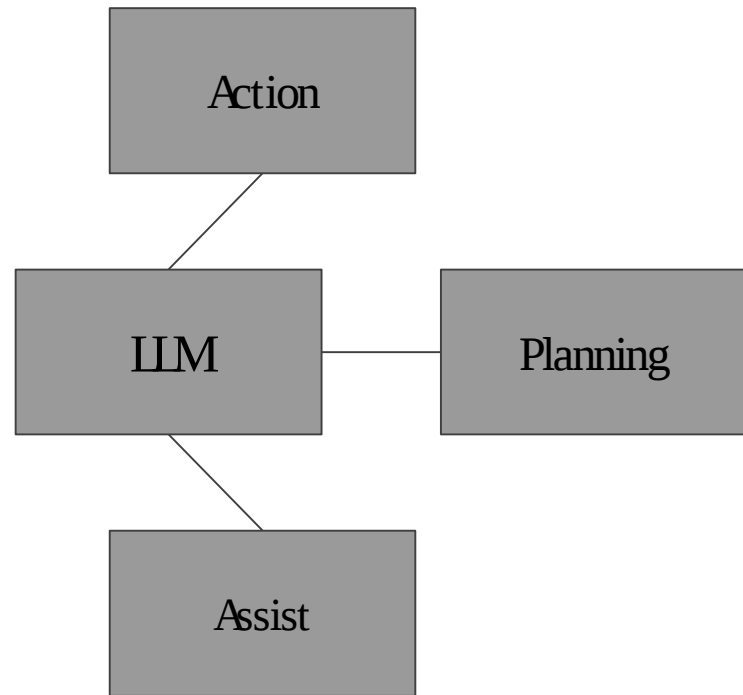


Застосування великих мовних моделей до розробки стратегій в настільних іграх

Виконав роботу: Найда Андрій Олегович

Керівник магістерської роботи: д.ф.м.н., доц. Ігнатенко О. П.

ІІМ, Розвиток та застосування в іграх



Існуючі тести продуктивності великих мовних моделей

T ▲	Model ▲	Average 📈 ▲	ARC ▲	HellaSwag ▲	MMLU ▲
◆	davidkim205/Rhea-72b-v0.5 📄	81.22	79.78	91.15	77.95
🗨	MTSAIR/MultiVerse_70B 📄	81	78.67	89.77	78.22
🗨	MTSAIR/MultiVerse_70B 📄	80.98	78.58	89.74	78.27
◆	abacusai/Smaug-72B-v0.1 📄	80.48	76.02	89.27	77.15
◆	ibivibiv/alpaca-dragon-72b-v1 📄	79.3	73.89	88.16	77.4
🗨	mistralai/Mixtral-8x22B-Instruct-v0.1 📄	79.15	72.7	89.08	77.77
🗨	MaziyarPanahi/Llama-3-70B-Instruct-DPO-v0.2 📄	78.96	72.53	86.22	80.41

