblasts, members of the genus Haeckelia, which feed mainly on jellyfish, incorporate their victims\' stinging nematocytes into their own tentacles – some cnidaria-eating nudibranchs similarly incorporate nematocytes into their bodies for defense. The tentilla of Euplokamis differ significantly from those of other cydippids: they contain striated muscle, a cell type otherwise unknown in the phylum Ctenophora; and they are coiled when relaxed, while the tentilla of all other known ctenophores elongate when relaxed. Euplokamis\' tentilla have three types of movement that are used in capturing prey: they may flick out very quickly (in 40 to 60 milliseconds); they can wriggle, which may lure prey by behaving like small planktonic worms; and they coil round prey. The unique flicking is an uncoiling movement powered by contraction of the striated muscle. The wriggling motion is produced by smooth muscles, but of a highly specialized type. Coiling around prey is accomplished largely by the return of the tentilla to their inactive state, but the coils may be tightened by smooth muscle.

Passage

What are the tentacles of cydippid ctenophores are usually fringed with? tentilla
What are colloblasts? specialized mushroom-shaped cells in the outer layer of the epidermis

What makes the tentilla of euplokamis different from other cysippids? they contain striated muscle
How many types of movements do euplokamis tentilla have? three types of movement
What does the euplokamis use three types of movement for? capturing prey

Human

What are the tentacles of cydippid ctenophores typically fringed with? Tentilla
What are the tentacles of cydippid ctenophores called? little tentacles
What are the tentacles of cydippid ctenophores called? tentilla
What are the tentacles of cydippid ctenophores? fringed with tentilla
What do the tentacles of cydippid ctenophores consist of? tentilla

Beam Search

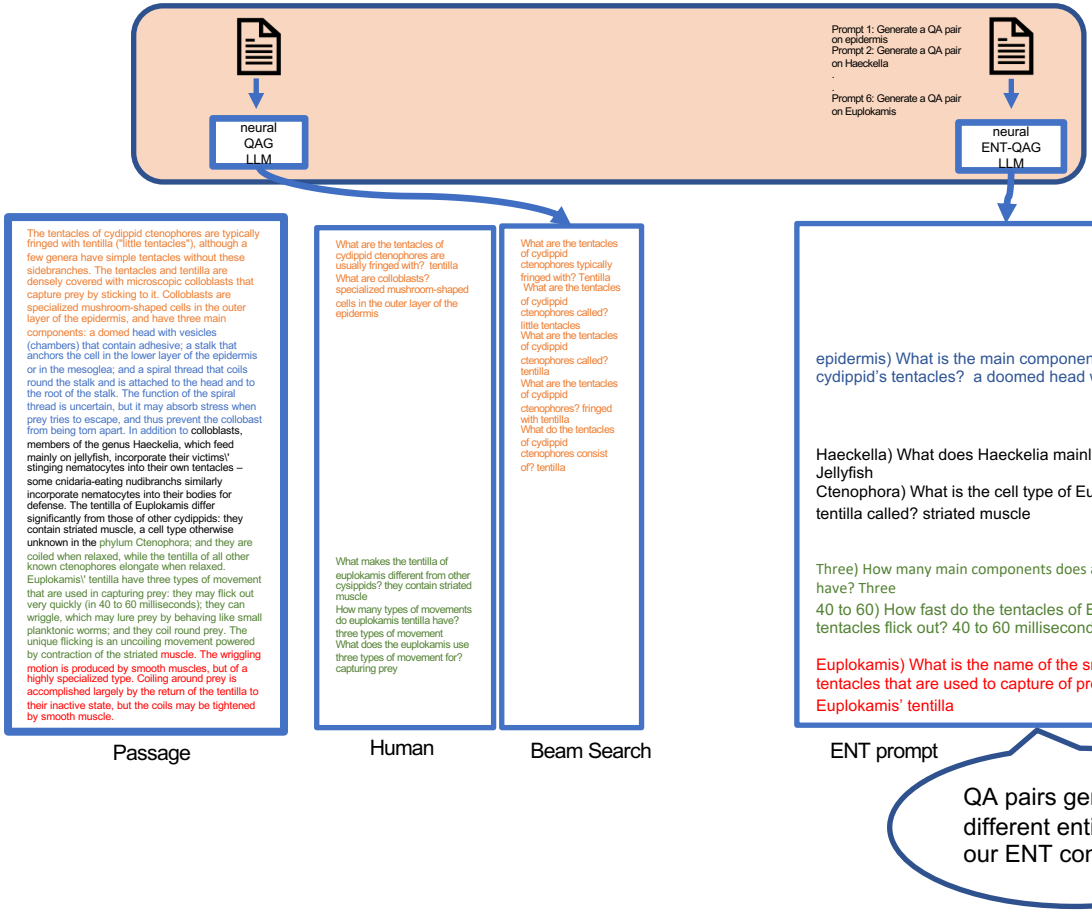
Where) Where are the tentacles of cydippid ctenophores fringed? tentilla ("little tentacles")
Which) Which group of ctenophores have small tentacles? cydippid ctenos

