

## A Details of Multitask Training

The seq2seq multi-task model was trained on a variety of tasks at training time. Each task is specified by a special token to delineate to the model which task it is. Tasks at training time include the following, in the form of (source, target) pairs. “+” represents a concatenation of inputs, separated by a special token.

- (empty, question)
- (empty, document)
- (empty, answer)
- (empty, question + document)
- (empty, question + document + answer)
- (question, answer)
- (question, document)
- (question + document, answer)
- (question, document + answer)
- masked word prediction: 15% of source words are replaced by a “[MASK]” token and the corresponding tokens must be predicted as the target in the correct order

## B Architectural Details

### B.1 Extractive BidAF

The BidAF model is trained using the AllenNLP<sup>8</sup> implementation, using the standard hyper-parameters (specified in the bidaf.jsonnet file<sup>9</sup>). We only change the batch size, since a 16GB GPU can only fit one example per batch, and as a result the Adam learning rate has to be changed to  $5e - 5$ . We provide the code to select the target span and sub-sample the input in our data, as well as to convert it to the SQUAD format required by the AllenNLP system.

### B.2 Abstractive Models

Models are trained with the Adam optimizer with beta values (0.9, 0.98), initial learning rate  $1e - 07$  with 4000 warmup steps to learning rate 0.0001. We follow the inverse square root learning rate scheduler described in (Vaswani et al., 2017). Models are trained with a label smoothing value of 0.1.

<sup>8</sup><https://allennlp.org/>

<sup>9</sup>[https://github.com/allenai/allennlp/blob/master/training\\_config/bidaf.jsonnet](https://github.com/allenai/allennlp/blob/master/training_config/bidaf.jsonnet)

Sequence to sequence models are trained with following architecture from (Vaswani et al., 2017): 6 encoder layers, 6 decoder layers, FFN dimension 4096, 16 attention heads, embedding dimension 1024. Gradient updating is delayed until 32 updates have been processed. Models are regularized with dropout 0.1 and attention dropout 0.1.

Language models are trained with same parameters described for seq2seq above, with 6 decoder layers. We did not train with 12 decoder layers, as we found the deeper Transformer model was harder to optimize and we achieved worse results compared to a 6-layer language model.

For generation, models generate a minimum of 200 words and a maximum of 500 words.

## C Comparison of Extractive and Abstractive Methods

Figure 13 displays an example of a generated answer for an example where the source document is of poor quality but the abstractive answer still has strong ROUGE. In comparison, the extractive answer is heavily affected by the poor document quality and derails in topic.

## D Test/Valid Similarity with Train

Figure 9 shows the performance of the Multi-task Seq2Seq and LM on Question + Document + Answer by the similarity of the question in the validation set to a question in the training set. The similarity is determined by TFIDF. There is very little effect of answer generation on a question more similar to a training question than less similar.

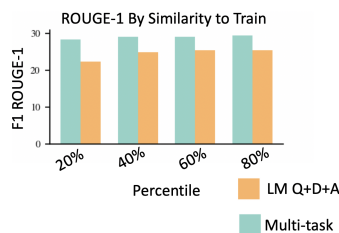


Figure 9: ROUGE of full answer generation is not strongly affected by similarity of the questions in the validation set to questions in the training set.

## E Variance in Human Evaluation Studies

We analyze the variation of our human evaluation study for answer generation fluency in Figure 10. We conduct 3 different trials of the same 100 randomly sampled question-answer pairs from the

test set for the selected models. Each trial is conducted on a different day. Our results show that standard deviation across the trials is small and not statistically significant.

Further, each answer is evaluated for fluency by 3 different crowdworkers. Figure 10 analyzes the agreement rate between crowdworkers that can choose on a scale of five options. We term “agreement” if all workers are positive, negative, or neutral about the answer fluency. We show that all three crowdworkers agree around 60% of the time for most models and almost 80% of the time for the language model. As the language model generation is significantly less fluent than the other models, most crowdworkers are in agreement. The agreement of at least two of the annotators is almost 100% for all of our evaluated systems.

## F Examples

We display randomly sampled examples from the training set in Figure 11 and examples of answers that do not answer the question in Figure 12 (an estimated 5% of the dataset).

To better understand the output of our models, we display example generations randomly sampled from the test set for the multi-task Seq2Seq model (Figure 14) and the Extractive BidAF model (Figure 15). We additionally show a set of poor generations for the multi-task Seq2Seq model (Figure 16) that display a representative set of problems for this abstractive model.

