THE PLACE OF USEFUL LEARNING

University of Strathclyde
Science

CS958 : Project

Explanation and Game-Play Capabilities of Modern AI and Large Language Models

- **ZERKSIS MISTRY**

Advanced Computer Science with Artificial Intelligence

Research Objectives

Aim:

To experiment and evaluate the game-playing and explainability competencies of the leading Large Language Models (LLMs).

Objectives:

- Assess Explanatory Performance
- Identify Factors Impacting Explainability
- Understanding the Quality of the Models
- Assessing Game-Play Logic of LLMs

Research Motivation

Acknowledging the ascent and central importance of LLMs in the AI domain.

Detailing the swift progress in NLP that has catalysed the development of sophisticated LLMs.

Discussing the need to understand LLM operations and their real-world applicability.

Exploring LLM capabilities beyond traditional NLP tasks.

Understanding the role of LLMs in the pursuit of Artificial General Intelligence (AGI).

The Turing Test's legacy and the evolution of machine language understanding.

Research Questions

- UNDERSTANDING OF THE INFERENCE QUALITY WITH FEW-SHOT PERSONALIZATION
- EVALUATION AND COMPARISON OF LLMS
- UNDERSTANDING MODEL COMPONENTS AND FEATURES
- EVALUATING LLMS IN THE AGE OF AGI
- INTERPRETABILITY AND EXPLAINABILITY IN LLMS

Literature Review Summary

HISTORICAL CONTEXT:

TURING TEST AND THE ASPIRATION FOR MACHINE LANGUAGE MASTERY.

IDENTIFIED RESEARCH GAP:

LLM PERFORMANCE IN STRATEGIC GAMES AND THEIR EXPLAINABILITY.

RECENT LLM INTRODUCTIONS:

DETAILED EXPLORATION OF LLM CAPABILITIES, HIGHLIGHTING THEIR TRANSFORMATIVE POTENTIAL IN VARIOUS APPLICATIONS.

EXPLAINABILITY IN AI:

WHILE THE TOPIC IS GAINING TRACTION, COMPREHENSIVE METHODOLOGIES AND FRAMEWORKS FOR ASSESSING LLM EXPLAINABILITY REMAIN SPARSE.

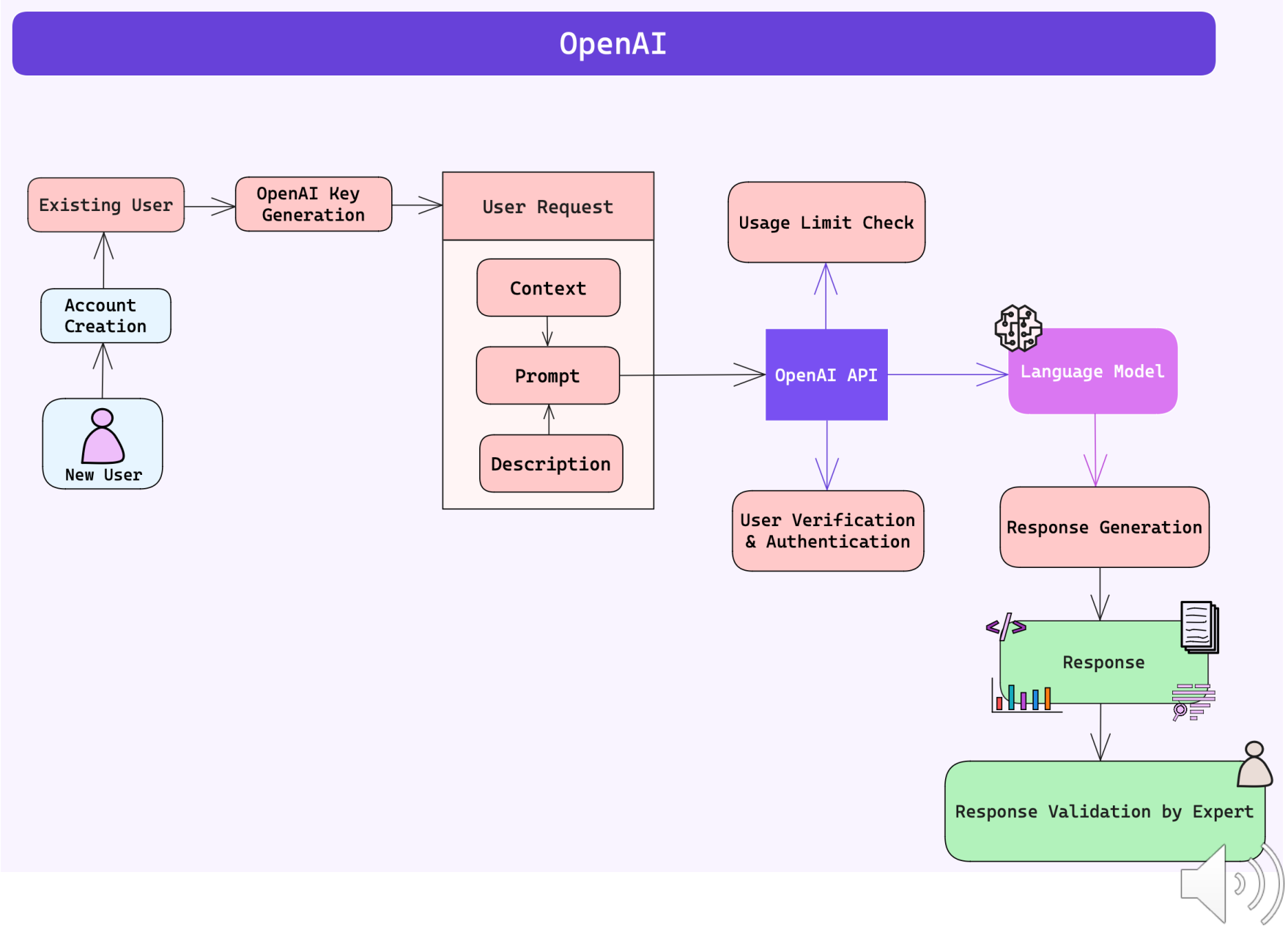
TOWARDS AGI:

