

HOMWORK 2

PRACTICES OF LARGE LANGUAGE MODELS¹

CS 678 ADVANCED NATURAL LANGUAGE PROCESSING (FALL 2024)

<https://nlp.cs.gmu.edu/course/cs678-fall24/>

OUT: Friday, September 20, 2024

DUE: Friday, October 11, 2024

Your name: Bhavana Kendhyala

Your GID: G01445952

IMPORTANT: The homework is accompanied by a Python notebook and a GitHub code repository. **For Part 1**, after copying this notebook to your Google Drive or One Drive, please paste a link to it below. To get a publicly accessible link, hit the Share button at the top right, then click “Get shareable link” and copy over the result. **For Part 2**, fork the public code repository to your own space and share a link to your code repository in the link below. When you are done, submit your completed .pdf report to Gradescope. **If you fail to do this, you will receive no credit for this homework!**

Note 1: You will only need to run lightweight code for this assignment, which can be done locally on your machine (i.e., ORC GPUs are not required).

Note 2: Both parts require using OpenAI API. Please make sure to **REMOVE** your private key from your notebook and code repository before submission.

Your Notebook solution for Part 1:

<https://colab.research.google.com/drive/1IBbtF5VgSDv-uEaj78AHiaKcNjQj61i1>

Your GitHub code repository for Part 2: <https://github.com/Bhavana0810-dev/HW2-files>

Graded Questions: 100 points

Bonus Questions: 20 points

Total Points Available: 120/100

Additional Notes:

- Upload the whole report PDF (including all pages, also if you’re scanning it).
- All questions only require written answers in this PDF. For coding problems, you will have to fill out all code blocks that say `YOUR CODE HERE` in your notebook or per instruction.
- For text-based answers, you should replace the text that says “Write your answer here...” with your

¹Compiled on Sunday 6th October, 2024 at 09:17

actual answer.

Part 1: Introduction to Prompts [50 pts]

This section uses the paired Python notebook. Below, provide the answers to the questions as generated by the completed code blocks in the similarly marked questions. Please also make sure to set up the Python environment (Task 0) before you run the notebook.

Task 1: Story Generation with Different Sampling Strategies

1. (5 points) Can you use the ChatCompletion function to generate a story about an Indian student studying abroad (e.g., at George Mason University)? Please use the default setting and generate only one story.

Give the output of your gpt-4o-mini here. Response: Title: A Journey of Dreams

Arjun Mehta stood in front of the sprawling campus of George Mason University, the crisp autumn air swirling around him. He had arrived a week ago from his hometown of Pune, India, and everything still felt like a dream. The vibrant colors of the leaves contrasted with the red brick buildings, and he felt both exhilarated and apprehensive. The journey that had brought him here was both long and arduous, yet filled with hope—a desire to pursue his studies in computer science, a field he had been passionate about since childhood.

Back in Pune, Arjun had spent countless hours coding, diligently practicing for the entrance exams, and dreaming of studying abroad. When he received his acceptance letter from George Mason, it felt surreal. It was a testament to his hard work, and he was determined to make the most of it.

On his first day of classes, Arjun walked into the lecture hall, filled with students from diverse backgrounds. As he took a seat, he noticed a girl with curly hair and a friendly smile sitting next to him. She introduced herself as Mia, a sophomore majoring in psychology. They quickly struck up a conversation, and Mia's warmth helped ease Arjun's initial shyness. They shared stories of their hometowns, cultures, and dreams.

"Studying here is amazing," Mia said, her eyes sparkling with enthusiasm. "You'll love it! Just remember to join some clubs. It's a great way to meet people."

Arjun nodded, excited by the prospect. However, as the classes progressed, he found himself struggling to keep up. The teaching style was different—more interactive, with discussions and group projects—and he often felt overwhelmed by the pace. There were moments during lectures when he would zone out, grappling with the new concepts while trying to suppress the nagging fear of being left behind.

Determined not to let these obstacles defeat him, Arjun turned to the resources available on campus. He spent hours in the library and attended tutoring sessions, gradually finding his footing. He joined a coding club where he met other students who were passionate about technology. One evening, while discussing an upcoming hackathon, his confidence began to bloom. The excitement was infectious, and he realized he was not alone in his struggles; everyone was adapting in their own way.

As the semester progressed, Arjun formed lasting friendships. He spent weekends exploring Washington, D.C., with Mia and other friends, visiting iconic landmarks like the Lincoln Memorial and indulging in local cuisine. One chilly afternoon, they huddled together in a cozy café, sipping hot chai lattes. Arjun shared stories of Diwali celebrations back home, and his friends listened intently, intrigued by the traditions he described.

Then came the day of the hackathon. Arjun felt a rush of adrenaline as he and his team brainstormed ideas. He proposed developing an app that helped students find study partners based on their schedules and subjects. With his coding skills and the team's creativity, they worked tirelessly over 24 hours, fueled by energy drinks and bleary excitement.

When the time came to present their project, Arjun's heart raced. He stepped up to speak, sharing their app's potential impact on improving academic collaboration. To his surprise, the judges were impressed. Although they didn't win first place, they received an honorable mention, and that recognition filled Arjun with pride.

By the time the semester came to an end, Arjun had not only adapted to his new life but had thrived. He realized how far he had come since that first day at George Mason. He had gained resilience, built a supportive circle of friends, and embraced the diverse culture around him. The experiences that had once felt daunting now became cherished memories.

As winter break approached, Arjun looked out over the campus covered in a fresh blanket of snow. He thought about his family back in India and how their sacrifices had brought him here. He vowed to make them proud by giving his best in his studies and continuing to chase his dreams. The journey was just beginning, filled with possibilities. With a heart full of gratitude, he knew that George Mason University would always hold a special place in his story—a chapter of growth, friendship, and the pursuit of passion.

2. (5 points) Now, can you do the same but try to get 2 generation outputs with "top_p" set to be 1?

The call returns 2 responses:

Response 0: **Title: A Journey Beyond Borders**

Aarav Sharma stood at the bustling Dulles International Airport, feeling a thrilling mix of excitement and nerves coursing through him. Arriving from India, he was on the cusp of a new chapter in his life as he set foot in the United States to pursue his Master's degree in Environmental Science at George Mason University. His journey hadn't been easy; it had been filled with late-night study sessions, countless applications, and the bittersweet farewells of friends and family back home. Aarav could still hear his mother's voice, reminding him to call often, and his father's encouraging words about how this was a tremendous opportunity to explore the world and all that it had to offer.

As Aarav stepped out into the crisp air of Virginia, a wave of reality washed over him. The sprawling campus of George Mason, with its modern buildings and lush green spaces, felt both intimidating and inviting. The echoes of laughter and conversations from groups of students mingled with the rustling of leaves, creating a symphony of college life that he couldn't wait to be a part of.

