Phase Transition In AI-Human Interaction Using Quartic Mean-Field Ising Model

Richard Kwame Ansah

Department of Mathematics and Statistics,University of Energy and Natural Resources

richard.ansah@uenr.edu.gh

DEEP LEARNING INDABA

Abstract

The essence of artificial intelligence (AI) machines in our society presents both opportunities and potential threats. To gain insights into this phenomenon, we propose a simulation model of a Human-AI ecosystem. This model considers various factors, including biased behaviour among agents, peer-to-peer interactions, three-body interactions, and also four-body interactions. The latter involves three humans interacting with an AI agent, two humans interacting with two AI agents, and three AI agents interacting with a human. Our main focus is to examine how the proportion of AI agents in the ecosystem influences its dynamics. Our analysis reveals compelling evidence that even slight changes in the percentage of AI agents can trigger significant and profound transformations within the system.

Introduction

In the realms of technological progress, few innovations have captured the world's imagination quite like Artificial Intelligence (AI). Unleashing the potential to simulate human intelligence in machines, AI has metamorphosed from a theoretical concept into a transformative reality. This cutting-edge technology has traversed multiple sectors, altering the way we live, work, and interact with our environment. From advanced natural language processing and computer vision to autonomous vehicles and personalized recommendations, AI's ever-expanding applications continue to shape the future in unprecedented ways[1].

Our approach involves constructing a sophisticated agent-based model, encompassing numerous parameters that can be refined through a complex fitting procedure. This endeavour aims to imitate reasonable behaviour within the system. The second approach, which we adopt in this paper, entails using a simplified model that retains essential elements of the original problem. With this analytical approach, we can push the boundaries of understanding[2].

Main Objectives

The main purpose of this research is to gain valuable insights into the behaviour of the human-AI ecosystem using the quartic mean field model.

Materials and Methods

We will analyse the Hamiltonian of an Ising model that consists of M spin configurations, defined as

$$T_M(\phi) = -\sum_{a,b,c,d=1}^{M} I_{a,b,c,d}\phi_a\phi_b\phi_c\phi_d - \sum_{a,b,c=1}^{M} J_{a,b,c}\phi_a\phi_b\phi_c - \sum_{a,b=1}^{M} K_{a,b}\phi_a\phi_b - \sum_{a=1}^{M} h_a\phi_a. \qquad (1)$$

$\Omega_M = \{-1,+1\}^M$ is the configuration space including all possible spins. In this context,

$$\phi_a = \begin{cases} +1, & \text{spin up,} \\ -1, & \text{spin down .} \end{cases}$$

The notation $\phi = (\phi_1, \phi_2, \ldots, \phi_M) \in \Omega_M$.

Mathematical Section

0.1 The one component quartic mean-field model

In continuation, our focus lies solely on the mean-field interaction between agents. By assuming a mean-field interaction, we fix $I_{a,b,c,d} = \frac{I}{4M^3}$, $J_{a,b,c} = \frac{J}{3M^2}$ $K_{a,b} = \frac{K}{2M}$ and $h_a = h$. The symbols I, J, K, and h represent the coupling of spins in a quartic, cubic, and binary spin, respectively, while h denotes the strength of an external magnetic field. Therefore, the Hamiltonian per particle is

$$T_M(\phi) = -M\left(\frac{I}{4}m^4 + \frac{J}{3}m^3 + \frac{K}{2}m^2 + hm\right), \qquad (2)$$

where

$$m = \frac{1}{M}\sum_{a=1}^{M}\phi_a \qquad (3)$$

is the magnetization per particle of the configuration ϕ. The probability distribution for the choice $\phi = (\phi_1, \phi_2, \cdots, \phi_M)$ is described by the joint probability distribution given as:

$$\psi_{M,I,J,K,h}(\phi) = \frac{e^{-T_M(\phi)}}{D_M}, \quad \text{where} \quad \phi \in \Omega_M, \qquad (4)$$

with D_M representing the partition function of the model, defined as:

$$D_M = \sum_{\phi\in\Omega_M} e^{-T_M(\phi)}. \qquad (5)$$

The function s_M that represents the pressure function of the model can then be expressed as

$$s_M = \frac{1}{M}\ln D_M. \qquad (6)$$

For a given observable f, the Boltzmann-Gibbs expectation $\tilde{w}_M(f)$ is defined as follows:

$$\tilde{w}_M(f) = \frac{\sum_\phi f e^{-T(\phi)}}{\sum_\phi e^{-T(\phi)}}. \qquad (7)$$