DISCUSSIONS ON AGI OFTEN REMAIN SPECULATIVE, WITH LIMITED EMPIRICAL EVIDENCE TO SUPPORT CLAIMS.

Methodology Overview

- Experimentation on LLMs: GPT-4, GPT-3.5, Google Bard, Claude-2, LLaMA-2 (based on prompts designed by us).
- Assessing ability of the LLMs to play Games such as Tic-Tac-Toe, Connect Four, Chess.
- Automation of mathematical tasks using zero-shot prompting.
- Evaluation metrics: Game outcomes, adherence to game rules, and clarity of explanations.
- Proposed OpenAI's LLM Framework Overview of the process.

OpenAI API Framework

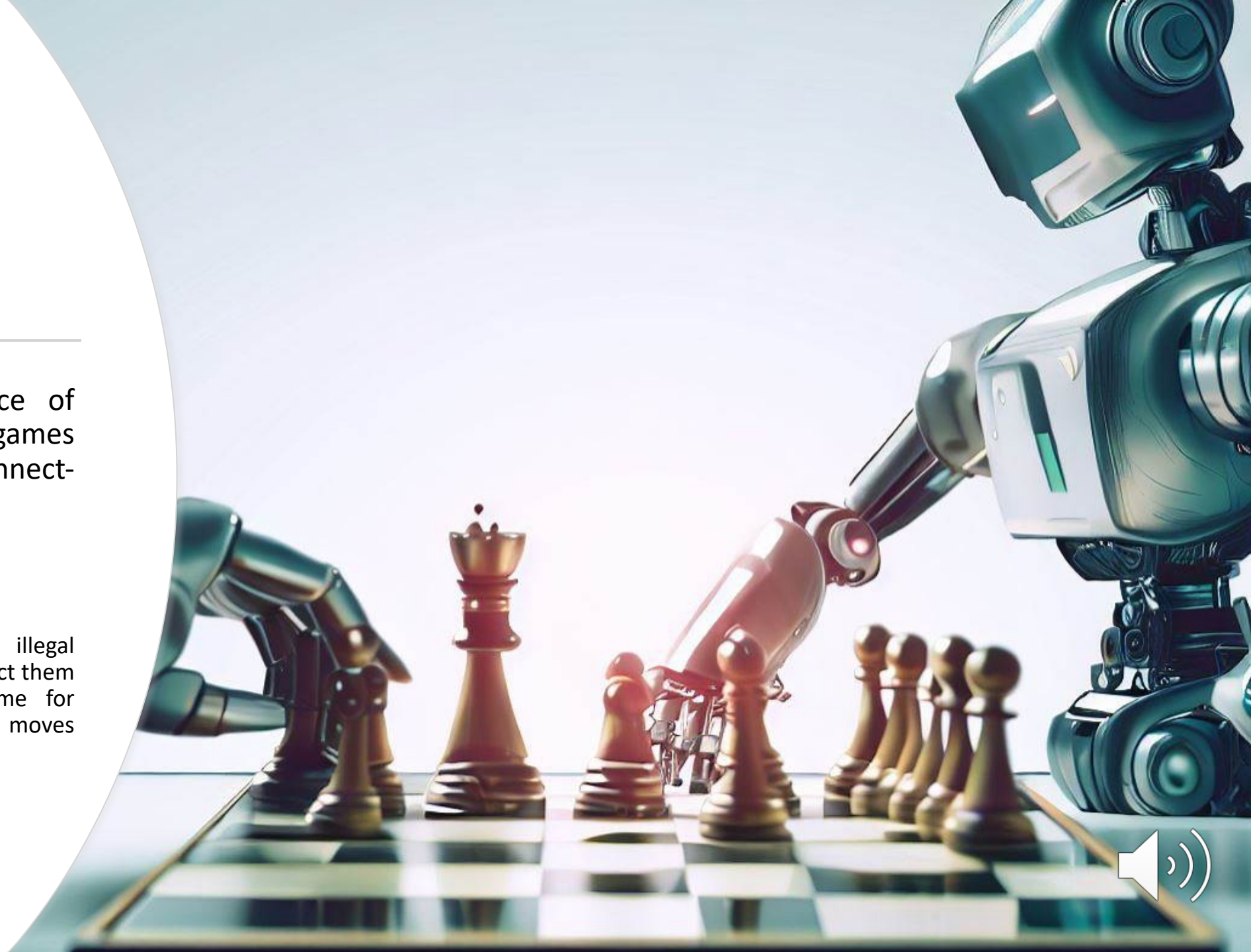


Analysis - Game Performance

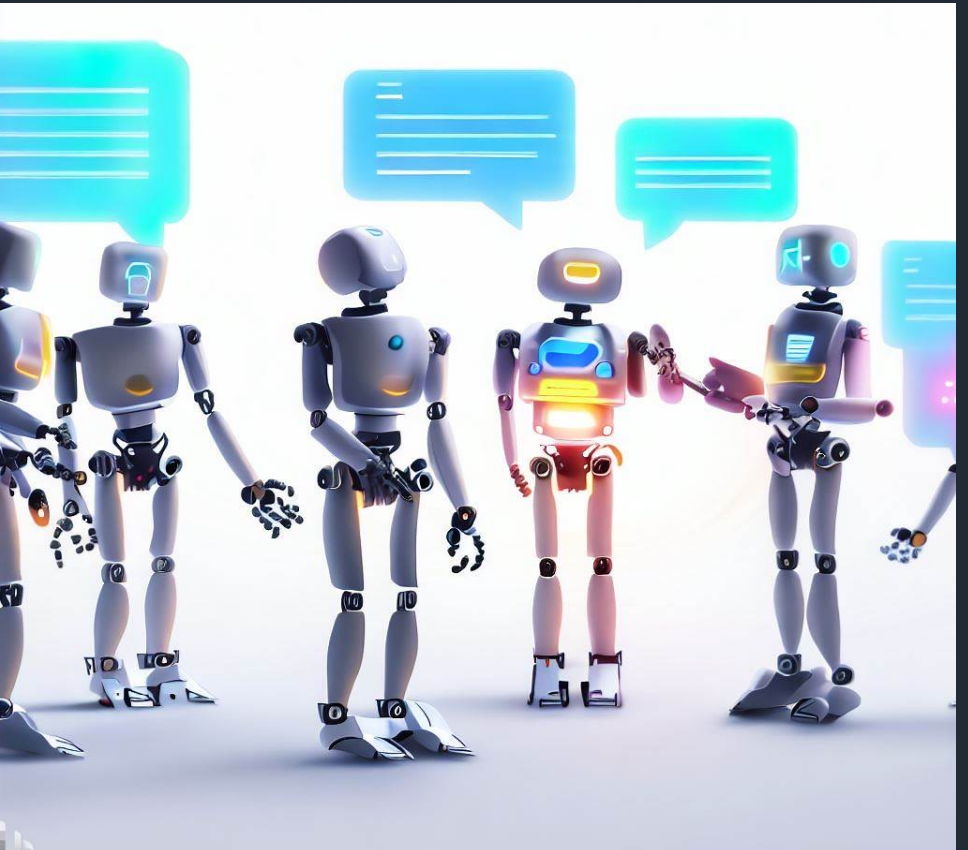
- Analyzed the performance of each LLMs in strategic games such as Tic-Tac-Toe, Connect-Four and Chess.

- Challenges:

  LLMs' tendencies to make illegal moves and attempting to correct them for continuation of the game for further analysis on the legal moves played by the LLM.



Analysis - Explainability



- Rationale for Explainability
- Differential Performance Across LLMs
- Insightfulness of Explanations
- Presentation of Explanatory Content

Key Conclusions and Achievements

- LLMs, including GPT-4, GPT-3.5, Google Bard, Claude-2, and LLaMA-2, exhibited diverse capabilities, both in game-play and in providing explanations.
- While LLMs showed promise in strategic games, they occasionally made illegal moves, highlighting the need for further refinement in game-based tasks.
- Our unique automation tasks, particularly the follow-up question system and game-play setup, provided deeper insights into GPT-3.5 LLM interactions and their potential for real-world applications.
- The project emphasized the significance of LLMs' explainability, with models varying in their depth and clarity of explanations.

Recommendations

REFINEMENT OF PROMPT ENGINEERING

TRANSPARENCY AND DOCUMENTATION

AGI VISION

ENHANCED GAME TRAINING

ROBUST EVALUATION FRAMEWORK

Closing Remarks

Significant Milestones

Quest for AGI

Harmonious Blend


Thank You

THE PLACE OF USEFUL LEARNING

University of Strathclyde

Science