His first few weeks were a whirlwind of orientation sessions, finding his way around campus, and meeting new people. Aarav quickly found himself gravitating towards a group of diverse students from different backgrounds, united by a shared passion for environmental issues. Among them was Maya, a spirited girl from Brazil with an infectious energy, and James, a thoughtful young man from South Africa who had an impressive depth of knowledge about sustainability.

One afternoon, in the library, Aarav sat surrounded by textbooks and notes, feeling the weight of the world's problems pressing down on him. As he looked around, he noticed Maya and James poring over a shared laptop, animatedly discussing a project on renewable energy. Drawn by their enthusiasm, he approached them, and within minutes, they were deep in conversation, exchanging ideas and perspectives.

"Why don't we work together on this project?" Maya suggested, her eyes sparkling. "We can combine our strengths and create something impactful!"

That project turned into an exploration of their combined cultural insights and academic knowledge, merging Indian, Brazilian, and South African perspectives on sustainability. They poured their hearts into the research, spending countless nights finalizing presentations and prepping for discussions. Aarav thrived in this new collaborative environment, feeling a deep sense of belonging.

One weekend, they decided to visit Washington, D.C., to gain inspiration from the heart of American politics and environmental advocacy. As they wandered the majestic National Mall, Aarav was struck by the vastness of the world and the interconnectedness he often theorized about in class. They visited museums and monuments, discussing the significance of each visit over steaming cups of coffee at a nearby café.

"Learning about these global issues makes me realize we're all in this together," Aarav said, looking out over the horizon, where the Lincoln Memorial stood proudly against a backdrop of the setting sun.

As the semester progressed, the trio's friendship deepened. They celebrated Diwali at Aarav's apartment, lighting diyas and sharing sweets. In turn, Aarav participated in a vibrant Carnival celebration at Maya's place and learned the rhythms of South African music during a cultural night hosted by James's family. These experiences enriched Aarav's life in ways he never imagined, blending their traditions into a colorful tapestry of friendship.

Weeks turned into months, and the end of the semester arrived. Their hard work paid off as they presented their project at an environmental conference on campus, receiving accolades and interest from local organizations. Standing on stage, Aarav took a deep breath, confidence surging within him.

"Together, we have shown that our diverse backgrounds only strengthen our solutions to global issues," he said, recalling the various cultures that had shaped their minds. The applause that followed was a testament to not only their work but also the friendships and connections they had formed along the way.

As the year came to a close, Aarav reflected on how much he had grown. He was no longer just a student from India; he was now a global citizen, enriched by the experiences and friendships he had built at George Mason.

On the final night before his return home, Aarav gathered with Maya and James, reminiscing about their journey. "This isn't just the end; it's the beginning of a much larger adventure," he said, raising a toast with homemade chai.

With the sun setting, painting the sky in hues of orange and purple, Aarav felt a profound sense of gratitude. He had arrived in a foreign land as an uncertain student and was now returning home with a heart full of memories, friendships, and a renewed passion for making a difference in the world. The journey beyond borders had not only transformed his studies but had also journeyed deeply into his spirit, igniting a flame that he would carry with him long after he left George Mason.

——— *Response 1*: **Title: A Journey Across Continents**

Aarav Patel stood on the campus of George Mason University, a slight flutter of excitement and nervousness in his stomach. Having left his hometown of Ahmedabad, India, just a few months earlier, he was now experiencing the vibrant culture and academic rigor of a foreign land. The sprawling green lawns, red-brick buildings, and the hum of American student life were all too different yet exhilarating.

Aarav had come to study computer science, fueled by dreams of innovation and technology. His parents had worked hard to provide him with this opportunity, and he wanted to make them proud. Yet, with the excitement also came the weight of responsibilities and the fear of the unknown.

The first few weeks were a blur of orientation events, new faces, and a plethora of information to absorb. Aarav made quick friends with fellow international students from Brazil, South Korea, and Nigeria. They bonded over their shared experiences of homesickness and the challenges of adapting to a new culture. His roommate, Miguel, was an aspiring musician who often strummed his guitar late into the night, filling their dorm with soft melodies.

As classes began, Aarav quickly found a rhythm in his studies. He was captivated by his computer programming classes and was particularly drawn to artificial intelligence. It was a field that spurred his creativity, and he often spent late nights coding in the lab, losing track of time as he immersed himself in projects. However, the academic pressure sometimes felt overwhelming, and he occasionally missed the comfort of home-cooked meals and the sound of his mother's laughter.

To ease his homesickness, Aarav discovered a local Indian grocery store where he could buy spices and ingredients to recreate his favorite dishes. One weekend, he invited his friends over for a small dinner party. The fragrance of masala chai and biryani wafted through the apartment, drawing curious neighbors and bringing everyone together in a delightful mix of cultures. That night, laughter filled the room as they shared stories from their homelands, reminding Aarav that, despite the geographical distance, they were all connected in their aspirations and dreams.

One Saturday morning, Aarav woke up to a text from Miguel suggesting a hike to Great Falls Park. Although initially hesitant, he agreed, eager for a change of scenery. The majestic waterfalls and the calming sounds of nature provided a welcome relief from the chaos of studies. They laughed and joked, capturing moments with selfies, and for the first time in a while, Aarav felt a sense of freedom that transcended his worries about exams and projects.

As the semester progressed, Aarav faced challenges, both academically and personally. A particularly tough project left him feeling stressed, and he began to question if he was cut out for this rigorous academic environment. During one of his more challenging days, he visited the university's counseling center where a kind advisor reminded him of the story of his own journey. "It's not about how fast you run, but how resilient you are," she said, letting him know that seeking help was a sign of strength, not weakness.

Inspired, Aarav refocused his efforts and began studying in study groups, which not only helped him academically but also strengthened his friendships. He learned to embrace imperfections, realizing that every student in his program was on a unique journey of growth and learning.

As spring arrived, Aarav found himself more confident and immersed in the campus life. He joined the South Asian Student Association, where he participated in cultural events and even performed a traditional dance at the annual festival. The applause from the audience filled him with joy, showcasing the beauty of his culture abroad.

By the time summer rolled around, Aarav had secured an internship at a tech startup in Arlington, just a short train ride from campus. It was a dream come true, and he felt the culmination of his hard work and dedication. As he walked through the bustling office filled with innovation and creativity, a sense of belonging washed over him.

On the last day of his internship, his boss, recognizing Aarav's talent and dedication, offered him a chance to continue working part-time while finishing his studies. Excitement bubbled inside him; his dreams were unfolding, and he was carving his own path in a world that initially felt so foreign.

As he strolled across the university campus one final evening before summer break, Aarav looked up at the stars shining brightly overhead and smiled. The journey of studying abroad had shaped him into a more adaptable, resilient individual. He had forged friendships, embraced new cultures, and discovered his own strength.

