YouTube's Silent Echo

Team Members: Simon Li, Vincent Tran, Isaac Ibidun
GitHub Repo: https://github.com/mayowaibi/EECS4461-Project/

§1. Phenomenon Overview

1.1 Phenomenon Statement

Echo chambers are a common issue on social media platforms like YouTube, where recommendation algorithms and social bots reinforce people's existing beliefs while also filtering out different perspectives. This process, as described in the Brookings Report (Brown et al., 2023), creates mild ideological echo chambers, which means users gradually see more content that aligns with their views, and opposing viewpoints become less visible over time. Hussein et al. (2020) further highlight that YouTube's recommendation system can contribute to filter bubbles, where engagement-driven visibility amplifies specific narratives, often including misinformation. Social bots take advantage of this by boosting certain videos (using likes, comments, and shares) to make content seem more popular than it actually is. This manipulation of engagement-based ranking algorithms makes it easier for echo chambers to form (Grusauakaite et al. 2024). This tricks the algorithm into recommending the same kinds of videos repeatedly, making it harder for users to come across different viewpoints. Over time, this reinforces selective exposure, pushing people into tighter ideological bubbles (Brown et al., 2023). Our project simulates how social bots and YouTube's recommendation system interact to reinforce echo chambers. Using agent-based modelling, we will analyze how bots, algorithms, and users contribute to the formation of ideological clusters, which is similar to what happens in real life.

1.2 Problem Statement

Understanding how AI-driven recommendations and social bots influence user exposure is crucial because engagement-based ranking algorithms shape public discourse, contribute to political polarization, and spread misinformation. While YouTube's algorithm is designed to maximize watch time, Ledwich and Zaitsev (2019) argue that this reinforces algorithmic extremism, where users are increasingly exposed to ideologically aligned content. Cinelli et al, (2021) further emphasize that engagement-based visibility favours content that generates interaction, even if that content is polarizing. This self-reinforcing exposure cycle leads to deeper ideological segregation over time. The issue is compounded by social bots, which manipulate engagement to boost specific content artificially, influencing organic user interactions. To address this, our project seeks to answer three key research questions:

- How does YouTube's recommendation algorithm and social bots influence what users see?
- To what extent do AI-driven interactions reinforce ideological bubbles?
- Are there ways to reduce ideological reinforcement and improve content diversity?

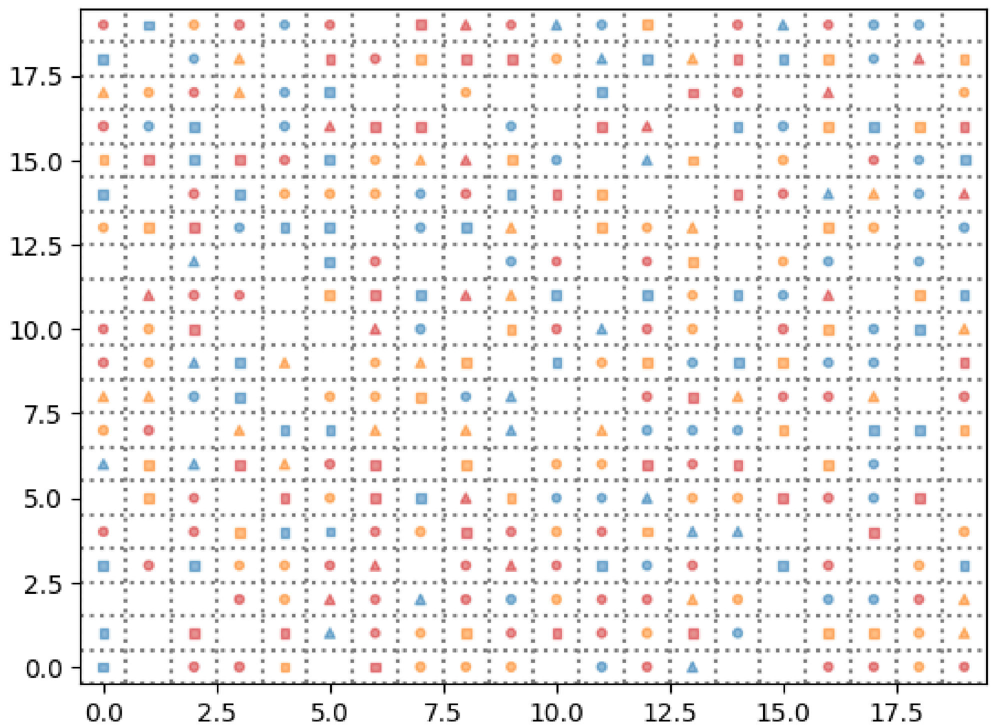
1.3 Why ABM?

