# Project Name: Multimodal Chain-of-Thought Assignment

Group Members: Kenneth Wang, Kevin Cai, Qingyuan Liu, Sewon Sohn

Kevin: 3035839832

Sewon: 3034884295

Kenneth: 3034634747

Qingyuan: 3034657757

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# Part 0: Abstract

Chain of thought (CoT) refers to the mental process of reasoning and inference that humans use to arrive at an answer or solution to a problem. It involves synthesizing information from multiple sources, making logical connections between ideas, and integrating them into a coherent line of reasoning. In the context of natural language processing and deep learning models, CoT prompting involves generating intermediate reasoning steps to arrive at the final answer to a question. This process allows the model to break down complex questions into simpler, more manageable steps, and to leverage information from multiple sources, such as text and images, to arrive at a more accurate answer. This homework assignment is heavily adapted from this [paper](https://arxiv.org/abs/2302.00923). This homework assignment uses the [Science Question Answering (ScienceQA) dataset](https://github.com/lupantech/ScienceQA), comprising a total of 21,000+ multiple-choice science questions sourced from elementary and high school curricula. Through the [attached notebook](https://colab.research.google.com/drive/19FdymmEqGXbvtW8A5mAN_iPx_IU1VvRm#scrollTo=LqDbUDfSZ6zW), students will explore a subset of this dataset, which consists of questions that only contain a text context as well as questions that have both text and image contexts. This homework assignment will walk you through the sequential steps towards building a multimodal chain-of-thought model that utilizes both text and image inputs and chain-of-thought reasoning to solve ScienceQA problems. In addition, this assignment seeks to simplify the model and build stronger intuition in students by visually showing the students the data as they implement the data loading, tokenizing, and architecture of the model.

# Part I: Copy of all the review comments

**Question 1: Content and Correctness**

Reviewer 1:

There are some small glitches with some variable names (actual\_trainer here when training the language model the first time) as well as some GPU problems with the notebook that prevent the possibility of training the model fully. I'm currently running into RAM issues with the notebook due to the model being so large, so that's a part of the notebook that's not usable. I'm using Google Colab, so this is an issue that should be fixed. Other than that, everything is great!

* Medium improvement

Reviewer 2:

The coding todo question engages with key concepts in the model and it's possible for someone with CS182 level mathematical maturity to under the paper concept, while I think the instruction and hint is not enough.  
Code solutions are provided but with some bugs and errors.  
The full pdf of assignment and separated solution are fully exported, while LaTeX source code is not exported. The problems and solutions are completely correct and engage with the material in a pedagogically useful way.  
Doing the full problem (coding and written parts) would take an average CS 182 student 1.5-2hr to complete.

* Small improvement needed

Reviewer 3:

This project did a great job of writing out all the homework questions and explaining how everything works. All the explanations are very detailed and it was not too difficult to follow along with the code. All the concepts required to understand the assignment are covered (outside of concepts we should know from CS 182 itself). Overall, this part of the project was very well done.

* Excellent work, no actions needed.

**Question 2: Scaffolding**

Reviewer 1:

Implementing the actual model of the notebook may be difficult for people not very familiar with NLP. For example, some of the variable names can be confusing to implement (i.e self.gate\_dense or the use of t5stack) that were not explained in the written portion of the homework. Some more background in each coding section would be useful since it's more architecture dependent rather than actual coding logic dependent.  
The homework was very nicely self contained, and had no dependency issues. However, I did have some hiccups with downloading the dependencies, though this can vary from person to person.  
  
Other than that, I appreciate that there is an autograder embedded in the notebook that checks answers! That was very nice. Furthermore, the code explaining the intuitions behind each portion of the dataset as well as the architecture was a nice touch.

* Small improvement needed

Reviewer 2:

Project provides some scaffolding for a student to engage with the material. Some of the code that students are expected to implement is explained not clearly through the use of text cells in Jupyter and/or code comments. The comments and text cells in Jupyter are too few  
The project group uses some external packages and requires students to use in the todo part, but there is no tutorial link and instruction on how to use them.

There are autograding tests/sanity checks to make sure a student’s implementation is correct.

* Medium improvement needed

Reviewer 3:

This is the part of the project that I think needs the most work. For the most part this project does have the scaffolding it needs. It provides a very large amount of scaffolding code that handles most of the model creation, data-loading, etc....  
  
However, the way that the data downloading is handled currently does not work. When running the code blocks that should be downloading the data and pretrained models, I get this error: """  
Access denied with the following error:  
  
Cannot retrieve the public link of the file. You may need to change  
the permission to 'Anyone with the link', or have had many accesses.  
  