What) What are the tentacles of cydippid ctenophores typically fringed with? tentilla ("little tentacles")
Who) Who is the only cnidria-eating nudibranch that incorporate nematocytes into their own tentacles? cnids
Why) Why do cnidaria-eating nudibranchs incorporate nematocytes into their own tentacles? Defense
When) When do Euplokamis' tentacles become coiled? When relaxed
Other) How long do the tentacles of cydippid ctenophores last? 40 to 60 milliseconds

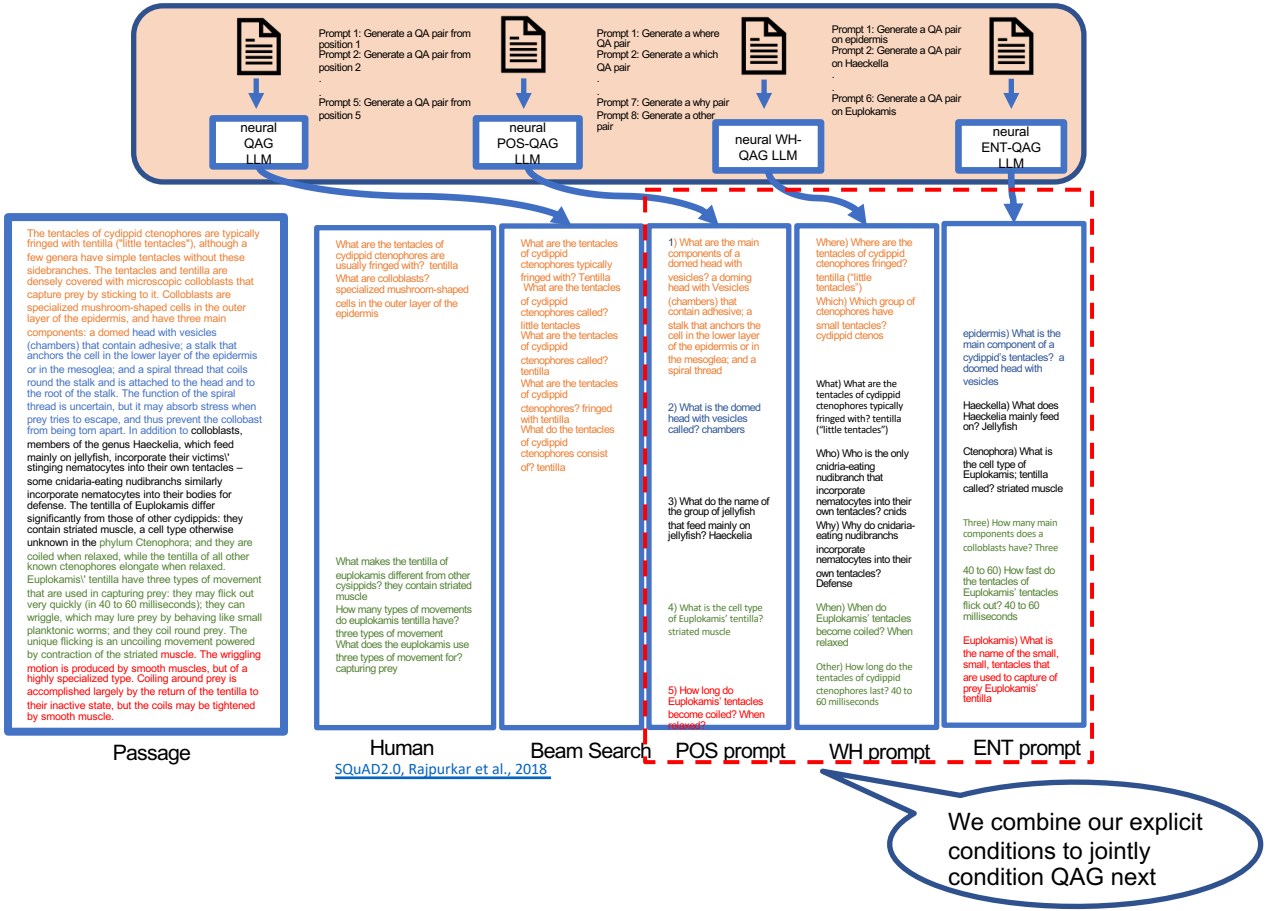
WH prompt

QA pairs generated with different WH types with our WH condition

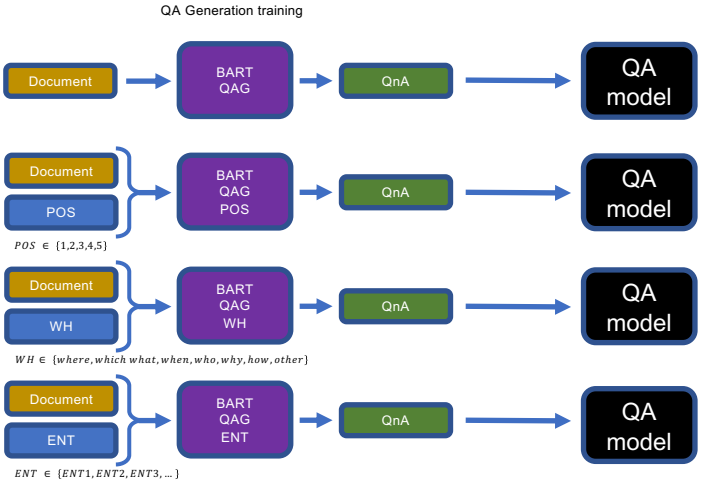
Explicit Diverse QAG with ENT condition



QAG pairs from explicit diversity conditions



Training a QA model to evaluate synthetic QA pairs from the QAG model



Improvements in Downstream QA model accuracy

We train BERT-large-whole-word-masked based QA model on the synthetic data generated by our QAG models. Then we evaluate the performance of the QA model on the SQuAD held out test set which is summarized below.

Source	Method	Approach	BART Orig Size EM	BART Orig Size F1	LLaMa-7B Orig Size EM	LLaMa-7B Orig Size F1
Syn		Greedy	64.76	76.66	71.41	83.26
Syn	Implicit	Nucleus (0.95)	64.44	77.15	71.92	83.53