In the thermodynamic limit, which refers to very large values of M, the behaviour of the model is dictated by the pressure function. Therefore, we analyse how the pressure function behaves as M approaches infinity.

Let T_M be a Hamiltonian as defined in 2 and let I, J, K, and h be fixed parameters. Define the pressure per particle as $s_M = \frac{1}{M}\ln D_M$, where D_M is the grand canonical partition function for a system with M particles. Then, if the thermodynamic limit $\lim_{M\to\infty} s_M = s$ exists, then it can be expressed in the following equivalent form: $s = \sup_{\bar{\mu}} s_{lower}$, with

$$s_{lower} = \frac{I}{4}\bar{\mu}^4 + \frac{J}{3} + \frac{J}{3}\bar{\mu}^3 + \frac{K}{2}\bar{\mu}^2 + h\bar{\mu} - \frac{1+\bar{\mu}}{2}\ln\left(\frac{1+\bar{\mu}}{2}\right) + \frac{1-\bar{\mu}}{2}\ln\left(\frac{1-\bar{\mu}}{2}\right), \quad \bar{\mu} \in [-1,1].$$

0.1.1 Exact solution of the model

Even though s_{lower} and s_{upper} are different functions, they share the same local maxima. This can be shown by differentiating both functions with respect to the parameter $\bar{\mu}$ and noticing that the Mean Field Equations yield identical extremality conditions for both functions:

$$\bar{\mu} = \tanh\left(I\bar{\mu}^3 + J\bar{\mu}^2 + K\bar{\mu} + h\right) \text{ where } \bar{\mu} \in [-1,1]. \qquad (8)$$