You may still be able to access the file from the browser:"""  
  
It seems that the files in the google drive do not have the proper access rights, meaning that the provided code doesn't work. Furthermore, when trying to manually fix this by moving the files manually, it seemed to not always recognize that the files were in fact in the google drive folder.  
  
Furthermore, every time I restarted the runtime, it might sometimes mess up the google drive folder (idk, how or why) and it wouldn't recognize files there anymore, and forced me to start from the beginning. Overall, I think I wasted like two hours fighting with google drive before finally giving up when I just could not get the pre-trained model weights to recognize.  
  
For the most part, this bad experience seems to be due to the buggy-ness of google drive and google colab notebooks and how they handle runtime state, and not due to anything the project authors did. However, improper file access rights still are an issue and ended up causing a lot of trouble. Please make sure to fix this, and actually do a full "end-to-end" test of your code on another google account to make sure everything works. This error could have been caught.

* Medium improvement needed

**Question 3: Readability/Clarity**

Reviewer 1:

The homework assignment is clear and concise, with very intuitive examples and explanations, though this is to be expected with an NLP task since there isn't much mathematical background needed. However, when discussing the fusion of the language encoder and the vision extractor, it would be helpful to explain the meaning of the lambda here, like explaining what happens when the lambda is big or small here, and why we use such an equation. It would be nice if there were more mathematical based problems that furthered the understanding of the architecture (for example perhaps some analysis on the loss function, or the derivatives of the model).

* Small improvement needed

Reviewer 2:

The HW assignment and commentary are easy to read and follow. Any mathematical notation used is understandable, and any non-standard notation is clearly explained. The assignment and commentary are free of spelling/grammar errors.

* Excellent work, no actions needed.

Reviewer 3:

In this category the project is fantastic. I found all the explanations very well made and the formatting on all documents was very clear. The commentary talks about the entire assignment and does a good job of explaining each component.

* Excellent work, no actions needed.

**Question 4: Commentary on HW**

Reviewer 1:

The commentary is excellent, giving a brief overview of the goals of the entire assignment. The commentary clearly outlines which parts of the assignment do what, and how it connects it to what we've learned in class in both the notebook and the written assignment. One thing I might add to improve is just section dividers as well as section titles, to make it clearer which part the reader is reading about, but other than that the commentary on the assignment is excellent.

* Excellent work, no actions needed.

Reviewer 2:

Project has a 2-3 page commentary on the HW that explains the key concepts in the paper and how the assignment engages with them.  
The key concepts that are engaged with in the HW assignment are explained briefly in the commentary. The commentary is good enough for students to understand and helpful.

* Excellent work, no actions needed.

Reviewer 3:

The commentary, like the rest of the project, is well written and explains everything it needs to very well.

* Excellent work, no actions needed.

**Question 5: Going above and beyond**

Reviewer 1:

This submission simplifies the paper well, making the process very intuitive for this model architecture. (i.e have 1 model output reasoning, then have another model use that reasoning to create an answer). This was very useful for somebody who has never heard of this model before, so having clear intuition is important here. Furthermore, this submission uses handpicked examples in code to push the intuition of the model architecture (I especially enjoyed the bottom feeders example). Well Done!

* Small improvement needed

Reviewer 2:

Creating synthetic datasets that illustrate certain ideas  
Simplifying the paper  
Not enough visualization

* Small improvement needed

Reviewer 3:

This project does have a section at the end where the student is able to get a deeper understanding of the material through visualizations. However, due to technical issues with the scaffolding and data fetching explained above, I was not able to run this section of the code, and therefore, cannot properly review it.

* Excellent work, no actions needed.

# Part II: Point-by-Point Response to Review Comments

**Question 1: Content and Correctness**

Reviewer 1:

There are some small glitches with some variable names (actual\_trainer here when training the language model the first time) as well as some GPU problems with the notebook that prevent the possibility of training the model fully. I'm currently running into RAM issues with the notebook due to the model being so large, so that's a part of the notebook that's not usable. I'm using Google Colab, so this is an issue that should be fixed. Other than that, everything is great!

* Medium improvement

The variable names are now set correctly and the eval set size is reduced to 20 from 424 in order to fix the GPU and RAM issues. We also tested the notebook on different google accounts and different laptops several times in order to ensure it would work for another student.

Reviewer 2:

The coding todo question engages with key concepts in the model and it's possible for someone with CS182 level mathematical maturity to under the paper concept, while I think the instruction and hint is not enough. Code solutions are provided but with some bugs and errors.  
The full pdf of assignment and separated solution are fully exported, while LaTeX source code is not exported. The problems and solutions are completely correct and engage with the material in a pedagogically useful wayDoing the full problem (coding and written parts) would take an average CS 182 student 1.5-2hr to complete.

