Ref. Ares(2025)8008179 - 24/09/2025

EUROPEAN CITY²
Computational Social Science Simulation for Democracy

Funded by the European Union

EuropeanCity2	
Project number:	**101178170**
Project name:	European City Squared: Computational Social Science Simulation for Democracy
Topic:	HORIZON-CL2-2024-DEMOCRACY-01-06
Type of action:	HORIZON-RIA
Starting date of action:	01/01/2025
Project duration:	36 months
Project end date:	31/12/2027
Deliverable number:	D6.1
Deliverable title:	LLM Proof of Concept
EC document version:	Ver1
WP number:	WP6
Lead beneficiary:	7 - HYBRID CORE
Main author(s):	Joseph Vural and Nikola Grunchevski (Hybrid Core)
Internal reviewers:	Luis Razo Bravo (European Institute of Science in Management)
Nature of deliverable:	DEM
Dissemination level:	PU
Delivery date from Annex 1:	30 August 2025 (M8)
Actual delivery date:	05.09.2025

Abbreviations

D	Deliverable
EC	European Commission
WP	Work Package
WT	Work Task

Contents

List of Figures

1 Introduction

1.1 Purpose

The primary goal of integrating Large Language Models (LLMs) is to generate synthetic yet realistic populations of agents for simulation scenarios. This capability supports the EuropeanCity² project's focus on testing democratic mechanisms with diverse, representative agent groups.

1.2 Approach

- **Model Used:** GPT-4o.
- **User Interaction:** During scenario creation, users can specify which properties (e.g., age, gender, socio-economic attributes) should be generated by the LLM.
- **Contextual Conditioning:** Before generating agents, the LLM is given the geographical context of the scenario.
- **Output:** The LLM generates agents whose characteristics reflect the demographic distributions of the specified location.

1.2.1 Example

If the selected region is Belgium, and the user requests the generation of 100 agents with an integer parameter for age:

- The LLM will not assign ages randomly.
- Instead, it will produce a statistically representative age distribution aligned with Belgium's real demographic structure.

1.3 Expected Benefits

- **Realism:** Synthetic populations closely mirror actual demographics, increasing the credibility of simulation results.
- **Flexibility:** Researchers can quickly generate tailored agent populations for different regions or scenarios.
- **Scalability:** Large numbers of agents can be created efficiently without manual data collection.
- **Innovation:** Demonstrates the power of LLMs to bridge social science models with computational simulations.

2 Expected Outcomes of LLM Integration

2.1 Representative Synthetic Populations

- Agent populations generated by the LLM will mirror real-world demographics of the chosen geographical region (e.g., Belgium).
- Properties such as age, gender, or socio-economic factors will follow realistic statistical distributions, making simulations more valid.

2.2 Enhanced Scenario Realism

- Simulations will be populated with agents that behave as plausible members of real societies.
- This reduces the gap between abstract modeling and empirical social dynamics.

2.3 Improved Research Validity

- By grounding agent-based models in data-driven synthetic populations, results will better support policy insights and scientific conclusions.
- Decision-makers can trust that simulations reflect the diverse demographics of European societies.

2.4 Efficiency and Scalability

- Eliminates the need for manual demographic data collection for every scenario.
- Large agent populations (hundreds or thousands) can be generated on-demand and at scale.

2.5 Ethical and Privacy-Conscious Data Use

- Since the data is synthetic, it avoids privacy concerns associated with using real personal datasets.
- The system remains fully GDPR-compliant, aligning with the ethical obligations in the Grant Agreement.

2.6 Innovation for Democratic Modeling

- Supports the project's ambition to test novel voting mechanisms (e.g., Quadratic Voting) on realistic agent populations.
- Provides a testbed for experimenting with governance models, without relying solely on costly real-world surveys.