In that moment, Aarav realized that home wasn't just a place on a map—it was the blend of experiences and connections he had created along the way, a beautiful tapestry of cultures woven together by shared stories and dreams. With every step he took, he knew he was not just studying abroad; he was building a future filled with endless possibilities. ———

3. (5 points) How about 2 generation outputs with “top_p” set to be 0.5?

The call returns 2 responses:

Response 0: **Title: A Journey Beyond Borders**

Riya Sharma stood at the entrance of George Mason University, her heart racing with a mix of excitement and nervousness. Having traveled over 8,000 miles from her hometown in Pune, India, to Fairfax, Virginia, she felt like a tiny fish in a vast ocean. The sprawling campus buzzed with students from all over the world, each one seemingly confident and at ease in their surroundings. Riya took a deep breath, adjusted her backpack, and stepped forward into her new life.

The first few weeks were a whirlwind of orientation sessions, new classes, and meeting people from diverse backgrounds. Riya was pursuing a Master's in Information Technology, a field she had always been passionate about. Back in India, she had excelled in her undergraduate studies, but the academic environment at George Mason was unlike anything she had experienced before. The professors encouraged open discussions, and students were expected to contribute their thoughts freely. Riya found herself grappling with the differences in teaching styles and cultural expectations.

One afternoon, as she sat in the campus library, poring over her notes, a friendly voice interrupted her concentration. "Hey, mind if I join you?" It was Jake, a tall student with tousled hair and an easy smile. He was studying computer science and had noticed Riya's focused demeanor. They struck up a conversation, and Riya soon found herself laughing and sharing stories about her life in India. Jake was curious about her culture, and Riya enjoyed explaining the vibrant festivals, the delicious food, and the rich history of her homeland.

As weeks turned into months, Riya and Jake became close friends. They explored the city together, visiting historical sites like the National Mall and indulging in the local cuisine. Riya introduced Jake to the spicy flavors of Indian food, while he introduced her to the joys of American barbecue. They often studied together, and Jake helped Riya navigate the intricacies of group projects, where she sometimes felt overshadowed by her more outspoken classmates.

However, not every day was easy. Riya struggled with homesickness, especially during the Diwali festival, which she celebrated with her family back home. The vibrant lights, the sweet treats, and the laughter of loved ones felt so far away. To lift her spirits, Jake surprised her with a small Diwali celebration in his apartment. He decorated the space with fairy lights, and they made a simple feast together, complete with samosas and chai. Riya felt a warmth in her heart as she shared her traditions with Jake, realizing that friendship could bridge the gap between cultures.

As the semester progressed, Riya's confidence grew. She began to participate more actively in class discussions and even joined a student organization focused on technology and innovation. Her unique perspective as an international student brought fresh ideas to the group, and she found herself thriving in the collaborative environment.

By the time finals approached, Riya felt a sense of belonging she hadn't anticipated. She had formed a supportive network of friends, both international and American, who encouraged one another through the challenges of academia. The late-night study sessions, the shared meals, and the laughter created a tapestry of memories that Riya cherished.

On the last day of finals, Riya and Jake celebrated their achievements with a picnic at a nearby park. As they sat on the grass, enjoying the warmth of the sun, Riya reflected on her journey. She had come to George Mason as a shy student, unsure of herself in a foreign land, but she was leaving with newfound confidence and a deeper understanding of herself and the world around her.

"Thank you for being such a great friend, Jake," Riya said, her eyes sparkling with gratitude. "I couldn't have done this without you."

Jake smiled, "And thank you for sharing your culture with me. You've made this year unforgettable."

As they packed up their picnic, Riya realized that her journey was just beginning. She had learned to embrace the challenges of studying abroad, and she was excited about the future. With her heart full of hope and her mind brimming with ideas, Riya knew that she would carry the lessons of her time at George Mason with her, no matter where life took her next. ———

Response 1: **Title: A Journey Beyond Borders**

Aarav Sharma stood at the entrance of George Mason University, a sprawling campus nestled in the heart of Virginia. The crisp autumn air filled his lungs as he took in the vibrant colors of the leaves, a stark contrast to the tropical greens of his hometown in India. It was a dream come true for him to study abroad, but it also felt like stepping into a new world.

Aarav had always been fascinated by the United States. Growing up in Pune, he spent countless nights watching Hollywood movies and reading about American culture. He was determined to pursue a degree in Computer Science, and after months of hard work, he had secured a scholarship to George Mason. As he adjusted his backpack, he felt a mix of excitement and anxiety about what lay ahead.

On his first day, Aarav attended an orientation session for international students. He was greeted by a diverse group of students from all over the world, each with their own stories and dreams. As he introduced himself, he felt a sense of camaraderie with his peers. They were all navigating the same challenges of being far from home, and it was comforting to know he wasn't alone.

Classes began, and Aarav quickly found himself immersed in a rigorous academic environment. His professors were knowledgeable and approachable, and he was particularly inspired by Dr. Patel, a fellow Indian who taught Artificial Intelligence. Dr. Patel encouraged Aarav to think critically and creatively, pushing him to explore ideas beyond the textbook. Aarav felt a spark of passion igniting within him, and he began to envision a future where he could contribute to technology in meaningful ways.

Outside of classes, Aarav joined several student organizations, including the Indian Student Association. Through this group, he connected with other Indian students who shared his culture and experiences. They celebrated festivals like Diwali and Holi, bringing a piece of home to the campus. Aarav felt a sense of belonging as they shared stories of their families and traditions, and he found comfort in the familiar flavors of home-cooked meals.

However, not everything was easy. Aarav faced challenges adjusting to the American education system. The emphasis on participation and discussion was different from the rote learning he was used to. There were moments of self-doubt when he struggled to express his ideas in class, but he reminded himself that growth often comes from discomfort. With each passing week, he became more confident, engaging in discussions and sharing his unique perspective.

One chilly evening, Aarav decided to take a walk around the campus. The glow of the streetlights illuminated the path as he strolled past the library and the student center. He spotted a group of students gathered around a fire pit, laughing and roasting marshmallows. Drawn by the warmth of their camaraderie, he approached them. They welcomed him with open arms, and soon he found himself sharing stories about India, his family, and his aspirations.

As the weeks turned into months, Aarav's life at George Mason flourished. He made friends from different backgrounds, learned to navigate the complexities of American culture, and even took up hiking in the nearby national parks. Each experience broadened his horizons and enriched his understanding of the world.

Midway through the semester, Aarav faced a pivotal moment. He was invited to present a project he had been working on in Dr. Patel's class at a regional tech conference. The project involved using machine learning to improve healthcare accessibility in rural India. Aarav poured his heart into the presentation, determined to make an impact.