* Small improvement needed

More instructions and hints are provided with documentation links to certain packages and additional scaffolding. Specifically, scaffolding for gated fusion mechanism is now included under part d and a documentation to T5Stack is linked in the notebook. For more background of the architecture, we included a computational graph. Bugs and errors have been fixed by the solution stated above (Reviewer 1). LaTex source code has been created for the final submission.

Reviewer 3:

This project did a great job of writing out all the homework questions and explaining how everything works. All the explanations are very detailed and it was not too difficult to follow along with the code. All the concepts required to understand the assignment are covered (outside of concepts we should know from CS 182 itself). Overall, this part of the project was very well done.

* Excellent work, no actions needed.

◡̈

**Question 2: Scaffolding**

Reviewer 1:

Implementing the actual model of the notebook may be difficult for people not very familiar with NLP. For example, some of the variable names can be confusing to implement (i.e self.gate\_dense or the use of t5stack) that were not explained in the written portion of the homework. Some more background in each coding section would be useful since it's more architecture dependent rather than actual coding logic dependent.  
The homework was very nicely self contained, and had no dependency issues. However, I did have some hiccups with downloading the dependencies, though this can vary from person to person.  
  
Other than that, I appreciate that there is an autograder embedded in the notebook that checks answers! That was very nice. Furthermore, the code explaining the intuitions behind each portion of the dataset as well as the architecture was a nice touch.

* Small improvement needed

Again, more instructions are provided with documentation links to certain packages and additional scaffolding to gated fusion mechanism and T5Stack. Scaffolding for gated fusion mechanism is now included under part d “gated fusion mechanism” and a documentation to T5Stack is linked in the notebook. For more background of the architecture, we have created and included a visualization in Figure 4 where it shows the end-to-end model in more detail to help drive understanding. For dependency issues, we have tested the notebook on different google accounts and different laptops several times in order to ensure it would work for another student.

Reviewer 2:

Project provides some scaffolding for a student to engage with the material. Some of the code that students are expected to implement is explained not clearly through the use of text cells in Jupyter and/or code comments. The comments and text cells in Jupyter are too few  
The project group uses some external packages and requires students to use in the todo part, but there is no tutorial link and instruction on how to use them.

There are autograding tests/sanity checks to make sure a student’s implementation is correct.

* Medium improvement needed

Again, more instructions are provided with documentation links to certain packages and additional scaffolding. Specific ways we have done that have been stated above (Reviewer 1; Question 1 Reviewer 2)

Reviewer 3:

This is the part of the project that I think needs the most work. For the most part this project does have the scaffolding it needs. It provides a very large amount of scaffolding code that handles most of the model creation, data-loading, etc....  
  
However, the way that the data downloading is handled currently does not work. When running the code blocks that should be downloading the data and pretrained models, I get this error: """  
Access denied with the following error:  
  
Cannot retrieve the public link of the file. You may need to change  
the permission to 'Anyone with the link', or have had many accesses.  
  
You may still be able to access the file from the browser:"""  
  
It seems that the files in the google drive do not have the proper access rights, meaning that the provided code doesn't work. Furthermore, when trying to manually fix this by moving the files manually, it seemed to not always recognize that the files were in fact in the google drive folder.  
  
Furthermore, every time I restarted the runtime, it might sometimes mess up the google drive folder (idk, how or why) and it wouldn't recognize files there anymore, and forced me to start from the beginning. Overall, I think I wasted like two hours fighting with google drive before finally giving up when I just could not get the pre-trained model weights to recognize.  
  
For the most part, this bad experience seems to be due to the buggy-ness of google drive and google colab notebooks and how they handle runtime state, and not due to anything the project authors did. However, improper file access rights still are an issue and ended up causing a lot of trouble. Please make sure to fix this, and actually do a full "end-to-end" test of your code on another google account to make sure everything works. This error could have been caught.

* Medium improvement needed

We handled the data downloading issue with gdown by converting our files into zip files. This solved the issue of downloading features and weights for the notebook. We also moved the ‘download dependencies’ part, where students would have to restart the runtime, to the beginning of the notebook so that they wouldn’t have to run the prior cells again after having to restart the runtime. We also tested the notebook on different Google accounts and different laptops several times in order to ensure it would work for another student. In addition, we’ve added extra comment instructions to the gdown parts (since gdown seems to break for various reasons, not related to correct permissions sharing) that direct students to manually download the data if needed.

**Question 3: Readability/Clarity**

Reviewer 1:

The homework assignment is clear and concise, with very intuitive examples and explanations, though this is to be expected with an NLP task since there isn't much mathematical background needed. However, when discussing the fusion of the language encoder and the vision extractor, it would be helpful to explain the meaning of the lambda here, like explaining what happens when the lambda is big or small here, and why we use such an equation. It would be nice if there were more mathematical based problems that furthered the understanding of the architecture (for example perhaps some analysis on the loss function, or the derivatives of the model).