Results

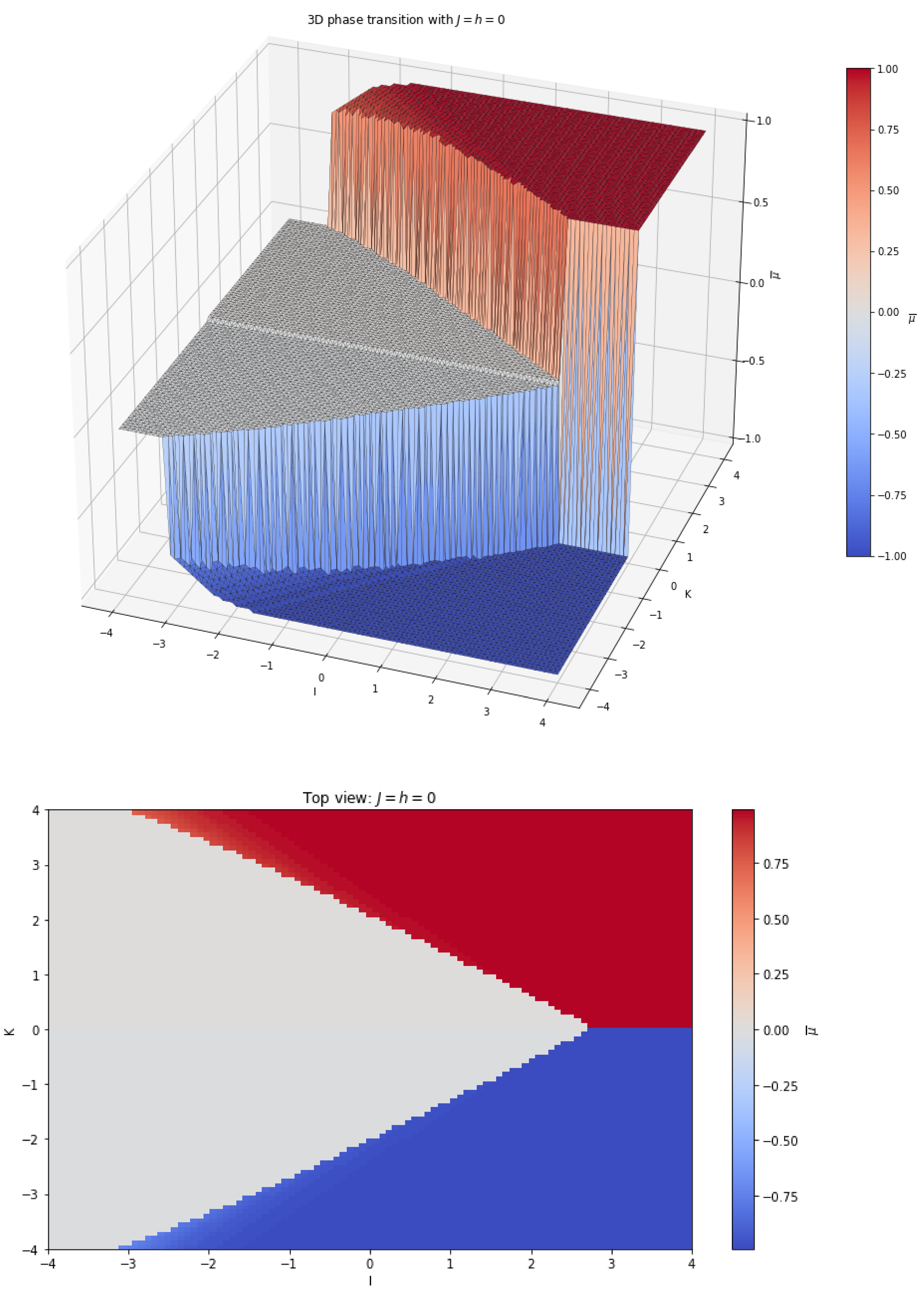


Figure 1: For $h = J = 0$ 3D and $h = J = 0$ top view, the phase diagram of the stable solutions in equation 8 illustrates the coexistence curves. For $I < 2.8$, three distinct phases are observed: the negatively polarized phase (in blue), the zero or unpolarized phase (in gray), and the positively polarized phase (in red). Consequently, if K increases progressively from negative to positive values in this region, it encounters two consecutive jumps.

0.2 The two component quartic mean-field model

Let's divide the group of M agents into two distinct subsystems, referred to as AI and H, containing M_1 and M_2 agents respectively and let AI $\cap$ H $= \emptyset$. Consider $\bar{\mu}_1$ and $\bar{\mu}_2$ as the mean opinions regarding the AI and H subsystems, respectively. Additionally, let's define γ_1 and γ_2 as the relative sizes of AI and H agents, represented by $\frac{M_1}{M}$ and $\frac{M_2}{M}$, respectively. In the context of the two-component quartic mean-field model, the energy contribution can be expressed as follows:

$$\begin{aligned} G(\bar{\mu}_1, \bar{\mu}_2) &= \frac{1}{4}\left[I_{1111}\gamma_1^4\bar{\mu}_1^4 + 4I_{1112}\gamma_1^3\gamma_2\bar{\mu}_1^3\bar{\mu}_2 + 6I_{1122}\gamma_1^2\gamma_2^2\bar{\mu}_1^2\bar{\mu}_2^2 + 4I_{1222}\gamma_1\gamma_2^3\bar{\mu}_1\bar{\mu}_2^3 + I_{2222}\gamma_2^4\bar{\mu}_2^4\right] \\ &+\frac{1}{3}\left[J_{111}\gamma_1^3\bar{\mu}_1^3 + 3J_{112}\gamma_1^2\gamma_2\bar{\mu}_1^2\bar{\mu}_2 + 3J_{122}\gamma_1\gamma_2^2\bar{\mu}_1\bar{\mu}_2^2 + J_{222}\gamma_2^3\bar{\mu}_2^3\right] + \frac{1}{2}\left[K_{11}\gamma_1^2\bar{\mu}_1^2 + 2K_{12}\gamma_1\gamma_2\bar{\mu}_1\bar{\mu}_2 + K_{22}\gamma_2^2\bar{\mu}_2^2\right] \\ &+ \left[h_1\gamma_1\bar{\mu}_1 + h_2\gamma_2\bar{\mu}_2\right] \end{aligned} \tag{9}$$