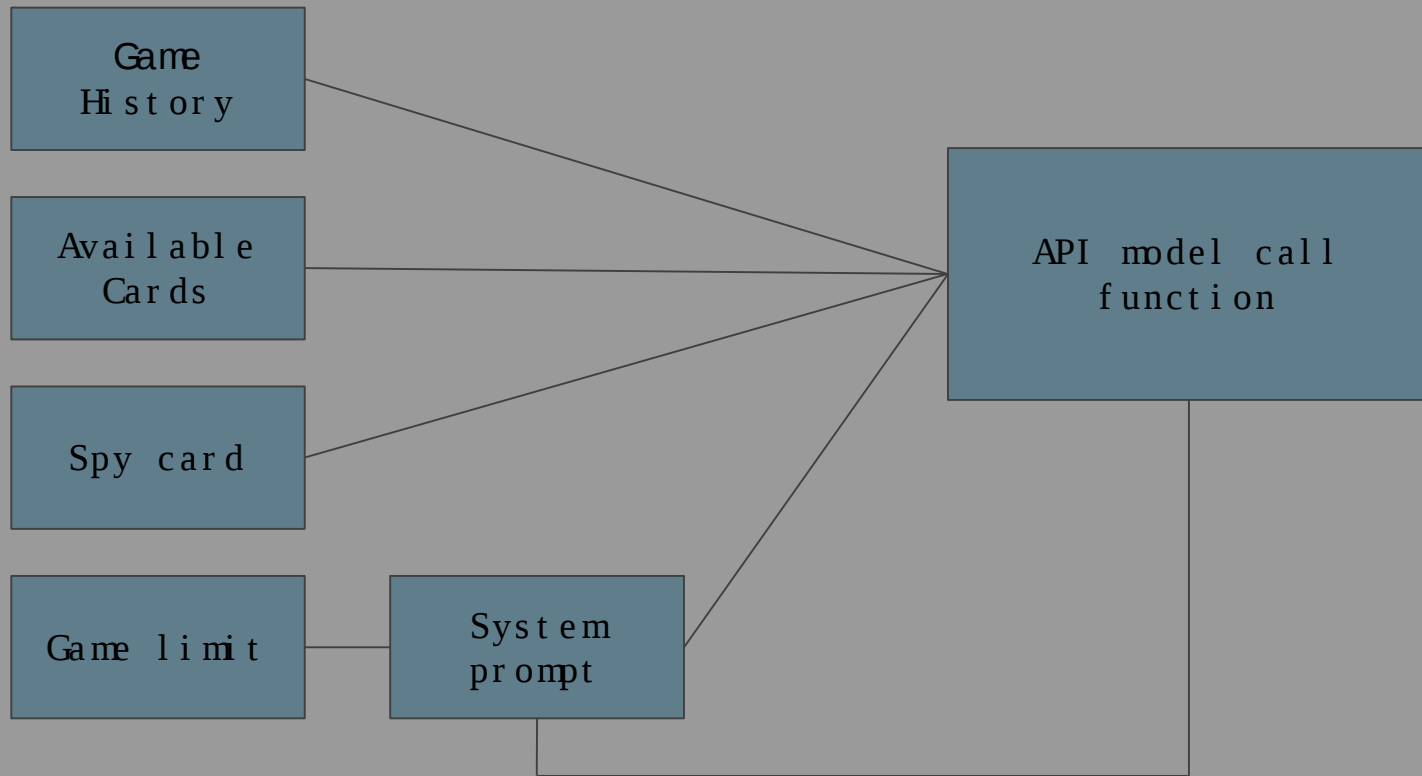
Відважні щури. Правила, складність гри



Генерування високоефективних стратегій для еталону гри.

Стратегія	Перемоги	Програші	Нічії	Успішність
(0, 7, 1, 4, 6, 5, 3)	355	61	995	0.9567
(3, 0, 7, 1, 4, 5, 6)	717	87	1293	0.9585
(3, 0, 7, 1, 6, 5, 4)	538	60	1084	0.9643
(3, 7, 0, 1, 4, 5, 6)	518	79	1363	0.9596
(3, 7, 0, 1, 6, 5, 4)	419	55	1186	0.9668
(7, 0, 1, 4, 6, 5, 3)	402	56	1184	0.9658
(7, 0, 3, 1, 5, 4, 6)	723	93	1341	0.9568
(7, 3, 0, 1, 6, 5, 4)	407	61	1082	0.960
(7, 3, 0, 1, 4, 5, 6)	555	77	1148	0.9567

Формулювання запиту до моделі



Таблиці результатів

Модель	Виграші	Програші	Нічії	Частка не програшу
gpt-3.5-turbo	48	11	41	0.89
gpt-4-turbo	3	29	68	0.71
gemini-1.5	13	15	72	0.85

Таблиці результатів

Модель	Виграші	Програші	Нічії	Частка не програшу
gpt-3.5-turbo	0	94	6	0.06
gemini-1.5	13	14	73	0.86

Таблиці результатів

Модель	Виграші	Програші	Нічії	Частка не програшу
gpt-3.5-turbo	0	98	2	0.02
gemini-1.5	1	99	0	0.01

Висновки

Поставлену задачу виконано - розроблено систему тестів продуктивності для великих мовних моделей з використанням настільних ігор.

Практична цінність - можливість оцінювання продуктивності моделей в контексті розробки стратегій в іграх для подальшого вибору моделі в контексті інтеграції в ігри.