* Small improvement needed

More scaffolding for gated fusion mechanism is included in the pdf file under part d “gated fusion mechanism”. We also created a visualization of the model including the gated fusion model to better help students.

Reviewer 2:

The HW assignment and commentary are easy to read and follow. Any mathematical notation used is understandable, and any non-standard notation is clearly explained. The assignment and commentary are free of spelling/grammar errors.

* Excellent work, no actions needed.

◡̈

Reviewer 3:

In this category the project is fantastic. I found all the explanations very well made and the formatting on all documents was very clear. The commentary talks about the entire assignment and does a good job of explaining each component.

* Excellent work, no actions needed.

◡̈

**Question 4: Commentary on HW**

Reviewer 1:

The commentary is excellent, giving a brief overview of the goals of the entire assignment. The commentary clearly outlines which parts of the assignment do what, and how it connects it to what we've learned in class in both the notebook and the written assignment. One thing I might add to improve is just section dividers as well as section titles, to make it clearer which part the reader is reading about, but other than that the commentary on the assignment is excellent.

* Excellent work, no actions needed.

◡̈

Reviewer 2:

Project has a 2-3 page commentary on the HW that explains the key concepts in the paper and how the assignment engages with them.  
The key concepts that are engaged with in the HW assignment are explained briefly in the commentary. The commentary is good enough for students to understand and helpful.

* Excellent work, no actions needed.

◡̈

Reviewer 3:

The commentary, like the rest of the project, is well written and explains everything it needs to very well.

* Excellent work, no actions needed.

◡̈

**Question 5: Going above and beyond**

Reviewer 1:

This submission simplifies the paper well, making the process very intuitive for this model architecture. (i.e have 1 model output reasoning, then have another model use that reasoning to create an answer). This was very useful for somebody who has never heard of this model before, so having clear intuition is important here. Furthermore, this submission uses handpicked examples in code to push the intuition of the model architecture (I especially enjoyed the bottom feeders example). Well Done!

* Small improvement needed

We’re not sure why this was graded “small improvement needed” as the comment does not address any point of improvement. We assume this is an overall summary (i.e. we need small improvement overall to address other parts they mentioned) or a misclick. However, we’ve added improvements on this front (“Going above and beyond”) in response to another reviewer’s comments.

Reviewer 2:

Creating synthetic datasets that illustrate certain ideas  
Simplifying the paper  
Not enough visualization

* Small improvement needed

This comment is not very clear what it means. We are interpreting this comment as saying, “Your team hit the rubric guidelines by ‘creating synthetic datasets’ and ‘simplifying the paper’, but your team has some small improvement needed on the ‘visualization’ front.” The other interpretation would be that we need to work on “synthetic datasets, simplifying the paper, and adding visualizations”, which seems like an unlikely interpretation.

As such, the place we needed to improve on was to add more visualizations. Since our model works with language and images, we thought it would be more appropriate to add visualizations by showing students what the data actually looks like throughout the model (as we’ve done so in parts a, b, and c, where students generate prompts, show the questions and image inputs, and also see the tokenized data). In response to this review comment, we’ve added parts e, f, and g, where the students use their trained model to generate rationales and see how experimentation or changes to the inputs to the models may change the rationales. The point of this is for the students to visualize how the two-stage framework model actually works, and what the rationales (which are in the ‘middle’ of the entire model) look like, especially in response to certain perturbations. In addition, we’ve gone beyond the paper and have added visualization of the actual model from end-to-end by drawing out the simplified computational graph.

Reviewer 3:

This project does have a section at the end where the student is able to get a deeper understanding of the material through visualizations. However, due to technical issues with the scaffolding and data fetching explained above, I was not able to run this section of the code, and therefore, cannot properly review it.

* Excellent work, no actions needed.

The technical problems are addressed above and have been resolved.

# Part III: Final submission

# Part IV: Team member contribution

Kevin worked on the trainin of the model, including the pretraining part. He also worked on implementing the notebook.

Qingyuan worked on fully understanding the paper and translating the code to the notebook.

Sewon and Kenneth worked on the LaTeX, the Homework PDF, making figures and diagrams, and adding scaffolding and commenting throughout the notebook.

Everyone worked together well and filled in when needed!

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# Part V: Link to code/dataset/supplementary materials

Notebook is linked here: <https://colab.research.google.com/drive/19FdymmEqGXbvtW8A5mAN_iPx_IU1VvRm#scrollTo=np0vbrioX2rY>

Everything needed is linked here.

In addition, here is the Google Drive (where we’ve uploaded a Zip file containing all our required things).

<https://drive.google.com/drive/u/0/folders/1gxFmhNqXB79-zCK7uLL2sDZ_w7Lue65->