The large number limit of the generating functional (6) linked to the two-component quartic mean-field model (9) can be expressed in its variational form as follows:

$$\sup_{\bar{\mu}\in[-1,1]^2} \nabla(\bar{\mu}) = \sup_{\bar{\mu}\in[-1,1]^2} \left[G(\bar{\mu}_1, \bar{\mu}_2) - (\gamma_1 Q(\bar{\mu}_1) + \gamma_2 Q(\bar{\mu}_2))\right] \tag{10}$$

where

$$Q(\bar{\mu}) = \frac{1+\bar{\mu}}{2}\ln\left(\frac{1+\bar{\mu}}{2}\right) + \frac{1-\bar{\mu}}{2}\ln\left(\frac{1-\bar{\mu}}{2}\right). \tag{11}$$

The stationary solutions $\bar{\mu}$ of ∇ can be described as follows

$$\bar{\mu}_d = \tanh\left(h_d + \sum_{t,s,r=1}^{2}\gamma_t\left(K_{dt} + \gamma_s J_{dts}\bar{\mu}_s + \gamma_s\gamma_r I_{dtsr}\bar{\mu}_s\bar{\mu}_r\right)\bar{\mu}_t\right) \text{ for } d = 1, 2. \tag{12}$$

Figure 1 illustrates the surfaces representing the solutions of equation (12), which correspond to the global maxima of the function ∇ as described in equation (10). These surfaces are displayed with respect to the free parameters γ and $I_{1112} = I_{1122} = I_{1222} = I$, while keeping the remaining parameters fixed at specific values.

Figures 2 exhibit the total average opinion surfaces obtained from a two-component quartic mean-field model.

