Token-Based Decision-Making in Language Models for Interactive Environments

1. Project Summary

This project explores how **causal language models (LLMs)** can be used to interact with and learn policies in **interactive, strategic environments** (e.g., grid-based RPG-like or tactical decision games) by **treating both environmental observations and actions as sequences of discrete tokens**.

Unlike conventional RL methods, which rely on explicit state representations and value-based learning, our approach investigates the extent to which a **language model can act as a decision-making agent**—by learning to predict actions in a purely **autoregressive token prediction** paradigm.

We focus on modeling **non-textual, symbolic signals** (such as spatial layouts, object locations, player states) as tokens and studying how LLMs can interpret them within a unified sequence modeling framework.

---

2. Problem Statement

Can a language model, trained purely to predict tokens in a sequence, learn to make optimal decisions in a strategic environment by interpreting:

- **Observations** (the current game state) as a sequence of **tokens**, and
- **Actions** (agent's choices) as **tokens** to be predicted?

This raises several challenges:

- Designing a **tokenization scheme** for complex, non-linguistic observations.
- Maintaining **contextual memory** over long sequences of states and actions.
- Ensuring **generalization** to unseen environments and puzzles.
- Supporting **continual learning** as new levels and rules are introduced over time.

---

3. Learning Objectives

Students will gain hands-on experience in the following areas:

- **Language modeling for control**: Framing agent decisions as next-token prediction.
- **Tokenization of structured environments**: Translating spatial or symbolic data into discrete, sequential representations.
- **Reinforcement learning via imitation and feedback**: Learning from rollouts using language modeling loss, not reward-based policy gradients.
- **Continual learning and catastrophic forgetting**: Studying how models adapt or forget over evolving environments.
- **Evaluation of agents**: Using both language metrics (e.g., perplexity) and task-specific success metrics (e.g., completion rate, efficiency).

---

4. Nature of the Environment

The environments chosen for this project are **strategic, symbolic games** involving:

- **Grid-based world representations**.
- A player agent that can navigate, interact with objects, and reach defined goals.
- **Turn-based decision-making** where each action affects the state in discrete steps.
- **Partial observability or memory-dependent strategies**, which require reasoning over past actions and future consequences.

Examples may include:

- Tactical puzzle-solving games.
- Dungeon-style or RPG-like settings with resource management.
- Grid-based maps with evolving enemy or obstacle dynamics.

---

5. Key Innovations

- **Token-as-signal abstraction**: Unlike prior LLM applications focused on natural language or structured APIs, we reinterpret **environmental state changes** as sequences of discrete signals. These signals, though non-linguistic in nature, are treated identically to language tokens by the model.

- **Language modeling for decision-making**: Instead of using policy/value networks or sampling actions from logits over a state-action space, we pose decision-making as **sequence completion**, where action tokens are predicted as a continuation of observation tokens.

- **Unified architecture**: No separate modules for perception, planning, and action; the Transformer autoregressively handles all reasoning.

---

6. Relevant Topics Covered

- Large Language Models (Causal Transformers)
- Reinforcement Learning (Behavior Cloning, Reward Modeling)
- Representation Learning (Tokenization of structured environments)
- Continual Learning (Online updates, forgetting, curriculum design)
- Evaluation of decision models
- Imitation Learning and Planning as Inference

---

7. Related Work and References

Language Modeling for Agents

- **SayCan: Do As I Can, Not As I Say**
  Ahn et al., 2022 (Google Research)
  https://arxiv.org/abs/2204.01691

- **Decision Transformer: Reinforcement Learning via Sequence Modeling**
  Chen et al., 2021 (OpenAI)

https://arxiv.org/abs/2106.01345

Token-Based Representation in RL

- **Offline Reinforcement Learning as One Big Sequence Modeling Problem**
  Janner et al., 2021
  https://arxiv.org/abs/2106.02039

- **Plan4MC: Skill Reinforcement Learning and Planning for Open-World Long-Horizon Tasks**
  Yuan et al., 2023
  https://arxiv.org/abs/2303.16563

- **Solving Continual Offline Reinforcement Learning with Decision Transformer**
  Huang et al., 2024
  https://arxiv.org/abs/2401.08478

Continual Learning

- **Fine-tuned Language Models are Continual Learners**
  Scialom et al., 2022
  https://arxiv.org/abs/2205.12393

- **Avalanche RL: a Continual Reinforcement Learning Library**
  Lucchesi et al., 2022
  https://arxiv.org/abs/2202.13657

---

8. Final Notes

This project bridges **LLMs and RL** by viewing the entire agent-environment loop as a sequence to be learned. It encourages students to think beyond traditional action-value formulations and engage with cutting-edge ideas at the frontier of **token-based reasoning and language-driven control**. The abstraction of **signals as tokens** opens new paths for LLM-based agents in rich, symbolic domains like games, robotics, and planning.