On the day of the conference, he stood nervously in front of a room full of experts and peers. As he spoke, he saw nods of understanding and interest from the audience. When he finished, the applause was thunderous. Aarav felt a rush of pride and accomplishment. He realized that he had not only represented his ideas but also his culture and his homeland.

By the end of the academic year, Aarav had transformed. He had grown from a shy student into a confident individual ready to take on the world. He forged lasting friendships, gained invaluable knowledge, and developed a deeper appreciation for diversity. As he prepared to return to India for the summer, he reflected on his journey with gratitude.

Standing at the same entrance where it all began, Aarav took a moment to breathe in the fresh air. He knew that this experience had changed him forever. With a heart full of dreams and a mind brimming with ideas, he was ready to embrace whatever came next, knowing that the world was now a little smaller and more connected than before. ———

4. (10 points) What did you observe from Q1 - Q3? Did the different “top_p” configurations give you the same or different results? Why?

Experiment Results

Q1 (Default settings): Generated one coherent and detailed story.

Q2 (top_p=1, n_samples=2): Produced two diverse stories with different plots and expressions.

Q3 (top_p=0.5, n_samples=2): Generated two stories that were more similar in language and structure.

Did the different top_p configurations give the same or different results? Why?

Different Results: The stories varied depending on the top_p value used.

- With **top_p=1**: The model considered a broader range of word choices, leading to more diverse and creative stories.
- With **top_p=0.5**: The model's choices were limited to more probable words, resulting in less diverse and more predictable stories.

Why Did This Happen?

Effect of top_p: The top_p parameter controls the diversity of the generated text by limiting the pool of potential next words.

- **Higher top_p** (e.g., 1): Allows the model to use a wider range of words, increasing creativity and variability.
- **Lower top_p** (e.g., 0.5): Restricts the model to the most probable words, leading to more conservative and similar outputs.

Conclusion

Changing the top_p value affects the creativity and diversity of the generated content. Lower top_p values produce more uniform and predictable results, while higher values allow for more varied and creative outputs.

Task 2: gpt-4o-mini for Solving Mathematical Problems

5. (5 points) You are presented a mathematical problem: *Melanie is a door-to-door saleswoman. She sold a third of her vacuum cleaners at the green house, 2 more to the red house, and half of what was left at the orange house. If Melanie has 5 vacuum cleaners left, how many did she start with?* Can you use the ChatCompletion function and prompt gpt-4o-mini to work out the problem?

Response: Let x be the number of vacuum cleaners Melanie started with.

1. She sold a third of her vacuum cleaners at the green house:

$$\text{Sold at green house} = \frac{x}{3}$$

After this sale, she has:

$$x - \frac{x}{3} = \frac{2x}{3}$$

2. Next, she sold 2 more to the red house:

$$\text{Sold at red house} = 2$$

Now, she has:

$$\frac{2x}{3} - 2$$

3. Then, she sold half of what was left at the orange house. The amount left before this sale is:

$$\frac{2x}{3} - 2$$

Therefore, she sold half of this amount at the orange house:

$$\text{Sold at orange house} = \frac{1}{2} \left(\frac{2x}{3} - 2 \right)$$

After selling this, the amount of vacuum cleaners she has left is:

$$\left(\frac{2x}{3} - 2 \right) - \frac{1}{2} \left(\frac{2x}{3} - 2 \right)$$

We can factor it out:

$$\frac{2x}{3} - 2 - \frac{1}{2} \left(\frac{2x}{3} - 2 \right) = \frac{2x}{3} - 2 - \frac{1}{2} \cdot \frac{2x}{3} + 1$$

Now, simplifying the second term, we have:

$$\frac{2x}{3} - 2 - \frac{x}{3} + 1 = \left(\frac{2x}{3} - \frac{x}{3} \right) + (-2 + 1) = \frac{x}{3} - 1$$

4. We know from the problem that she has 5 vacuum cleaners left:

$$\frac{x}{3} - 1 = 5$$

Adding 1 to both sides, we get:

$$\frac{x}{3} = 6$$

Multiplying both sides by 3 gives:

$$x = 18$$

Thus, Melanie started with 18 vacuum cleaners.

Did gpt-4o-mini solve the problem correctly? If not, where did it go wrong?

Yes, gpt-4o-mini solved the problem correctly. The step-by-step solution accurately models the sales transactions and correctly calculates that Melanie started with 18 vacuum cleaners

6. (10 points) Now, try to get 10 solutions from gpt-4o-mini with "top_p" set to 0.7. You may see multiple

different answers produced by gpt-4o-mini. Summarize them in the table below and give one example output from gpt-4o-mini each. Did gpt-4o-mini do right in all of the solutions? If there are any mistakes, what are the common errors that gpt-4o-mini makes?

Answer	Count	Example Output
Ans1(18) Solutions that set up the equation $\frac{x}{3} - 1 = 5$	8 out of 10	<p>Response 0</p> <p>Let x be the number of vacuum cleaners Melanie started with.</p> <p>1. Sold at Green House:</p> $\text{Sold} = \frac{x}{3}$ $\text{Remaining} = x - \frac{x}{3} = \frac{2x}{3}$ <p>2. Sold at Red House:</p> $\text{Sold} = 2$ $\text{Remaining} = \frac{2x}{3} - 2$ <p>3. Sold at Orange House:</p> $\text{Sold} = \frac{1}{2} \left(\frac{2x}{3} - 2 \right)$ $\text{Remaining} = \left(\frac{2x}{3} - 2 \right) - \frac{1}{2} \left(\frac{2x}{3} - 2 \right) = \frac{x}{3} - 1$ <p>4. Set Up Equation:</p> $\frac{x}{3} - 1 = 5$ <p>5. Solve for x:</p> $\frac{x}{3} = 6 \implies x = 18$ <p>Answer: Melanie started with 18 vacuum cleaners. .</p>
Ans2(18) Solutions that set up the equation $\frac{2x-6}{6} = 5$	1 out of 10	<p>Response 4</p> <p>1. Sold at Green House:</p> $\text{Remaining} = x - \frac{x}{3} = \frac{2x}{3}$ <p>2. Sold at Red House:</p> $\text{Remaining} = \frac{2x}{3} - 2$ <p>3. Sold at Orange House:</p> $\text{Remaining} = \left(\frac{2x}{3} - 2 \right) - \frac{1}{2} \left(\frac{2x}{3} - 2 \right) = \frac{2x-6}{6}$ <p>4. Set Up Equation:</p> $\frac{2x-6}{6} = 5$ <p>5. Solve for x:</p> $2x - 6 = 30 \implies 2x = 36 \implies x = 18$ <p>Answer: Melanie started with 18 vacuum cleaners.</p>
Ans3(18) Solutions that set up the equation $\frac{1}{2} \left(\frac{2x}{3} - 2 \right) = 5$	1 out of 10	<p>respopnse 7- Let x be the number of vacuum cleaners. After simplifying, we use the equation $\frac{1}{2} \left(\frac{2x}{3} - 2 \right) = 5$, and solving this gives $x = 18$.</p>
<p>Observations and Findings: 1. Correctness of Solutions: All 10 responses correctly solved the problem and arrived at the answer of 18 vacuum cleaners. 2. Variation in Equation Setup: The majority (8 out of 10) used the equation:</p> $\frac{x}{3} - 1 = 5$ <p>while two responses used alternative equations. 3. Common Steps Across Solutions: All responses defined x as the initial number of vacuum cleaners and accounted for the sequential sales at each house. 4. Different Algebraic Approaches: Each method led to the correct answer through different equation setups, showcasing the model's flexibility.</p>		

7. (10 points) Can you try other ways to prompt gpt-4o-mini to give correct solutions more stably? Be creative! Include your prompt design and the answer in the report. Why do you think it works or not?

