BASIC PROMPT ENGINEERING WITH CHATGPT: AN INTRODUCTION

MODULE 6: Broader societal implications of LLMs and ChatGPT

OVERVIEW

In this module, we will explore the broader societal implications of Large Language Models and their associated interfaces like ChatGPT. As you learn to work with these technologies, it's crucial to consider not only their technical capabilities but also their potential impacts on society, both positive and negative. In this module, you will explore key topics related to the ethical, social, and economic aspects of LLMs and ChatGPT (and by extension, other AI chatbots), while developing an understanding of concepts such as societally responsible innovation, responsible AI, principled innovation, ethical AI and innovation, and principles of ethical and responsible use.

You will start by exploring the benefits and risks associated with LLMs, such as enhanced communication and productivity, the potential spread of misinformation, and societal disruption. This will provide you with an important perspective on the advantages and drawbacks of using these models in various contexts.

Next, you will examine the challenges surrounding LLMs, such as the potential impacts on employment, the ethics of LLM development and use, and responsible and principled innovation. By considering these wide-ranging consequences, you will be better prepared to make responsible decisions and contribute positively to the evolving AI landscape and how ChatGPT is used across multiple domains.

By the end of Module 6, you will have a foundational understanding of the broader societal implications of large language models and AI chatbots and their role in shaping our world. You will be equipped with the knowledge to engage in thoughtful discussions about AI ethics and the responsible use of LLMs, fostering a more inclusive, transparent, and accountable approach to AI development. This understanding will enable you to make more informed decisions when working with language models, and to contribute constructively to the ongoing conversation surrounding AI's impact on society.

The module should take between 3 – 7 hours to complete.

LEARNING OBJECTIVES

By the end of this module you will be able to:

6.1. Describe key potential benefits and risks/ethical concerns associated with LLMs and AI chatbots.
6.2. Explain the concepts of responsible innovation and how it applies to LLMs and AI chatbots like ChatGPT.
6.3. Apply ethical and responsible practices to the use of LLMs and AI chatbots, based on a set of personal Principles for Responsible Prompt Engineering.

FLOW

OVERVIEW

Intro blurb (above)

Intro video:

- This module is focused on the societally responsible and ethical use of AI chatbots like ChatGPT, together with the underlying large language models
- This is extremely important as, with a technology as powerful as this, mist-steps, even well intentioned ones, can cause substantial harm.
- This is not a module about existential risks of AI. Rather, it is designed to help you understand the landscape around the responsible and ethical use of ChatGPT as you become an expert in its use.
- The module also only touches the tip of the iceberg here. For instance, it doesn't directly address the use of ChatGPT in plagiarism or in intellectual property. But it does provide you with the understanding to make responsible and ethical decisions in new and often complex situations.
- And importantly, it helps you become one of a growing number of people who are helping ensure the emergence of socially responsible and beneficial uses of AI.

ADDITIONAL INFORMATION

As you explore the world of Large Language Models and their interfaces like ChatGPT, it's crucial to consider their impact on society, both positive and negative. This understanding will not only help you become a responsible and principled prompt engineer, but will help you be part of ensuring that emerging AI capabilities are used for the betterment of all.

Some key topics here include:

1. **Benefits of LLMs:** Large language models (and interfaces like ChatGPT) can bring numerous advantages, including:
    - Boosting productivity and efficiency across various professional domains.
    - Democratizing access to information and knowledge.

- Supporting language translation and fostering cross-cultural understanding.
- Enabling creative applications in writing, art, and design.
- Accelerating the discovery of new knowledge, and novel solutions to complex problems.
- Creating highly personalized and tailored environments and experiences that accelerate learning.
- Substantially extending the ability of individuals and organizations to innovate, invent, design, learn, manufacture, and otherwise do what humans do.
- (There are many other benefits of LLMs – these are just some).