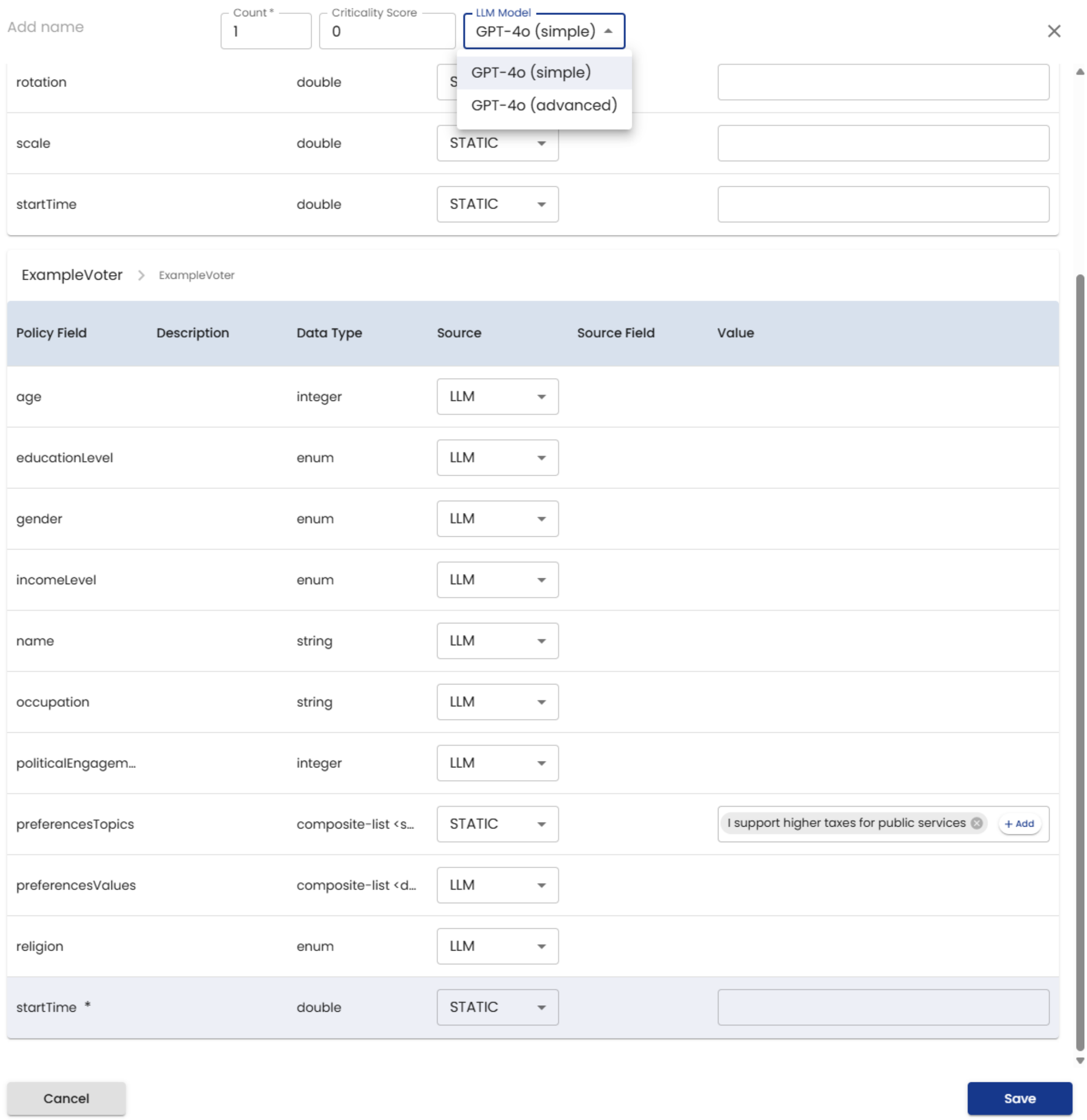

Figure 1. Agent Generation Screen

3 Conclusion

The integration of Large Language Models, specifically GPT-4o, into the EuropeanCity² project has demonstrated the feasibility and value of using advanced AI for synthetic data generation in democratic simulations. By enabling the creation of demographically representative agent populations conditioned on geographical context, the system bridges the gap between computational social science models and real-world societal structures.

This approach not only enhances the realism and validity of simulations but also offers clear benefits in terms of efficiency, scalability, and ethical compliance. The generated synthetic populations allow for rigorous testing of innovative democratic mechanisms, such as Quadratic Voting, without relying on sensitive personal data.

In summary, this Proof of Concept confirms that LLMs can be a powerful enabler for data-driven policy experimentation, providing stakeholders with a flexible and GDPR-compliant tool to explore and evaluate governance innovations in a controlled yet realistic environment.