Prompt:
To help gpt-4o-mini provide correct solutions more consistently, we can design a prompt that includes multiple example problems with detailed solutions. This method, known as few-shot prompting, guides the model to follow the demonstrated reasoning steps.

Prompt Design:

- *System Message: Instruct the model to act as an expert mathematician and solve problems step by step.*
- *Example Problems and Solutions: Provide one or more solved math problems to illustrate the expected format and level of detail.*
- *New Problem: Present the new problem for the model to solve, expecting it to mimic the reasoning style shown in the examples.*

Math Problems and Solutions:

- **Problem 1:** Melanie is a door-to-door saleswoman. She sold a third of her vacuum cleaners at the green house, 2 more to the red house, and half of what was left at the orange house. If Melanie has 5 vacuum cleaners left, how many did she start with? **Solution 1:** *in py notebook*
- **Problem 2:** A farmer has chickens and cows. If there are a total of 30 heads and 100 legs, how many chickens and cows are there? **Solution 2:** *in py notebook*
- **Problem 3:** John drives for 3 hours at a speed of 60 mph and then turns around because he realizes he forgot something very important at home. He tries to get home in 4 hours but spends the first 2 hours in standstill traffic. He spends the next half-hour driving at a speed of 30 mph, before being able to drive the remaining time of the 4 hours going at 80 mph. How far is he from home at the end of those 4 hours?

Output:
Solution from gpt-4o-mini:
 Hello! How can I assist you today? To solve this problem, we will break it down into parts based on John's driving and traffic experiences.

- Distance John traveled before turning around:**

$$\text{Distance} = \text{Speed} \times \text{Time} = 60 \text{ mph} \times 3 \text{ hours} = 180 \text{ miles}$$
- John's return journey:**
 - Traffic time: 2 hours with 0 miles traveled.
 - Time spent at 30 mph for 0.5 hours:

$$\text{Distance}_{30} = 30 \text{ mph} \times 0.5 \text{ hours} = 15 \text{ miles}$$
- Driving the remaining time (1.5 hours at 80 mph):**

$$\text{Distance}_{80} = 80 \text{ mph} \times 1.5 \text{ hours} = 120 \text{ miles}$$
- Total distance traveled back:**

$$\text{Total distance back} = 15 \text{ miles} + 120 \text{ miles} = 135 \text{ miles}$$
- Final distance from home:**

$$\text{Distance from home} = 180 \text{ miles} - 135 \text{ miles} = 45 \text{ miles}$$

Findings: The system breaks down each part of the problem methodically, providing a step-by-step explanation. The use of few-shot prompting and example problems ensures consistent reasoning in solving the problem.

Part 2: Build an LLM Agent with Gentopia [50 pts + 20 bonus pts]

In this part, we will switch to an extended topic called “LLM agents”. In Part 1, we have mainly used gpt-4o-mini as a question-answer system, but an LLM can be formulated to be a “vivid” agent who learns to use tools and helps us in broader tasks, just like a virtual assistant!

To this end, we will be using an open-source LLM agent implementation framework, called “Gentopia” [2]. The original code repository can be found at <https://github.com/Gentopia-AI>, but note that for this assignment, we will use this adapted “Gentopia-Mason” version at <https://github.com/LittleYUYU/Gentopia-Mason/tree/main>.

Prerequisite: Library Installation

As the first step, please make sure to install all required Python packages following the instructions on README. More specifically, it means to run the following lines of code:

```
1 # Clone the repository
2 git clone git@github.com:LittleYUYU/Gentopia-Mason.git
3 cd Gentopia-Mason
4
5 # Create a conda virtual environment
6 conda create --name gentenv python=3.10
7 conda activate gentenv
8 pip install -r requirements.txt
9
10 # Set up global environment
11 export PYTHONPATH="$PWD/Gentopia:$PYTHONPATH"
12
13 # Set up OpenAI API key
14 cd GentPool
15 touch .env
16 echo "OPENAI_API_KEY=<your_openai_api_key>" >> .env
```

Task 3: Build a Vanilla LLM Agent

In this task, you will follow the instructions at <https://github.com/LittleYUYU/Gentopia-Mason/tree/main?tab=readme-ov-file#quick-start-clone-a-vanilla-llm-agent> to create a vanilla LLM agent. This agent is essentially the same question-answer model you have tried in Part 1.

The agent, under the framework of Gentopia, is configured as follows (with annotations):

```
1 # ./GentPool/gentpool/pool/<your_agent_name>/agent.yaml
2 # Vanilla agent template
3
4 name: <your_agent_name>
5 version: 0.0.1
6 type: vanilla
7 description: A plain gpt-4 LLM. # this is a brief description of the agent
8 target_tasks:
9   - anything to do with an LLM
10 llm:
11   model_name: gpt-4 # indicating the API version
12   params:
13     temperature: 0.0 # both temperature and top_p are used to control the
14       sampling diversity
```

```

14     top_p: 0.9
15     repetition_penalty: 1.0 # this is for preventing repetitive tokens
16     max_tokens: 1024 # maximum generation length
17 prompt_template: !prompt VanillaPrompt # this is prompt we used for this agent,
    see ./Gentopia/gentopia/prompt/vanilla.py

```

8. (10 points) **Change the "model_name" field from "gpt-4" to "gpt-4o-mini".** Successfully run this agent, and copy its responses for the following conversations below:

User: Hi! Tell me something about George Mason NLP group!

my_agent:

The group typically collaborates with other departments and institutions, contributing to interdisciplinary research that addresses real-world problems. They may also engage in educational activities, such as teaching courses related to NLP and hosting workshops or seminars to share their findings with the broader academic community. For the most current information about their research projects, publications, and activities, it's best to visit their official website or check recent publications from the group. *The George Mason University Natural Language Processing (NLP) group is part of the university's broader research initiatives in computer science and artificial intelligence. The group focuses on various aspects of NLP, including but not limited to machine learning, text analysis, information retrieval, and computational linguistics. Researchers in this group often work on projects that involve developing algorithms and models to understand and generate human language, as well as applications in areas like sentiment analysis, language translation, and dialogue systems.*

The group typically collaborates with other departments and institutions, contributing to interdisciplinary research that addresses real-world problems. They may also engage in educational activities, such as teaching courses related to NLP and hosting workshops or seminars to share their findings with the broader academic community. For the most current information about their research projects, publications, and activities, it's best to visit their official website or check recent publications from the group.