2. **Risks and challenges:** These include:
    - Misinformation and manipulation: LLMs can generate plausible yet false information, which could be exploited for spreading misinformation or manipulating opinions.
    - Bias and discrimination: LLMs may unintentionally learn and perpetuate harmful biases present in their training data, leading to discriminatory or offensive outputs. As someone who is involved in prompt engineering, you should strive to recognize and mitigate such biases in your work.
    - Privacy concerns: LLMs can inadvertently disclose sensitive or private information. Be mindful of privacy implications when designing prompts and applications.
    - Job displacement: LLMs have the ability to disrupt jobs across multiple sectors and, through this, impact economic stability and growth.
    - Loss of autonomy and dignity: Used irresponsibly, LLMs have the potential to undermine the ability of individuals to have control over their lives, and even to undermine what gives individuals a sense of self and purpose.
    - (These are just some of the risks and challenges).
3. **Ethical considerations:** These include:
    - Fairness and equity: Ensuring that work with LLMs promotes fairness and equity by addressing potential biases and avoiding the exacerbation of existing inequalities.
    - Transparency and accountability: Fostering a culture of responsibility and openness by disclosing the development, use, and limitations of LLMs in your projects.
    - Loss of agency: Robbing people of the ability to be in control of their situation and their lives.
    - Environmental impact: Being conscious of the energy consumption and carbon footprint associated with training and deploying large-scale AI models.
    - (These are just some of the ethical considerations).
4. **The roles of responsible prompt engineering:** These include:

- Crafting clear, precise, and unbiased prompts, to help direct LLMs towards generating accurate, relevant, and safe outputs.
- Focusing on creating prompt templates that ensure that LLMs are adaptable and perform well across diverse contexts, reducing the likelihood of unintended consequences.
- Evaluating the performance of LLMs critically, considering metrics that reflect ethical and societal aspects in addition to technical performance.

5. **The impact of LLMs on learning and education:** These include:
    - Understanding how LLMs may undermine learning by providing inaccurate, misleading, or otherwise false information, and learning how to address these limitations.
    - Being aware of how LLMs might tempt students to take shortcuts in assignments that are classed as plagiarism or other forms of cheating.
    - Understanding how LLMs can transform education and learning if used in innovative and responsible ways (this course is one example).
    - Learning to develop prompts and prompt sequences that leverage the ability of LLMs to provide personalized and tailored learning environments.
6. **Governance and regulation:** These include:
    - Staying informed about ethical guidelines and best practices for LLM development and deployment.
    - Engaging in dialogue with peers, academics, and policymakers to contribute to a regulatory framework that balances innovation with societal well-being.
    - Advocating for public engagement and diverse perspectives to ensure that LLM development is guided by a broad range of values.
    - Contributing to the effective governance and regulation of LLMs, including their development and use.

By exploring these topics and embracing your role as a responsible prompt engineer, you can contribute to harnessing the power of LLMs for the greater good. Finally, remember to stay informed about the latest developments, engage in discussions about AI ethics, and always strive for responsible and principled AI practices in your work and community.

EXERCISE: Exploring the potential benefits of LLMs (15 points)

This exercise is designed to help you develop a broader understanding of the potential benefits of AI chatbots like ChatGPT and the underlying large language models.

It is also designed to demonstrate the use of ChatGPT in synthesizing rough notes.

- Spend some time exploring the potential benefits of LLMs, ChatGPT, and other AI chatbots. Use ChatGPT, Google, and any other sources at your disposal. Take notes.

- Ask ChatGPT to synthesize your notes into a coherent statement on the potential benefits of LLMs and ChatGPT. This should be between 300 – 500 words long. (Note that this is an exercise in asking ChatGPT to make sense of unstructured notes – you do no need to refine and format your notes first!)
- Submit documentation of your ChatGPT sessions below.

This exercise should take you no longer than an hour.

You will be given full points for submitting documentation that demonstrates you have spent an appropriate amount of time and effort on this exercise. Points will be given on submission. Points may be removed at a future date if it appears that you did not spend as much time and effort as expected on the exercise.

EXERCISE: Exploring the potential risks of LLMs (15 points)

This exercise is designed to help you develop a broader understanding of the potential risks, downsides, and ethical concerns associated with AI chatbots like ChatGPT and the underlying large language models.

It is also designed to demonstrate the use of ChatGPT in synthesizing rough notes.

- Spend some time exploring the potential risks and ethical concerns associated with LLMs, ChatGPT, and other AI chatbots. Use ChatGPT, Google, and any other sources at your disposal. Take notes.
- Ask ChatGPT to synthesize your notes into a coherent statement on the potential benefits of LLMs and ChatGPT. This should be between 300 – 500 words long. (Note that this is an exercise in asking ChatGPT to make sense of unstructured notes – you do no need to refine and format your notes first!)
- Submit documentation of your ChatGPT sessions below.

This exercise should take you no longer than an hour.

You will be given full points for submitting documentation that demonstrates you have spent an appropriate amount of time and effort on this exercise. Points will be given on submission. Points may be removed at a future date if it appears that you did not spend as much time and effort as expected on the exercise.