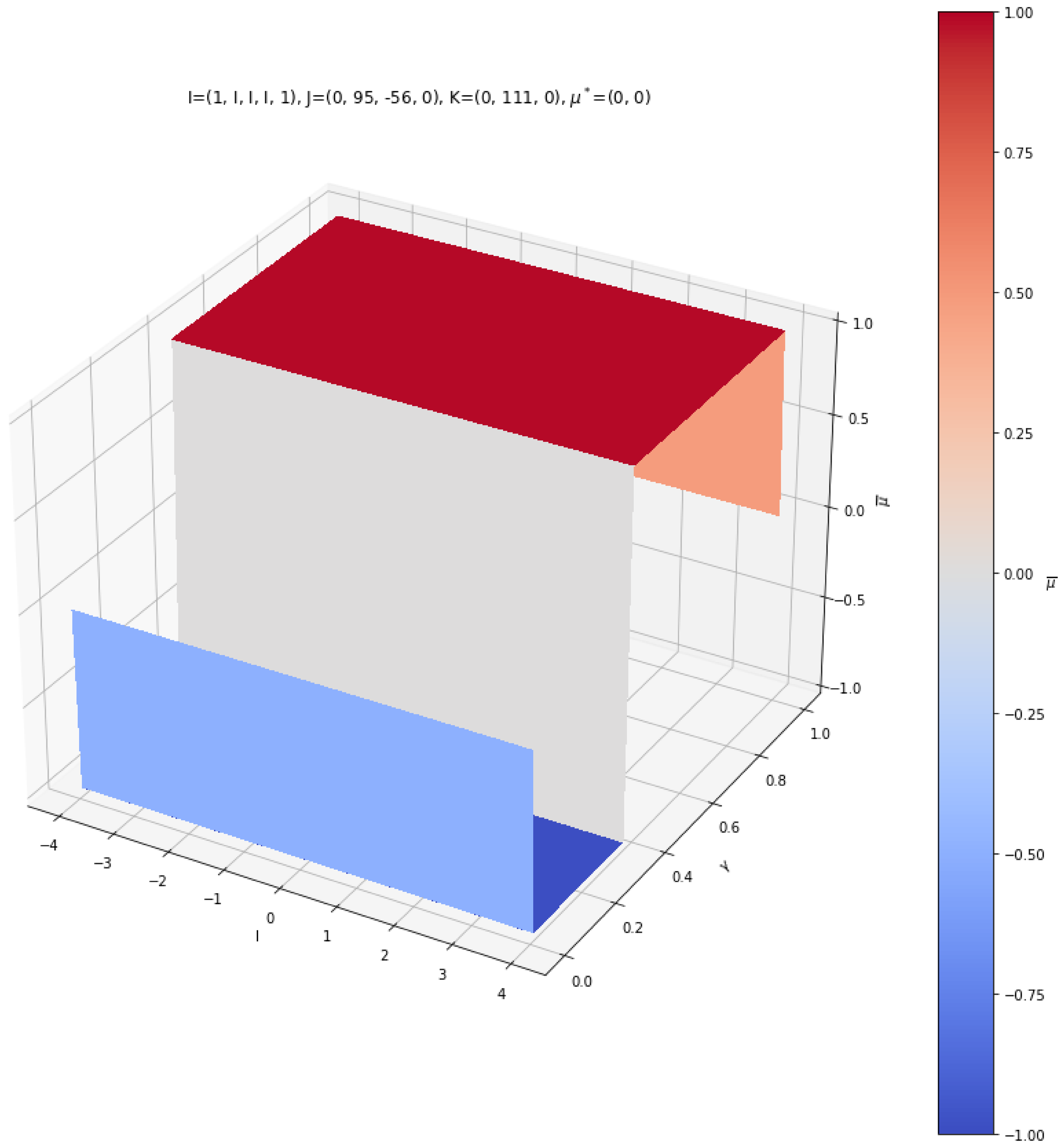
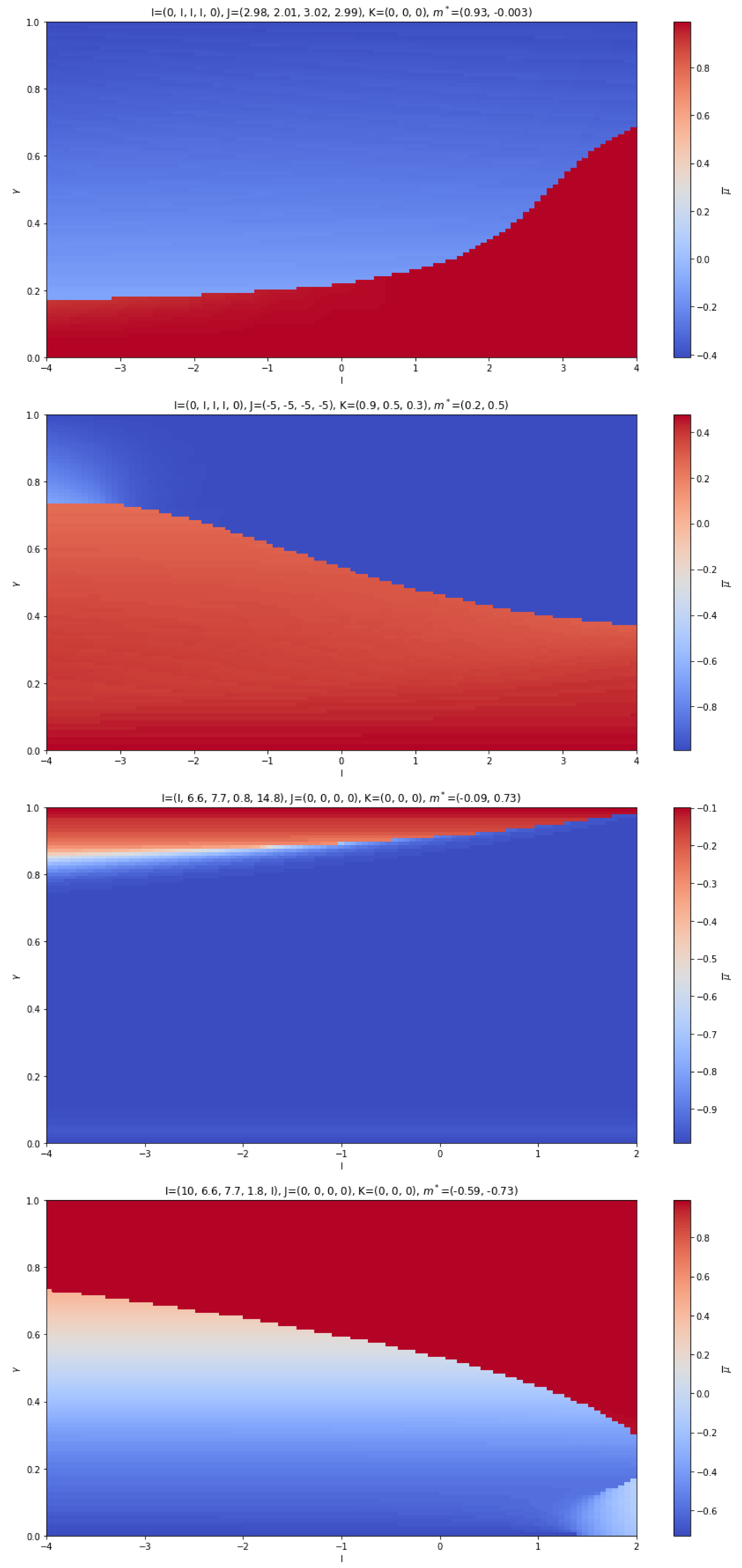


