UỶ BAN NHÂN DÂN TP. HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**



A blue circle with white text

Description automatically generated

**EXERCISE 1: TRẢ LỜI CÂU HỎI TRONG CHAPTER 1, 2, 3**

MÔN HỌC: **TRÍ TUỆ NHÂN TẠO NÂNG CAO**

**Giảng viên hướng dẫn** : Đỗ Như Tài

**Sinh viên thực hiện Nhóm 11**

Phạm Văn Nam 3122410251

Nguyễn Quan Tuấn Nghĩa 3122410260

Tạ Hồng Quí 3122410348

Vũ Quốc Vương 3120410629

*Thành phố Hồ Chí Minh - Tháng 09/2025*

Mục lục

[Discussion 1: Turing Test: Large Language Models (LLMs) 3](#_Toc208911783)

[Discussion 2: The AI Effect: AI gets no respect? 3](#_Toc208911784)

[Discussion 3: AI Safety 3](#_Toc208911785)

### Discussion 1: Turing Test: Large Language Models (LLMs)

* Would a modern LLM pass the Turing Test?
* Would you be fooled?
* Why does it or does it not pass your test?
* What does this mean for artificial general intelligence (AGI) or narrow AI?
* How do we currently test the performance of LLMs?
* See: Open LLM Leaderboard (Hugging Face)

### Discussion 2: The AI Effect: AI gets no respect?

* How do you think LLMs will affect the value of being able to write assays as taught in high school?
* LLMS write computer code. What does this mean for the value of learning to code?
* When should students be allowed to use the following tools? Give reasons for your decision.
  + A pocket calculator
  + LLMs (to answer homework questions and write assays)
  + LLMs to write or support writing code

### Discussion 3: AI Safety

* How are LLMs affected by:
* Robustness: Black swan vs. adversarial robustness
* Monitoring Al
* What about liability?
* Goal/reward alignment
* Reward hacking
* AGI and instrumental convergence
* Should the use of LLMs be regulated?
* How?
* What about copyright?