Syn	Sampl.	Nucleus+TopK	64.17	76.47	72.08	83.71
Syn		DiverseDecod.	65.21	77.37	71.87	83.66
Syn	Explicit	QuesType (WH)	67.25	79.60	72.97	84.46
Syn	Divers.	Position (POS)	69.62	81.49	72.74	84.25
Syn	prompt	Entity (ENT)	69.31	81.80	72.59	84.21
Syn		POS->WH	71.77	83.30	-	-
Syn	Learned	ENT->WH	70.46	81.78	-	-
Syn		Combined	-	-	**73.29**	**84.76**
H+Syn		$SQuAD_{dev}$+WH	74.30	85.62	75.11	86.42
H+Syn		$SQuAD_{dev}$+POS	74.53	85.61	**75.76**	**87.15**
H+Syn		$SQuAD_{dev}$+ENT	73.17	85.01	75.59	87.06
Hum(H)		$SQuAD_{dev}$	EM=74.08		F1=85.19	

Table 1: **Downstream QA performance on the QG-bench SQuAD DU test dataset. We use top_p=0.95 and topK=30.The fourth-row block settings refer to the learned combination of diversity conditions (section 3.2) where the first prompt predicts the second potential diversity prompt (separated by ->). The Orig Size indicates that the synthetic data size matches the original training size of SQuAD DU dataset of 10570 QA pairs. The eval dataset for all the rows is the SQuAD DU test split which contains 11877 QA pairs. Hum and syn refers to human annotated and synthetic QA dataset respectively.**

SQuAD_DU (Du et al., 2017) dataset from QG-bench (Ushio et al., 2022)