Figure 2: $I = (1, I, I, I, 1), J = (0, 95, -56, 0), K = (0, 111, -0.24), \mu^* = (0, 0)$

Conclusion

This section introduces phase diagrams of the model, exploring the correlation between parameter γ and one of the interacting variables. The demarcation between the different colours represents distinct opinion phases, with a special focus on highlighting the occurrence of first-order phase transitions. These transitions manifest as sudden shifts in opinions, leading to abrupt changes in the colour distribution, as visually depicted in Figure 3.



The analysis of Figure 3 indicates that the model parameter's appropriate values lead to simulations where even a tiny proportion of AI agents can trigger abrupt behaviours in opinion formation within the Human-AI ecosystem. This finding aligns with Figure 3 demonstrates, showing that in a system with limited interaction among Human agents $I_{2222} = I$, a smaller fraction of AI agents can still drive phase transitions and prevalent opinion formation. In Figure 3, when $I_{1112} = I_{1122} = I_{1222} = I$ is small the system require a smaller fraction (i.e. γ) of the AI agents to observe a phase transition and as $I_{1112} = I_{1122} = I_{1222} = I$ increases, the proportion of AI agents required for a phase transition increases. We observe from Figures 3 that when interaction among AI agents ($I_{1111} = I$) increases their proportion needed to observe a phase transition increases. We observe that interaction among human agents ($I_{2222} = I$) increases their proportion needed to observe a phase transition also increases.

References

[1] Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach.-New York, NY, USA: Pearson.

[2] Contucci, P., Kertész, J., & Osabutey, G. (2022). Human-AI ecosystem with abrupt changes as a function of the composition. PloS one, 17(5), e0267310.

[3] Alberici, D., Contucci, P., Mingione, E., & Molari, M. (2017). Aggregation models on hypergraphs. Annals of Physics, 376, 412-424.

Acknowledgements

The authors thank the staff at the Mathematics and Statistics Department of University of Energy and Natural Resources and Department of Mathematics, Kwame Nkrumah University of Science and Technology for their support and kindness during the period this paper was written.