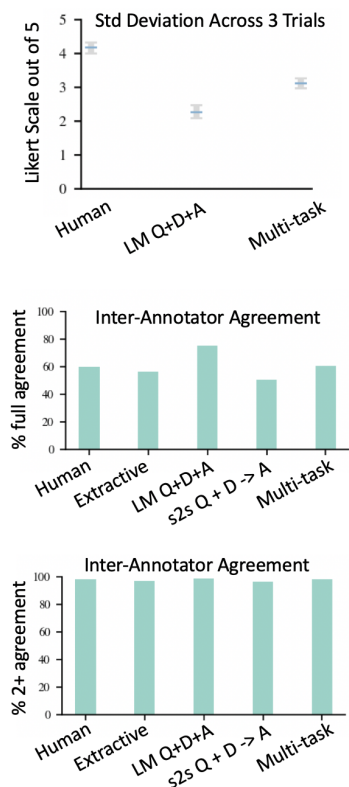


Figure 10: **Analysis of Human Fluency Study** (a) We analyze the variation between three iterations of the same experiment, conducted on three different days. We evaluate the fluency rating given to the human answers, LM answers, and Multi-task model answers, and find low variation across multiple trials.

(b) We calculate the inter-annotator agreement between the three evaluators that score the fluency of each of the models, and find that the % of time all three annotators agree is highest for the language model. Agreement is calculated using positive (Likert scale scores 4 and 5), neutral (3), and negative (scores 1 and 2).

(c) We calculate the inter-annotator agreement between at least two evaluators for the fluency of each of the models, and find that two annotators usually agree almost 100% of the time.

---

**Why do Oreos soften when they get stale, but most other foods (like bread) harden?** It has to do with how much moisture is in the food compared to how much moisture is in the air. Oreos? Drier than air, they gain moisture and turn soft. Loaf of bread? More moist than air, it loses moisture and dries up. Kitchen hack: put a slice of bread in with cookies to keep the cookies fresher a little longer.

**Why is it that cracks seemingly spread on their own?** Because of how the mechanics of the crack work. The stress in a material is related to the radius of corners, that is, rounded corners, like the bottom of the letter U, have less stress than sharp corners, like a V. Sharp corners are actually referred to in mechanical engineering as “stress raisers”. And what’s sharper than the edge of a crack? This means that even stress from thermal expansion of the material can be enough to cause the crack to spread. This is especially true of glass, whose brittle composition is quite susceptible to catastrophic failures.

**Discrete Math: reflexive, symmetric, antisymmetric, transitive** A \*binary relation\* is a function of two variables that is either true or false. I’m going to use  $Rxy$  to notate the relation  $R$  applied to  $x$  and  $y$  in that order. A relation is \*reflexive\* if  $Rxx$  is true; that is, if both variables are the same the relation always holds. A relation is \*symmetric\* if  $Rxy = Ryx$ ; that is, switching the order of the arguments doesn’t change the value. A relation is \*antisymmetric\* if  $Rxy$  never equals  $Ryx$  for distinct  $x$  and  $y$ ; that is, switching the order of the arguments always changes the value (unless the arguments are the same, in which case it obviously can’t). A relation is \*transitive\* if  $Rxz$  is true whenever  $Rxy$  and  $Ryz$  are. Equality is the simplest example of this; if you have  $x = y$  and  $y = z$ , then you can conclude that  $x = z$ .

**Why does bashing my crt tv make it work?** There are several reasons why “percussive maintenance” can work. The most likely is that there’s simply a poor electrical connection somewhere, and banging on it gets it into contact. Once the electricity starts to flow, things heat up and expand a bit, maintaining the connection. Until you turn it off and everything cools off again.

**Is it more efficient to leave the AC running on auto at 74F (in 85 degree whether) or turning it off when leaving the house, and turning it back on when returning when the ambient temp in the apartment is 85?** Turn it off, you will use less power. Thermodynamics tells us that heat loss is proportional to temperature difference, so if you let your house warm up the heat from outside enters more slowly. Essentially the product of time and temperature difference is your cooling energy. There is a minor subtlety with maintenance and life cycle, your AC unit may not be designed for continuous duty, so long cool down cycles could be hard on it. Not sure if that is the case in your unit, seems like a bad way to design anything but it could be. Edit: one non-thermodynamic factor is humidity and mold, which will be different at a constant temperature vs a cycling temperature.