User: Who is/are leading the group?

my_agent: Could you please provide more context or specify which group you are referring to? This will help me give you a more accurate answer.

What problems did you see from the second response?

The main issue with the agent's second response is its failure to maintain conversational context. When asked, "Who is/are leading the group?", the agent should have recognized that I was still referring to the George Mason NLP group. Instead, it requested clarification, indicating a lack of context awareness.

Conclusion:

The agent needs better context retention to provide accurate and relevant answers in follow-up questions.

9. (10 points) Now, let's do a trick here. Instead of asking two questions, let's ask a combined one. And similarly let's also try the second question asking about facts:

User: Who is/are leading George Mason NLP group?

my_agent: *As of my last update in October 2023, the George Mason University Natural Language Processing (NLP) group is led by Dr. Mark E. Nissenbaum and Dr. David A. Smith. However, faculty and research group leadership can change, so I recommend checking the official George Mason University website or the specific NLP group's page for the most current information.*

User: Find papers written by Ziyu Yao at George Mason University

my_agent:

Google Scholar: Go to Google Scholar and enter "Ziyu Yao George Mason University" in the search bar. This should bring up a list of publications associated with that name and institution. University Repository: Check George Mason University's institutional repository or library website. Many universities have a digital archive of faculty publications. ResearchGate or Academia.edu: These platforms allow researchers to share their publications. You can search for Ziyu Yao on these sites. Academic Databases: Use academic databases like PubMed, IEEE Xplore, or JSTOR. You can search for the author's name and filter results by affiliation. Contact the Author: If you have trouble finding the papers, consider reaching out directly to Ziyu Yao via email or through their university profile. If you need help with a specific topic or paper, feel free to ask!*I don't have direct access to databases or the internet to search for specific papers or authors. However, you can find papers written by Ziyu Yao at George Mason University by following these steps:*

Google Scholar: Go to Google Scholar and enter "Ziyu Yao George Mason University" in the search bar. This should bring up a list of publications associated with that name and institution. University Repository: Check George Mason University's institutional repository or library website. Many universities have a digital archive of faculty publications. ResearchGate or Academia.edu: These platforms allow researchers to share their publications. You can search for Ziyu Yao on these sites. Academic Databases: Use academic databases like PubMed, IEEE Xplore, or JSTOR. You can search for the author's name and filter results by affiliation. Contact the Author: If you have trouble finding the papers, consider reaching out directly to Ziyu Yao via email or through their university profile. If you need help with a specific topic or paper, feel free to ask!

How do the responses look now? (Hint: you may want to verify the agent output.) Why would this situation happen? Describe your observations and findings below:

The agent's responses contain inaccuracies and fabrications. It incorrectly names "Dr. Mark E. Nissenbaum" and "Dr. David A. Smith" as leaders of the George Mason NLP group, but there is no evidence that these individuals hold such positions at the university. This situation occurs because the language model is generating plausible-sounding but incorrect information—a phenomenon known as "hallucination." The model tries to provide a helpful answer even when it lacks accurate data, leading to the creation of fictitious details. This highlights the model's limitation in delivering precise factual information without access to verified sources.

Till now, you probably have realized that an LLM alone will be facing various problems, including not being able to access the latest knowledge. Therefore, we'd like to try something called "tool-augmented LLM agent", where an LLM agent can be equipped with external tools for tasks. This is what we will

do in Task 4!

Task 4: Build a *Tool-Augmented* LLM Agent

Now, we will follow the instructions at <https://github.com/LittleYUYU/Gentopia-Mason/tree/main?tab=readme-ov-file#implement-a-scholar-llm-agent-with-tool-augmentation> and create a tool-augmented LLM agent. In particular, this agent, called a “scholar agent”, has been equipped with tools to access Google Scholar.²

This tool use does not come for free; it relies on Python implementation using the “scholarly” library,³ which can be found within the gentopia source code at https://github.com/LittleYUYU/Gentopia-Mason/blob/main/Gentopia/gentopia/tools/google_scholar.py.

Again, **change the “model_name” field to “gpt-4o-mini”**.

10. (10 points) Now, let’s the same query for Prof. Ziyu Yao’s papers again:

²<https://scholar.google.com/>

³<https://pypi.org/project/scholarly/>

User: Find papers written by Ziyu Yao at George Mason University

my_scholar_agent:

1. ****Navigating the Shortcut Maze: A Comprehensive Analysis of Shortcut Learning in Text Classification by Language Models**** - Publication Year: 2024 - Venue: arXiv preprint arXiv:2409.17455, 2024 - Citations: 0
2. ****A Practical Review of Mechanistic Interpretability for Transformer-Based Language Models**** - Publication Year: 2024 - Venue: arXiv preprint arXiv:2407.02646, 2024 - Citations: 1
3. ****An Investigation of Neuron Activation as a Unified Lens to Explain Chain-of-Thought Eliciting Arithmetic Reasoning of LLMs**** - Publication Year: 2024 - Venue: ACL 2024, 2024 - Citations: 1
4. ****Look Further Ahead: Testing the Limits of GPT-4 in Path Planning**** - Publication Year: 2024 - Venue: IEEE CASE 2024, 2024 - Citations: 0
5. ****MathVC: An LLM-Simulated Multi-Character Virtual Classroom for Mathematics Education**** - Publication Year: 2024 - Venue: arXiv preprint arXiv:2404.06711, 2024 - Citations: 8

If you need more information or additional papers, feel free to ask!*Here are the recent papers written by Ziyu Yao at George Mason University:*

1. ****Navigating the Shortcut Maze: A Comprehensive Analysis of Shortcut Learning in Text Classification by Language Models**** - Publication Year: 2024 - Venue: arXiv preprint arXiv:2409.17455, 2024 - Citations: 0
2. ****A Practical Review of Mechanistic Interpretability for Transformer-Based Language Models**** - Publication Year: 2024 - Venue: arXiv preprint arXiv:2407.02646, 2024 - Citations: 1
3. ****An Investigation of Neuron Activation as a Unified Lens to Explain Chain-of-Thought Eliciting Arithmetic Reasoning of LLMs**** - Publication Year: 2024 - Venue: ACL 2024, 2024 - Citations: 1
4. ****Look Further Ahead: Testing the Limits of GPT-4 in Path Planning**** - Publication Year: 2024 - Venue: IEEE CASE 2024, 2024 - Citations: 0
5. ****MathVC: An LLM-Simulated Multi-Character Virtual Classroom for Mathematics Education**** - Publication Year: 2024 - Venue: arXiv preprint arXiv:2404.06711, 2024 - Citations: 8

If you need more information or additional papers, feel free to ask!

