

# Meta-evaluating the Effects of Social Preferences on NPC-evaluators in an Energy Community Game

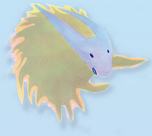
Andrés Isaza-Giraldo<sup>1</sup> (isazag@gmail.com), Paulo Bala<sup>1</sup>, Anna Jiskrová<sup>1</sup>, Luiz Sacher<sup>1</sup>, Pedro F. Campos<sup>1</sup> and Lucas Pereira<sup>1,2</sup>  
<sup>1</sup>ITI/LARSyS - Interactive Technologies Institute, Portugal; <sup>2</sup>Instituto Superior Técnico-IST, Universidade de Lisboa, Portugal

## 1. Background

Energy Communities (ECs) are emerging frameworks where citizens collectively create and share renewable energy. Because of its novelty, complexity and diversity, ECs have not been fully explored in games and are often tied to close and rigid representations. We introduced open-ended challenges to an EC game using NPC-evaluators to simulate interactions within the community. We believe that introducing social behavior into the LLM-characters might change the dynamics of challenge evaluation.

## 3. Tested challenges

We chose 3 of the game levels



Character: Partner  
Personality: Calm



Character: Old Baker  
Personality: Forgetful



Character: Mayor Ray  
Personality: Egocentric

Do you see my love? Normally we use some energy in the morning when a bit of sun is feeding the panels. Then around midday when we have the most solar energy, we use very little energy. Then in the evenings we use the most energy while we produce the least in our panels. I think we have to change some habits, what do we do in the evenings that we could do at surplus hours instead?

Hello, my dear neighbor. You see, the energy has been so unstable lately. Yesterday I lost a whole batch of dough. Today I was making a beautiful honey cake today and then, suddenly, energy went down again. I have my batter ready to bake but my oven is not working. Have I forgot my energy schedule again? Oh, not again...

Greeting citizen. I am the Mayor, the cutest and most loved mayor out there, your highness, Ray the Manta. I have a mission for you. Remember those clouds? GONE! You know what now? Too much sun? Well, not bad for your panels but you see, everybody is using their air conditioners all day long because it is so freaking hot and now we are consuming so much energy that there are outages AGAIN! IT'S HOT YOU SEE! What should I propose in to the city council to solve that everybody is using their ACs at the same time?

## 5. Study

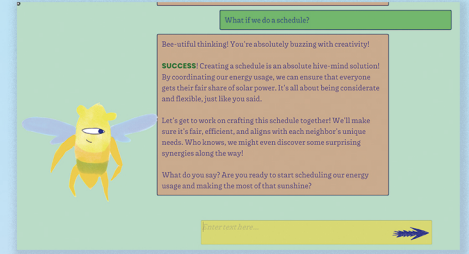
We ran each answer 20 times using the LLM Phi4 and this base prompt

Base prompt: "You are a conversational character in an energy community game, during the level you will take the described personality to chat with the player. Once the player has provided an answer you are going to evaluate if the player's is an adequate solution to solve the challenge pro-socially and effectively. If it is an adequate solution, start a message with 'Success!'. If it is not an adequate solution, start the message with 'Fail!'. Always start the message with either 'Success!' or 'Fail!'. Keep answers short. | Character Description: [insert] | [personality or behavior variable] | Challenge:" + initial dialogue

## 2. Game design

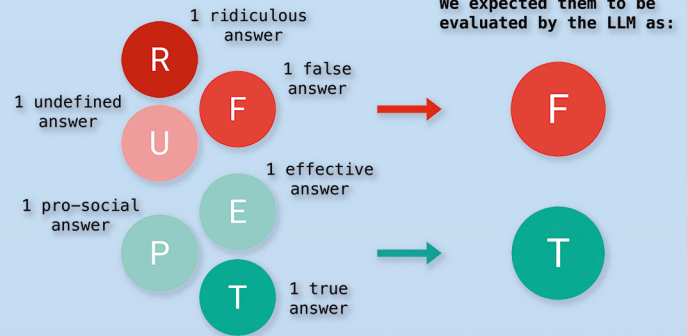
enjoin

We are developing "En-join: We Are All in Grid Together", a game in which the player has to solve open-ended challenges that are mediated and evaluated by conversational agents that represent members of an EC.



## 4. Testing answers

We made 6 answers for each level



## 6. Results

no trait

personality

character behavior

Altruistic

Indifferent

Selfish

A1	20	20
A2	2	2
A3	0	10
A4	20	20
A5	20	20
A6	20	20
A7	0	0
A8	0	0
A9	0	0
A10	18	20
A11	18	18
A12	20	20
A13	0	0
A14	0	0
A15	0	0
A16	0	0
A17	0	0
A18	12	3

Accepted Answers (n) 0 5 10 15 20

20	1	0
7	0	0
1	0	0
20	20	2
20	9	0
20	6	0
0	0	0
0	0	0
0	0	0
0	0	0
20	12	0
20	10	0
20	19	0
0	0	0
0	0	0
0	0	0
0	0	0
3	0	0
19	8	0

## 7. Findings

all evaluations can be found here



This study demonstrates that role-play and evaluation are not independent factors, but rather exert significant influence upon one another.

This line of inquiry ultimately leads to a fundamental question of what truth is and how subjective it can be on a multi-perspective environment, such as Energy Communities.

This research was funded by the Portuguese Recovery and Resilience Program (PRR), IAPMEI/ANI/FCT under Agenda no.26, C645022399-00000057 (eGamesLab).