---

Figure 11: Examples of Question-Answer Pairs randomly sampled from the training set

---

**The differences between Wii, PS, Xbox, etc., and which is considered the best system. I’m 40, out of the loop, and have to buy one for Christmas.** If he is 7 go for the Wii its /technically/ more kid friendly and has a more varied option of games for that age range... I think

**What is it when everything just looks really bright/your eyes white out for a moment, then goes back to normal?** What is it? Time to see a doctor.

**Neurologically what happens someone becomes insane** That is waaaaaaaay too broad a question to possibly answer since “insane” covers a multitude of different conditions and disorders, all with their own causes and that’s even assuming that we know the causes in the first place.

**If my spaceship could withstand the pressures of Jupiter, what would a trip through the planet be like?** Read the novel “2010” by Arthur C. Clarke. There’s a chapter that explains this far better than I could.

**If person A saves person B’s life through CPR but person B’s rib gets broken during chest compressions, why is person B able to sue person A? Why doesn’t the judge through the case out of court?** I know Good Samaritan laws play into this somehow, but I can’t remember how. I think that if you have the proper licensing, then they can’t sue you. Or maybe it’s the other way around. Sorry, this response has become useless

---

Figure 12: Examples of Answers in the training set that do not answer the question

---

**Question:** Why do TV shows hide logos ?

**Document:** Why is that? Why do logos get blurred on television, or in film, or even in music videos? The answer, it turns out, is complicated, but is mostly about money in various forms. A whole lot of dysfunction here. Quick Pick: HBO TV Shows Pick the missing word in the title of these HBO shows. September is season premiere month so these logos should be fresh in your mind. TV Shows Venn Diagram II Can you click on the most accurate section of the Venn Diagram for each of the following TV Shows? [...]

**ELI5 Answer:** nothing is free. In most cases, it is a prop for the show, but because apple did NOT pay them for the product placement, the show isn't going to give it away. In other cases, apple may not want their brand used in association with that media.

---

**Extractive:** The answer, it turns out, is complicated, but is mostly about money in various forms. There is only one reason they do not trust it and that is fear of losing their money. All the seals or say reassurance and credibility logos assure your user that a website is safe, protected and personal information is safe. A Lot Of New Disney TV Shows Disney is finally looking to break into the world of TV shows with this streaming service and that's reportedly going to include "episodic spinoffs" of beloved franchises like High School Musical (that's right, it's coming back...), Monsters Inc., and The Muppets.

**Seq2Seq Multi-task:** It's because they're paying for the right to use the show. If you're a TV show, you're paying to have the show shown for free, and you're not paying for a license to use that show for free. the reason they don't have to pay for the license is because the show doesn't want to be associated with the show, and the show isn't paying for it. It's not the show's fault, it's the company that makes the show pay for it, and it's not their fault.

---

Figure 13: Example of document with low document-answer overlap where the abstractive system still achieves a high ROUGE score. The extractive baseline correctly identifies where the question is reformulated in the source document and selects the following sentence, but then goes on to confuse different types of logos and goes off topic entirely with the Disney analysis.

---

**Who controls what the president is allowed to do? But then who controls what they're allowed to do? Etc ..** The president is the commander-in-chief of the armed forces. He is the head of the executive branch. He controls the military, the military and the military. The president controls the army, the navy, the air force, and the army. The President is the president. He has the authority to do whatever he wants, but he doesn't have the power to do all he wants. He doesn't want to do anything about it. He wants to do something about it, but it doesn't really matter what the president is doing. He's not the president, he's just a figurehead, and he's the president of the United States. It's not really a matter of who controls the government, it's the people who control the government. There are a lot of people who don't like the President, but they don't really like him. There's a lot more to it than that, but that's the basic gist of it. The US government is controlled by the US government, but not by the federal government. The federal government is the one that controls most of the US military, but there are a few that don't. They are the ones that control most of US military operations, and they are not the ones who control most.