EXERCISE: Responsible Innovation and ChatGPT (20 points)

This exercise is built around a conversation with ChatGPT that helps develop your understanding of responsible innovation in the context of responsible AI and ChatGPT. It is also an example of how to use prompt engineering to develop a responsive and tailored learning environment.

Note that if you are interested in further developing your understanding here, you can run the exercise multiple times!

- In a new session, provide ChatGPT (using GPT-4) with the following prompt:
    - "Hi ChatGPT. My name is [include full name] and I would like you to act as my personal tutor and teach me about responsible innovation in the context of using ChatGPT. I would like you to cover the field broadly. Please start by asking me a question that helps you gauge my level of understanding. Based on my response, ask me a follow-up question that is designed to increase my understanding. Continue to do this until I show a broad understanding of responsible innovation in the context of using ChatGPT."
- Submit documentation of your ChatGPT sessions below.

This exercise should take you no longer than an hour.

You will be given full points for submitting documentation that demonstrates you have spent an appropriate amount of time and effort on this exercise. Points will be given on submission. Points may be removed at a future date if it appears that you did not spend as much time and effort as expected on the exercise.

EXERCISE: Ethical dilemma (20 points)

In this exercise you will explore the nature of an ethical dilemma with ChatGPT, together with its resolution.

This is a sophisticated exercise where you will need to work out how best to craft the prompt you use with ChatGPT, how you develop the conversation that follows on from this, how you will assess the validity and usefulness of responses, and how you in turn will respond to this.

You are expected to use the skills you have developed so far, as well as your imagination and initiative.

- Open a new session of ChatGPT using GPT-4).

- Ask ChatGPT to provide you with examples of three complex ethical dilemmas that involve the use of ChatGPT.
- Select one of them and have a conversation with ChatGPT about the nature of the dilemma and how it might be addressed. Your conversation should include at least 5 exchanges with ChatGPT – ideally more. You should critically assess ChatGPT's responses and provide feedback, guidance, and your perspective as appropriate.
- Submit documentation of your ChatGPT sessions below.

This exercise should take you no longer than an hour.

You will be given full points for submitting documentation that demonstrates you have spent an appropriate amount of time and effort on this exercise. Points will be given on submission. Points may be removed at a future date if it appears that you did not spend as much time and effort as expected on the exercise.

ASSIGNMENT: Principles of Responsible Prompt Engineering (25 points)

This final exercise asks you to develop your own personal principles of Responsible Prompt Engineering with ChatGPT.

You can use ChatGPT to help this, as well as other resources, but the final set of principles should be your own work, and should reflect your interests and your potential future career path.

Your personal Principles of Personal Prompt Engineering should have between 8 – 15 principles. Each should be clearly identified and have a brief (1 – 2 sentence) description.

You should upload your final principles below, making sure that they include a title and your name.

This exercise should take you no longer than 1 – 2 hours.

You will be given full points for submitting documentation that demonstrates you have spent an appropriate amount of time and effort on this exercise. Points will be given on submission. Points may be removed at a future date if it appears that you did not spend as much time and effort as expected on the exercise.

ASSIGNMENT: Self-reflection and feedback (5 points)

As prompt engineering, ChatGPT, and this course, are so new, we would like to finish by asking for your feedback. This is separate to the ASU course evaluation, and is designed to help us understand what worked well with the course and what we could do better.

This should take no more than 30 minutes at most of your time.

Thank you – and thank you so much for being a part of this course. I hope that you have not only learned a lot, but that you will find this useful in your personal and professional life moving forward.

Cheers

Andrew Maynard (Instructor)

PS – Please also make sure you complete the separate course evaluation administered by ASU!

Questions:

Please indicate how much you feel your understanding of large language models and their limitations has improved over the class, from 1 (low) to 10 (high)

Please indicate how much your understanding of prompt formulation and refinement has increased over the class, from 1 (low) to 10 (high)

Please indicate how much your ability to develop and use prompt has improved over the class, from 1 (low) to 10 (high)

Please indicate how much your ability to evaluate prompts and responses has grown over the class, from 1 (low) to 10 (high)

Please indicate to what degree your awareness of emerging trends around LLMs and their applications has increased over the class, from 1 (low) to 10 (high)

Please indicate to what degree your understanding of responsible innovation and the broader societal implications of LLMs has increased over the class, from 1 (low) to 10 (high)

What has been most useful to you in this course (open ended)?

What's one thing you would change about the course (open ended)?

Thank you!