Does the response look reasonable now? Based on the meta information that popped out through the agent's responding process, can you describe how the agent made it right this time?

Yes, the response looks reasonable now. The agent correctly listed recent papers authored by Ziyu Yao at George Mason University with accurate titles, publication years, venues, and citation counts. This improvement occurred because the agent utilized the Google Scholar tool to fetch real-time scholarly data. By integrating this external resource, the agent accessed up-to-date information rather than relying solely on its internal knowledge base, leading to a more accurate and helpful response.

11. (5 points) Look at the configuration file of your scholar agent at `./GentPool/gentpool/pool/<your_agent_name>/agent.yaml` and see what other functions are supported. Try a few other things you found interesting! Some examples to consider: *Can you summarize the paper titled "Gentopia: A collaborative platform for tool-augmented llms"?*, *What papers have cited "Gentopia: A collaborative platform for tool-augmented llms"*, etc.

User: *What are some papers related to 'Large Language Models in Education'?*

my_scholar_agent:

1. ****ChatGPT for good? On opportunities and challenges of large language models for education**** - Authors: E Kasneci, K Seßler, S Küchemann, M Bannert - Year: 2023 - Venue: Learning and Individual Differences - Abstract: Large language models represent a significant advancement in the field of AI. The underlying technology is key to further innovations and, despite critical views and even bans. - Citations: 2857

2. ****Adapting large language models for education: Foundational capabilities, potentials, and challenges**** - Authors: Q Li, L Fu, W Zhang, X Chen, J Yu, W Xia - Year: 2023 - Venue: arXiv preprint - Abstract: Online education platforms, leveraging the internet to distribute education resources, seek to provide convenient education but often fall short in real-time communication with students. - Citations: 14

3. ****Large language models for education: A survey and outlook**** - Authors: S Wang, T Xu, H Li, C Zhang, J Liang, J Tang - Year: 2024 - Venue: arXiv preprint - Abstract: The advent of Large Language Models (LLMs) has brought in a new era of possibilities in the realm of education. This survey paper summarizes the various technologies of LLMs in education. - Citations: 21

4. ****ChatGPT is not capable of serving as an author: Ethical concerns and challenges of large language models in education**** - Authors: NL Rane, SP Choudhary, A Tawde - Year: 2023 - Venue: Research Journal of Education - Abstract: This research delves into the dynamic role of ChatGPT and similar large language models within the realm of education. It sheds light on their set of limitations, ethical concerns, and challenges. - Citations: 36

5. ****Assigning AI: Seven approaches for students, with prompts**** - Authors: E Mollick, L Mollick - Year: 2023 - Venue: arXiv preprint - Abstract: This paper examines the transformative role of Large Language Models (LLMs) in education and their potential as learning tools, despite their inherent risks and limitations. - Citations: 78

These papers explore various aspects of large language models in the educational context, including their opportunities, challenges, ethical concerns, and practical applications. *Here are some papers related to "Large Language Models in Education":*

1. ****ChatGPT for good? On opportunities and challenges of large language models for education**** - Authors: E Kasneci, K Seßler, S Küchemann, M Bannert - Year: 2023 - Venue: Learning and Individual Differences - Abstract: Large language models represent a significant advancement in the field of AI. The underlying technology is key to further innovations and, despite critical views and even bans. - Citations: 2857

2. ****Adapting large language models for education: Foundational capabilities, potentials, and challenges**** - Authors: Q Li, L Fu, W Zhang, X Chen, J Yu, W Xia - Year: 2023 - Venue: arXiv preprint - Abstract: Online education platforms, leveraging the internet to distribute education resources, seek to provide convenient education but often fall short in real-time communication with students. - Citations: 14

3. ****Large language models for education: A survey and outlook**** - Authors: S Wang, T Xu, H Li, C Zhang, J Liang, J Tang - Year: 2024 - Venue: arXiv preprint - Abstract: The advent of Large Language Models (LLMs) has brought in a new era of possibilities in the realm of education. This survey paper summarizes the various technologies of LLMs in education. - Citations: 21

4. ****ChatGPT is not capable of serving as an author: Ethical concerns and challenges of large language models in education**** - Authors: NL Rane, SP Choudhary, A Tawde - Year: 2023 - Venue: Research Journal of Education - Abstract: This research delves into the dynamic role of ChatGPT and similar large language models within the realm of education. It sheds light on their set of limitations, ethical concerns, and challenges. - Citations: 36

5. ****Assigning AI: Seven approaches for students, with prompts**** - Authors: E Mollick, L Mollick - Year: 2023 - Venue: arXiv preprint - Abstract: This paper examines the transformative role of Large Language Models (LLMs) in education and their potential as learning tools, despite their inherent risks and limitations. - Citations: 78

These papers explore various aspects of large language models in the educational context, including their opportunities, challenges, ethical concerns, and practical applications.

12. (15 points) You may have realized that the current scholar agent still fails to answer general questions such as “Tell me something about George Mason NLP group”, though they are also relevant to “scholar”. In addition, it cannot read PDFs for summarizing papers either, so it cannot provide answers to specific questions about the paper details. Can you figure out a way for the agent to do Google Search (5 pts) and PDF reading (10 pts), and present a few examples below? (Hint: Google Search has been provided in Gentopia’s source code, whereas you could opt for any Python tools you like, e.g., PyPDF2,⁴ for PDF reading.)

⁴<https://www.geeksforgeeks.org/working-with-pdf-files-in-python/>

User: Summarize the key points of the paper available at <https://arxiv.org/pdf/2106.01548.pdf>

my_scholar_agent:

1. **Objective**: The study aims to demonstrate that ViTs and MLP-Mixers can outperform ResNets in terms of accuracy and robustness without relying on large-scale pre-training or extensive data augmentations. 2. **Loss Geometry Analysis**: The authors analyze the loss landscapes of ViTs and MLP-Mixers, revealing that these models converge to sharp local minima, which can negatively impact generalization. In contrast, ResNets tend to converge to flatter regions. 3. **Sharpness-Aware Minimizer (SAM)**: To address the sharp minima issue, the paper introduces the sharpness-aware minimizer (SAM), which promotes smoother loss landscapes during training. This optimizer helps improve the generalization performance of ViTs and MLP-Mixers significantly. 4. **Performance Improvements**: The application of SAM leads to substantial improvements in accuracy on various tasks, including supervised learning, adversarial robustness, and transfer learning. For instance, ViT-B/16's top-1 accuracy on ImageNet increases from 74.65. 5. **Sparsity of Active Neurons**: The study finds that SAM increases the sparsity of active neurons in the initial layers of the networks, contributing to the smoother loss landscapes and better generalization. 6. **Comparison with ResNets**: The results indicate that ViTs can outperform ResNets of similar size and throughput when trained from scratch on ImageNet, achieving better accuracy and robustness without the need for extensive pre-training or strong data augmentations. 7. **Robustness Evaluation**: The paper evaluates the robustness of the models against various corruptions and adversarial attacks, showing that SAM-enhanced models exhibit improved robustness compared to their vanilla counterparts. 8. **Conclusion**: The findings suggest that with the right optimization techniques like SAM, ViTs and MLP-Mixers can be competitive with or even surpass traditional convolutional networks like ResNets, particularly in scenarios where data and computational resources are limited.