**What makes the stomach "growl"?** The sound of your stomach growling is caused by the movement of air in your stomach. The sound you hear when your stomach is empty is called peristalsis. This is the sound that your stomach makes when it is empty. When you are hungry your stomach will make the sound of the growling sound, but it will not be as loud as the sound you are making when you are not hungry. It is the same thing that makes your stomach growl when it's empty. You are not actually hungry, but you are actually contracting the muscles in your intestines to make the noise. The growling is the result of the muscles contracting and contracting to make a loud noise. It's the same reason why you feel the urge to eat when you're hungry. Your stomach is trying to empty itself, so it makes the noise as if you aren't hungry. The same thing happens with your stomach when you have a cold. The muscles in the stomach are contracting and making the noise, but they are not contracting. They are contracting, but not contracting, so they aren't contracting. The noise you feel when you feel it is the stomach making is your stomach making the sound. Source : I am a gastroenterologist, and I have a lot of experience with gastroenterology.

**How does a de-humidifier work?** A humidifier works by heating the air in the room. When the air is warm, it cools down. When it is cold, it warms up. This cools the air around it, which cools it down. This is why you can see a lot of cool things in a room if you look at it from a distance. The air around the room is warmer than it is, so it rises. This causes it to rise, and cool down. The same thing happens with a humidifier. The water in the air condenses out of the air, and condenses back into liquid water. This water condenses back to liquid water, which is what you see when you blow on a hot air balloon. The cold air cools down, and the warm air condenses back out of it, making the air colder. It's the same thing as a humidifier, except that it doesn't have any water in it, it just has a bunch of tiny droplets in it that make up the air. It does have some water, but it has a lot more in it than it does with a cold air balloon, so the air gets colder. The downside to a humidifier is that it isn't as efficient as a regular humidifier, and it is much more efficient than a regular one.

**Could anyone explain the suffix nomenclature for diseases/conditions/ailments? Ex. -itis vs -osis vs -algia** There are a few different types of suffix. -ous suffix comes from the Latin "-ary", which means "-ate". The suffix "-er" comes from a Latin root that means "something" or "something." "-ous" is a suffix that is used to refer to things that are related to something that is related to that thing. For example, "-es" means "a". So "-e" is the suffix that comes from "-ide", and "-o" is "-ic". In this case, the "-i" is from the Greek "-os", so "-ia" is just the Latin root of "-ian". It's the same way that "-s" is also the suffix for "-ies", but it's not the same thing as "-y". A lot of the words that are used in the English language don't have the same suffix, so they are used differently. In English, the word "-us" has a different suffix than "-u". This means that the -ic suffix is more likely to be used in English than the -ian suffix . In other words, it's more likely that the suffix will be used more often than it is in English.

**How do live visualization websites work?** I'm not sure what you're referring to, but I'll give it a shot. First of all, you have to understand what they mean by the term "visualization". Basically, they use a technique called "visualization." Basically, you take a picture of a room, look at it, and compare it to what it looks like to you. Then, you look at the room, and see what the room looks like, and then compare that to what you see in the room. You can then use the visualization to figure out what it actually looks like. For example, let's say you're in a room with a camera, and you want to know what the camera looked like. You know that the camera looks like the camera, but you don't know what it looked like, so you use it to look at what it was like to see what you were looking at. Then you use the technique to find out what the person was looking at, and figure out how to use that to get a sense of what they were actually looking at. The visualization is pretty simple, but it's a lot more complicated than that, so I'll try to explain it in a bit more detail. First, you need to understand how the visualization works. The first thing to understand is that it's very easy to use. Let's say that you're looking at a room and you're really interested in the camera. You want to be able to see how far away the camera is from you, so that you can figure out the distance between the camera and your camera. The second thing to remember is that the visualization isn't very accurate. It's just a way of looking at things that aren't directly related to each other. For instance, if you're interested in learning how to read a book, you might be interested in reading a book. You might also want to read the book in a way that is easy to understand, and it's easy to get used to it.

---

Figure 14: Example Question-Generated Answer pairs from Seq2Seq multi-task model, randomly sampled from the test set

---

**Why doesn't the color red compress well in JPEGs?** PNG will take longer to load on a website, but its sometimes simply worth it if it means the difference between a good or bad image. So with all this hate toward JPEGs, you might be asking yourself why the format continues to even exist when so many better options are available. Also important to note, JPEGs do more than compress the file, they also lose color and contrast range. These numbers and ratios are examples for the sake of easy explanation, but lets say a picture has 100 colors and 100 contrast points. Straight out of the camera, JPEGs often look much more vibrant than raw les, because the colors have been enhanced and sharpening applied in-camera. If you need to archive a large number then you could try placing them in a zip file, but you probably won't save more than 5%.

