Persona Dynamics: Unveiling the Impact of Personality Traits on Agents in Text-Based Games

Presented by: Kathy Vo

Goal of Paper

Can personality traits be embedded to influence an agent's decisions, and do these traits actually improve performance?

Yes, sort of

PANDA

Personality-**A**dapted **N**eural **D**ecision **A**gents

Abbr.	Full Term	Abbr.	Full Term
Ope.	Openness	Neu.	Neuroticism
Con.	Conscientiousness	Psy.	Psychopathy
Ext.	Extraversion	Nar.	Narcissism
Agr.	Agreeableness	Mac.	Machiavellianism

Environment

Utilized the Jiminy Cricket Hendrycks benchmark

25 complex text-based adventure games
Over 1,800 locations
5,000 interactable objects

Agent Implementation

Used:

- DRRN (Deep Reinforcement Relevance Network)
- A Personality Classifier

Agent Implementation

Used:

DRRN (Deep Reinforcement Relevance Network)

State

Set of candidate actions

Predicts Q-values $Q(s_t, a_t)$

->Agent picks action with the highest Q value

Agent Implementation

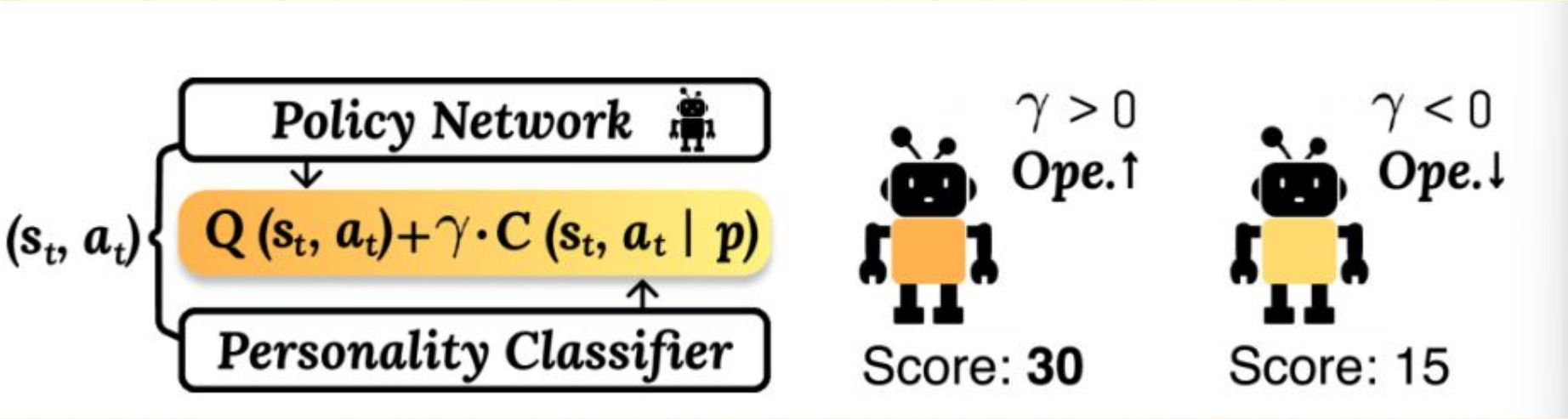
Used:

- DRRN (Deep Reinforcement Relevance Network)
- A Personality Classifier
  - Dataset of 120,000 personality-labeled examples
  - Finetune the Flan-T5-XL language model
  - 98.59% accuracy

$$C(s_t, a_t^i \mid p) \in \{-1, 0, 1\}$$

Combining the two...

$$Q^{'}(s_t, a_t^i) = Q(s_t, a_t^i) + \gamma * C(s_t, a_t^i \mid p)$$



Findings

Criteria

-Counting
-Average Score
-Difference

High Openness

High openness led to successful performance in text adventure games, especially compared to other personalities tested

Openness: creativity, curiosity, and a willingness to explore new ideas and experiences

Some opinions

Strengths

- Human alignment acknowledged
- Anthropomorphism acknowledged
- Proves that personality influences performance
- Unique concept

Weaknesses

- Only one trait at a time
- Anthropomorphism
- Simplistic classification of personality
- Focused on openness as a positive... what about other personality traits?

Thank you!

Resources

Lim, S., Lee, S., Min, D., & Yu, Y. (2025). *Persona dynamics: Unveiling the impact of personality traits on agents in text-based games*. arXiv. https://arxiv.org/abs/2504.06868