Overall, the paper highlights the potential of ViTs and MLP-Mixers in computer vision tasks and emphasizes the importance of optimization strategies in enhancing model performance. *The paper titled "When Vision Transformers Outperform ResNets Without Pre-Training or Strong Data Augmentations" investigates the performance of Vision Transformers (ViTs) and MLP-Mixers in comparison to ResNets, focusing on their training efficiency and generalization capabilities. Here are the key points:*

1. **Objective**: The study aims to demonstrate that ViTs and MLP-Mixers can outperform ResNets in terms of accuracy and robustness without relying on large-scale pre-training or extensive data augmentations. 2. **Loss Geometry Analysis**: The authors analyze the loss landscapes of ViTs and MLP-Mixers, revealing that these models converge to sharp local minima, which can negatively impact generalization. In contrast, ResNets tend to converge to flatter regions. 3. **Sharpness-Aware Minimizer (SAM)**: To address the sharp minima issue, the paper introduces the sharpness-aware minimizer (SAM), which promotes smoother loss landscapes during training. This optimizer helps improve the generalization performance of ViTs and MLP-Mixers significantly. 4. **Performance Improvements**: The application of SAM leads to substantial improvements in accuracy on various tasks, including supervised learning, adversarial robustness, and transfer learning. For instance, ViT-B/16's top-1 accuracy on ImageNet increases from 74.65. 5. **Sparsity of Active Neurons**: The study finds that SAM increases the sparsity of active neurons in the initial layers of the networks, contributing to the smoother loss landscapes and better generalization. 6. **Comparison with ResNets**: The results indicate that ViTs can outperform ResNets of similar size and throughput when trained from scratch on ImageNet, achieving better accuracy and robustness without the need for extensive pre-training or strong data augmentations. 7. **Robustness Evaluation**: The paper evaluates the robustness of the models against various corruptions and adversarial attacks, showing that SAM-enhanced models exhibit improved robustness compared to their vanilla counterparts. 8. **Conclusion**: The findings suggest that with the right optimization techniques like SAM, ViTs and MLP-Mixers can be competitive with or even surpass traditional convolutional networks like ResNets, particularly in scenarios where data and computational resources are limited.

Overall, the paper highlights the potential of ViTs and MLP-Mixers in computer vision tasks and emphasizes the importance of optimization strategies in enhancing model performance. **User:** Recent research papers on large language models in education

my_scholar_agent:

1. **Large Language Models in Education: Vision and Opportunities** by W Gan (2023) - This article investigates and summarizes the application of LLMs in smart education, introducing the research background and motivation of LLMs.

2. **Large Language Models for Education: A Survey** (2024) - This paper conducts a systematic review of LLMs in education, focusing on current technologies, challenges, and future developments.

3. **ChatGPT for good? On opportunities and challenges of educational applications of large language models** by E Kasneci (2023) - This commentary presents the potential benefits and challenges of educational applications of LLMs from student and teacher perspectives.

4. **The impact of large language models on higher education** by IC Peláez-Sánchez (2024) - This paper discusses the advancements in general artificial intelligence (GAI) represented by LLMs and their effectiveness based on extensive textual training.

5. **Exploring large language models as an integrated tool for research** by SN Jyothy (2024) - This research explores how LLMs can assist researchers in improving their research skills and identifying challenging topics.

6. **Practical and ethical challenges of large language models in education** by L Yan (2024) - This systematic scoping review identifies the current state of research on using LLMs to automate educational processes.

7. **Large language models and medical education: a paradigm shift** by Z Li (2024) - This paper discusses how LLMs can refine teaching strategies based on students' learning patterns and challenges.

These papers provide insights into the applications, challenges, and transformative potential of large language models in the field of education.

Here are some recent research papers on large language models (LLMs) in education:

1. **Large Language Models in Education: Vision and Opportunities** by W Gan (2023) - This article investigates and summarizes the application of LLMs in smart education, introducing the research background and motivation of LLMs.

2. **Large Language Models for Education: A Survey** (2024) - This paper conducts a systematic review of LLMs in education, focusing on current technologies, challenges, and future developments.

3. **ChatGPT for good? On opportunities and challenges of educational applications of large language models** by E Kasneci (2023) - This commentary presents the potential benefits and challenges of educational applications of LLMs from student and teacher perspectives.

4. **The impact of large language models on higher education** by IC Peláez-Sánchez (2024) - This paper discusses the advancements in general artificial intelligence (GAI) represented by LLMs and their effectiveness based on extensive textual training.

5. **Exploring large language models as an integrated tool for research** by SN Jyothy (2024) - This research explores how LLMs can assist researchers in improving their research skills and identifying challenging topics.

6. **Practical and ethical challenges of large language models in education** by L Yan (2024) - This systematic scoping review identifies the current state of research on using LLMs to automate educational processes.

7. **Large language models and medical education: a paradigm shift** by Z Li (2024) - This paper discusses how LLMs can refine teaching strategies based on students' learning patterns and challenges.

These papers provide insights into the applications, challenges, and transformative potential of large language models in the field of education.

Bonus: Build Your Own Agent! (20 pts)

We will provide up to 20 points to students who implement a different tool-augmented LLM agent *for positive use (i.e., the agent cannot be designed for unethical purposes)*. To claim bonus points, the agent needs to be substantially different from the scholar agent and others included in Gentopia's agent pool. Be creative! Partial credits may be provided based on student effort.

Acknowledgment

The mathematical problems used in this assignment come from the GSM8k dataset [1], and the source code of Gentopia comes from the referred EMNLP'23 paper [2].

References

- [1] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.
- [2] Binfeng Xu, Xukun Liu, Hua Shen, Zeyu Han, Yuhan Li, Murong Yue, Zhiyuan Peng, Yuchen Liu, Ziyu Yao, and Dongkuan Xu. Gentopia.AI: A collaborative platform for tool-augmented llms. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 237–245, 2023.