**When reading weather reports and it says 50% chance of rain, what does that actually mean and how is it calculated?** I have always maintained this is a confusing concept and its the main reason that I will rarely if ever use a percent chance in a forecast. When they say there is a 50% chance of rain, does that mean that there is a 50% chance it is not going to rain? Then, why does it always rain when the chance of rain is 50%? So, maybe the 50% chance means that it will rain on only 50% of the land while the other 50% rains on the water. This is important to keep in mind because when making claims about the impact of global warming, you need to look at the big picture, not just the last 150 years. Well, there are two input variables you have to keep in mind: first, the geographic location — where youre looking for a forecast, and second, the time window youre looking at.

**Why does my skin turn a paler white when pressed? Kinda random. Always wanted to know.** There is a darker shade, but the shade Sunkissed is perfect for the lighter skin wearers. It doesn't irritate eyes, and it's gentler on skin than some of their other powders — it's also very finely milled and thick enough that you can use it as a foundation and it covers even dark broken capillaries. What I don't like: This is very light peach when it starts out, and it doesn't turn paler on skin; it also oxidizes. It's a light peach when it starts out, and then it turns darker. If you are unsure if you have cool skin tone, check if you have bluish coloured veins inside your wrist (just under your forehead). Spots or a rash that do not fade under pressure will still be seen when the side of a clear drinking glass is pressed firmly against the skin.

**Can psychoactive drugs trigger schizophrenia in people who are at a genetic disposition to the disease already? If so, how high is the risk for different family members that have it? Do you have a higher chance if a parent has it or if a grandparent has it? What about cousins? Aunts and Uncles?** The identical twin of a person with schizophrenia is most at risk, with a 40 to 65 percent chance of developing the disorder. Some doctors think that the brain may not be able to process information correctly; and it is believed that genetic factors appear to play a role , as people who have family members with schizophrenia may be more likely to get the disease themselves. As Schizophrenia has a tendency to run in families, scientists already know there is a genetic link but that doesnt mean that if you do have someone in your family that has Schizophrenia that you will too, neither does it mean that if you dont, you wont, so there are other factors involved too. At the moment people with Schizophrenia are usually prescribed anti-psychotic medication, some of which can carry unpleasant side effects. If you have a pre-existing risk for schizophrenia (which most people at risk are unaware of), theres a much higher chance that using cannabis will trigger a schizophrenic episode. Again, it is extremely important to note that this risk applies primarily to people who are already at risk of developing schizophrenia.

**Why has the Canadian dollar gone down in value over the few years?** So far in 2016, the Canadian dollar has lost a lot of value. The days of the Canadian dollar at parity with the US dollar are long gone. A lot of that increase in book value is because of the loss of value of the Canadian dollar. What we have to remember however is that it is not really the Canadian dollar that has gone up in value, it is the American dollar which has gone down. Since the beginning of the Iraq War the American economy has stumbled and one could make the argument that it is because there is no one at the wheel. This means Canadians can now come into the US and get a great deal on just about anything , and they are buying up real estate, to the tune of more than \$20 billion a year.

**How do they make Jello with pineapple bits in it if pineapple ruins Jello? I have a store-bought cup of generic Jello with pineapple bits in it. According to all my google searching and middle school science classes, this shouldn't be possible because pineapple enzyme just destroys Jello's jellystuff. So what is this thing in my cupboard?** Enzymes: whats in your spit fruit and jello an hour before the lab begins it is possible to make jello with canned pineapple chunks but not fresh. Well, just as a little science experiment, I stuck a piece of fresh, uncooked pineapple in a cup of the jello, and this is what it looked like after about 20 minutes. In the cooking world, on the other hand, bromelain is regarded as the enemy of the gelatin dessert if you use fresh pineapple in gelatin, the enzyme eats the. Experiment demonstrates denaturation of an enzyme using an enzyme called catalase which is found in the pineapple and jelly experiment background. In this lab students will investigate a few of the different enzymes from our body pineapple contains the enzyme bromelain, and should break down the jello. Well, to begin with, it's not true you can't put any pineapples in jello jello is alive — at least according to a 1974 experiment performed by dr.