Improvements in Downstream QA model accuracy

Source	Method	Approach	BART Orig Size EM	BART Orig Size F1	LLaMa-7B Orig Size EM	LLaMa-7B Orig Size F1
Syn		Greedy	64.76	76.66	71.41	83.26
Syn	Implicit	Nucleus (0.95)	64.44	77.15	71.92	83.53
Syn	Sampl.	Nucleus+TopK	64.17	76.47	72.08	83.71
Syn		DiverseDecod.	65.21	77.37	71.87	83.66
Syn	Explicit	QuesType (WH)	67.25	79.60	72.97	84.46
Syn	Divers.	Position (POS)	69.62	81.49	72.74	84.25
Syn	prompt	Entity (ENT)	69.31	81.80	72.59	84.21
Syn		POS->WH	71.77	83.30	-	-
Syn	Learned	ENT->WH	70.46	81.78	-	-
Syn		Combined	-	-	**73.29**	**84.76**
H+Syn		$SQuAD_{dev}$+WH	74.30	85.62	75.11	86.42
H+Syn		$SQuAD_{dev}$+POS	74.53	85.61	**75.76**	**87.15**
H+Syn		$SQuAD_{dev}$+ENT	73.17	85.01	75.59	87.06
Hum(H)		$SQuAD_{dev}$	EM=74.08		F1=85.19	

Table 1: **Downstream QA performance on the QG-bench SQuAD DU test dataset. We use top_p=0.95 and topK=30.The fourth-row block settings refer to the learned combination of diversity conditions (section 3.2) where the first prompt predicts the second potential diversity prompt (separated by ->). The Orig Size indicates that the synthetic data size matches the original training size of SQuAD DU dataset of 10570 QA pairs. The eval dataset for all the rows is the SQuAD DU test split which contains 11877 QA pairs. Hum and syn refers to human annotated and synthetic QA dataset respectively.**

Baseline approach with QA model trained on synthetic QA pairs from State of Art QAG method

Diverse QA pair generation improves downstream QA performance

SQuAD_DU (Du et al., 2017) dataset from QG-bench (Ushio et al., 2022)

Improvements in Downstream QA model accuracy

Source	Method	Approach	BART Orig Size EM	BART Orig Size F1	LLaMa-7B Orig Size EM	LLaMa-7B Orig Size F1
Syn		Greedy	64.76	76.66	71.41	83.26
Syn	Implicit	Nucleus (0.95)	64.44	77.15	71.92	83.53
Syn	Sampl.	Nucleus+TopK	64.17	76.47	72.08	83.71
Syn		DiverseDecod.	65.21	77.37	71.87	83.66
Syn	Explicit	QuesType (WH)	67.25	79.60	72.97	84.46
Syn	Divers.	Position (POS)	69.62	81.49	72.74	84.25
Syn	prompt	Entity (ENT)	69.31	81.80	72.59	84.21
Syn		POS->WH	71.77	83.30	-	-
Syn	Learned	ENT->WH	70.46	81.78	-	-
Syn		Combined	-	-	**73.29**	**84.76**
H+Syn		$SQuAD_{dev}$+WH	74.30	85.62	75.11	86.42
H+Syn		$SQuAD_{dev}$+POS	74.53	85.61	**75.76**	**87.15**
H+Syn		$SQuAD_{dev}$+ENT	73.17	85.01	75.59	87.06
Hum(H)		$SQuAD_{dev}$	EM=74.08		F1=85.19	

Table 1: **Downstream QA performance on the QG-bench SQuAD DU test dataset. We use top_p=0.95 and topK=30.The fourth-row block settings refer to the learned combination of diversity conditions (section 3.2) where the first prompt predicts the second potential diversity prompt (separated by ->). The Orig Size indicates that the synthetic data size matches the original training size of SQuAD DU dataset of 10570 QA pairs. The eval dataset for all the rows is the SQuAD DU test split which contains 11877 QA pairs. Hum and syn refers to human annotated and synthetic QA dataset respectively.**

Learning which QA pairs to generate further improves QA performance

Almost achieves same downstream QA performance as that from Human annotated datasets

SQuAD_DU (Du et al., 2017) dataset from QG-bench (Ushio et al., 2022)

