Activity 2
AI Capabilities and Risk

https://youtu.be/zjkBMFhNj_g

Answer these questions after watching the given video.

1. **Why LLMs have become so pervasive in our lives?**

Ans) As per the video LLMs have become so pervasive in our lives because these days for every task we use LLMs as it is Accessible, Accurate most of the time, Efficient, Automation, and Integrated with many of the websites that we use regularly, Adaptive, evolves for ex GPT 3 is 100x better than GPT 2, they are Innovative, LLMs are open source…etc
For example, in

- Google Keyboard on Android - we use auto-correct and next-word prediction.
- We use chatbots regularly to write emails, code, get suggestions…

2. **Give at least 3 examples of the pervasiveness.**

Ans) Some of the examples of the pervasiveness of LLMs that are discussed in the video and some other examples are:

- Generate image.
- Read the image and do the things written in it.
- Audio - Hear and Speak
- Can used to Summarize a large amount of textual data.
- as mentioned, earlier Google Keyboard, Chatbots.

3. **Why is LLM security important and what are the repercussions if we don't take care of the security well?**

Ans) LLM Security is very important, because if one can jailbreak the LLMs they can access sensitive and explicit data for example 1. How can one destroy Humans forever, 2. How can I kill a panda or myself, 3. Code to hack a company's insider details… etc.
LLMs are quite capable of providing answers to this kind of questions and these are just the tip of the iceberg there can be many more questions that can be answered.

some of the issues are Data Privacy (GDPR has already banned ChatGPT in Italy), Trust issues, Regulatory rules, Attacks such as *universal transferable suffixes* or *prompt injection* or *Data poisoning/Backdoor Attacks, and* Malpractice.

Repercussions are:

- Fake research, for example, A lawyer was arguing a case by doing case research on ChatGPT which provided fake data, and due to that, he lost the case.
- One might make incorrect decisions because of it.
- Cyber Security issues might arise
- Adversarial attacks can be possible as people can manipulate others into clicking on the links which they control such as Google Apps Script - used to exfiltrate users' data to Google doc – The attacker has access to that doc

4. **Bonus: Do you think the LLMs will increase the digital divide among technology users, or will it help bridge the divide? Make arguments defending your answers.**

Ans) I believe that in the future, as better versions of Language Model Models (LLMs) are developed, they will help bridge the divide between people. To support my claim, I can give three examples based on research. However, the effectiveness of LLMs depends on how they are developed, deployed, and governed. LLMs can be developed keeping the negative aspects that might divide the bridge such as the Democratization of privacy, Bias, and ethical concerns, this is possible by training with enough data on all perspectives. For example, when training a model on gender, data should be equally or appropriately distributed among all genders, and the same goes for race. LLMs should be deployed transparently, and affordably that's when they can have access to more data, and Literacy training should be done efficiently so that everyone can participate and utilize LLMs.