---

Figure 15: Example Question-Generated Answer pairs from the extractive BidAF model, randomly sampled from the test set

---

**Why is “PC Gamer” Hardware so Flashy?** It’s because they don’t want to. They want to make sure that they have a decent amount of battery life, and that they can make a decent battery life. They also want to be able to make a small battery that will last for a very long time, and make sure they have enough power to last a long time. That said, there are a lot of people who don’t know what they’re talking about, and don’t really know what’s going on. There’s also the fact that they’re not going to make it to the end of the day, and it’s going to be a lot harder for them to make the battery last longer than it is for the average person. It’s not like they’re going to have to make their own battery last a lot longer than they are for a PC, but they’re still going to get a lot more battery life than they’re for a gaming PC, so it’s a lot easier to make them last longer, and they’re also going to need a lot less power than they were for a desktop PC, which means they’ll need more power to run the game, and so they’ll have to have more power for the game to last longer. That’s why it’s so important to keep in mind that there’s a ton of people out there who aren’t very good at it, and there’s no reason to be afraid of it.

**What are the benefits of an American city being it’s own county? i.e. San Francisco** There are a lot of benefits to an American city being “own” a county. It’s not like there’s a whole lot of money to be made in a city, but it’s a lot more complicated than that. There’s also the fact that there are **lots** of people who aren’t part of the city, and there aren’t **many** people who **are** part of that city. There are also **a lot** more people who don’t want to be part of a city than there are people who just want to live there. There is also a **lot** of **money** in that city, so it’s easier for them to get there than it is for the people who want to stay there. This is why there are so many people in San Francisco, and why it’s so hard to find a city that doesn’t have a whole bunch of people in it. **Edit:** [Here’s a link] (URL\_0) to an article that explains it better than I can. I’m not sure if this is the right place to ask this, but I’m sure someone else can give you a better answer than me, so I’ll give it a shot.

**How do cruise ships get wifi in the middle of the ocean?** There are a couple of ways to do this. One is to turn on the radio and turn off the wifi. The other is to make sure that the wifi is turned on. If you turn the wifi on, the radio will turn on and the wifi will turn off. This is done by turning on the wifi and turning off the radio. If the WiFi is turned off, then the WiFi will turn back on, but the radio won’t turn on. There are also other ways to turn off wifi, such as turning on wifi, or turning off wifi. I’m not sure how to explain this to a 5 year old, but I’ll try to explain it to a five year old. I hope this helps! Edit: I forgot to mention that WiFi is not a wireless network, it is a network of wireless networks. WiFi networks are connected to the internet, but they are not connected to WiFi networks. The internet is connected to a WiFi network, but it isn’t connected to Wifi networks. They are connected by WiFi networks, but WiFi networks can not be connected to wifi networks. So, if you turn on WiFi, the WiFi network will be turned on and off. The WiFi network can be turned off by turning off WiFi, but not by turning it back on.

**How do they know exactly how tall Mt. Everest is?** There are a lot of different ways to measure the height of an object. For example, you can measure the distance from the top of a mountain to the bottom of the mountain, and you can calculate the distance between the top and the mountain. The easiest way to do this is by measuring the angle between the mountain and the ground, and then measuring the height between the ground and the top. You can also measure the angle of the ground from the bottom to the top, and the height from the ground to the ground. The most common method is to use a ruler, which is a ruler that measures the length of the base of a piece of paper. The height of the top is measured in meters, and it’s measured in feet. The simplest way to measure height is by using a ruler. A ruler is a measure of the distance to the base, and a ruler measures the height at which the base is measured. A good ruler is also a good ruler, but it’s not a good one. It’s a good measure of how far away a mountain is, and how far it is away from the base. The same is true for measuring height, but you can’t measure it with a ruler because it’s too far away.

**How/where was the first language invented?** It wasn’t until the late 19th century that the first language was invented. It was just a way for people to communicate with each other and to communicate. For example, if you were to talk to someone, they would talk to you, but you wouldn’t know what they were talking about. If you were talking to them, they’d talk to each other, but they’d never know what you’re talking about, so they wouldn’t be able to tell you what they’re talking to. The same thing happens with languages. For instance, you might say “I’m talking to you”, but if you’re not talking to me, you won’t know how to say it. You might say, “I want to know you,” but you don’t know where you’re going to be talking to, so you might not know what to say. But you might also say “you want to understand me,” and so on. And so on and so forth. Eventually, people started to learn how to communicate, and eventually, they started to figure out how to use their own language to communicate in a way that they couldn’t understand, and then they developed a new language that they didn’t understand. It’s like how you can tell a friend that you’re in the middle of a conversation, and you can’t tell them how to tell them that they’re in a conversation.

---

Figure 16: Example poor quality Question-Generated Answer pairs from the Seq2Seq multi-task model