Improvements in Downstream QA model accuracy

Source	Method	Approach	BART Orig Size EM	BART Orig Size F1	LLaMa-7B Orig Size EM	LLaMa-7B Orig Size F1
Syn		Greedy	64.76	76.66	71.41	83.26
Syn	Implicit	Nucleus (0.95)	64.44	77.15	71.92	83.53
Syn	Sampl.	Nucleus+TopK	64.17	76.47	72.08	83.71
Syn		DiverseDecod.	65.21	77.37	71.87	83.66
Syn	Explicit	QuesType (WH)	67.25	79.60	72.97	84.46
Syn	Divers.	Position (POS)	69.62	81.49	72.74	84.25
Syn	prompt	Entity (ENT)	69.31	81.80	72.59	84.21
Syn		POS->WH	71.77	83.30	-	-
Syn	Learned	ENT->WH	70.46	81.78	-	-
Syn		Combined	-	-	**73.29**	**84.76**
H+Syn		$SQuAD_{dev}$+WH	74.30	85.62	75.11	86.42
H+Syn		$SQuAD_{dev}$+POS	74.53	85.61	**75.76**	**87.15**
H+Syn		$SQuAD_{dev}$+ENT	73.17	85.01	75.59	87.06
Hum(H)		$SQuAD_{dev}$	EM=74.08		F1=85.19	

Table 1: **Downstream QA performance on the QG-bench SQuAD DU test dataset. We use top_p=0.95 and topK=30.The fourth-row block settings refer to the learned combination of diversity conditions (section 3.2) where the first prompt predicts the second potential diversity prompt (separated by ->). The Orig Size indicates that the synthetic data size matches the original training size of SQuAD DU dataset of 10570 QA pairs. The eval dataset for all the rows is the SQuAD DU test split which contains 11877 QA pairs. Hum and syn refers to human annotated and synthetic QA dataset respectively.**

Combining with human annotations further improves QA performance

SQuAD_DU (Du et al., 2017) dataset from QG-bench (Ushio et al., 2022)

Improvements in Downstream QA model accuracy

We compute the lexical token overlap between the generated QA pairs for each document. We generated 5 questions with each approach and report the average pairwise token lexical overlap between all 5 Choose 2 QA pairs over SQuAD_DU dev split. We also compute the average lexical coverage of the 5 generated QA pairs by assessing answer text position (POS), entity and *wh* in question text.
The average lexical overlap token overlap and coverage of generated QA pairs are reported in the table below.

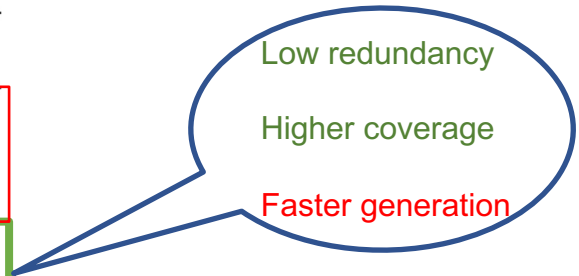
Analysis	**Overlap**	**Coverage**			**Time**
		POS	WH	ENT	(ms)
Greedy	63.07	36.84	31.33	32.18	223.1
Nucleus (0.95)	57.44	57.15	45.59	29.80	372.1
Nucleus+TopK	59.93	58.62	48.23	30.21	451.4
DiverseDecoding	46.85	49.83	42.76	35.38	388.2
POS	36.10	77.56	34.62	50.62	231.5
WH	30.67	60.41	97.81	48.06	218.7
ENT	34.59	75.89	55.34	63.90	227.9
Human	28.04	65.82	56.32	44.96	-

Table 3: **Pairwise lexical overlap between generated QA tokens, QA generation coverage, and average time for generating 5 QA pairs from SQuAD_{DU}.**

Improvements in Downstream QA model accuracy

Analysis	Overlap	Coverage			Time
		POS	WH	ENT	(ms)
Greedy	63.07	36.84	31.33	32.18	223.1
Nucleus (0.95)	57.44	57.15	45.59	29.80	372.1
Nucleus+TopK	59.93	58.62	48.23	30.21	451.4
DiverseDecoding	46.85	49.83	42.76	35.38	388.2
POS	36.10	77.56	34.62	50.62	231.5
WH	30.67	60.41	97.81	48.06	218.7
ENT	34.59	75.89	55.34	63.90	227.9
Human	28.04	65.82	56.32	44.96	-

Table 3: Pairwise lexical overlap between generated QA tokens, QA generation coverage, and average time for generating 5 QA pairs from SQuAD_{DU}.



Note – Our approach uses just 1 beam for each of the explicit joint conditions leading to much faster QA text generations

Conclusions

- We presented a detailed study of implicit versus explicit conditioning techniques for diverse QA generation. Existing implicit techniques lack diversity in generations.

- Our work empirically shows the clear benefits of explicit diversity conditions with substantial improvements in diverse generations leading to improved downstream QA task performance

- Our explicit conditions for diverse QAG also maximizes the information coverage from the input document

Thank you! Questions?