Agent-Based Modelling (ABM) is well-suited for studying emergent social media dynamics, where large-scale ideological shifts result from small individual interactions. Hussein et al. (2020) argue that algorithmic bias in content exposure is not deterministic, but rather emerges from user engagement behaviours. ABM allows us to model heterogeneous agents (humans, social bots, and AI-driven recommendations) to simulate how their interactions lead to systemic content clustering over time. Cinelli et al. (2021) note that homophily and engagement reinforcement drive real-world echo chamber formation, making ABM an effective tool for replicating these processes. Using the Mesa agent-based framework, our model tracks cluster formation, engagement trends, and AI influence over time, providing insight into how algorithmic bias and social bots contribute to ideological reinforcement.

1.4 Phenomenon Illustration

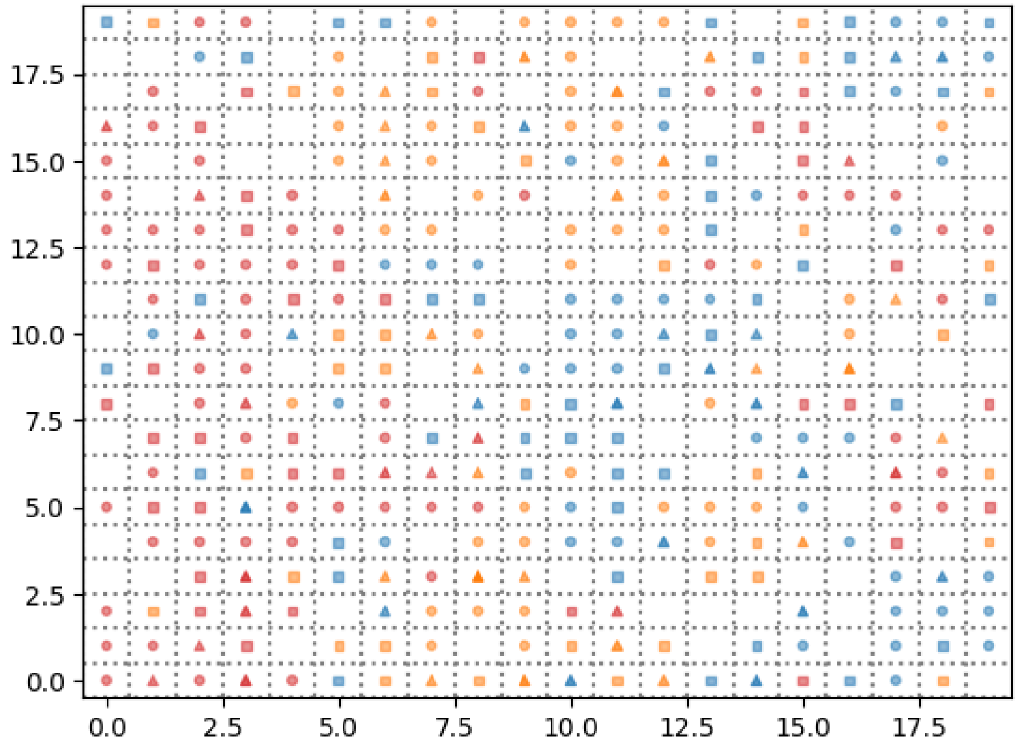
Initial State:

- The simulation begins with a randomized grid of agents. Human users (circles), social bots (squares), and recommendation algorithms (triangles) are distributed across the grid, each assigned a colour corresponding to their content preference (orange for video games, blue for sports, and red for politics).
- At this early stage, agents have not yet begun forming clusters, and there is no clear content alignment or ideological segregation. Similar to a new YouTube user, everyone is exposed to a diverse range of content with no personalized recommendations or significant bot influence.



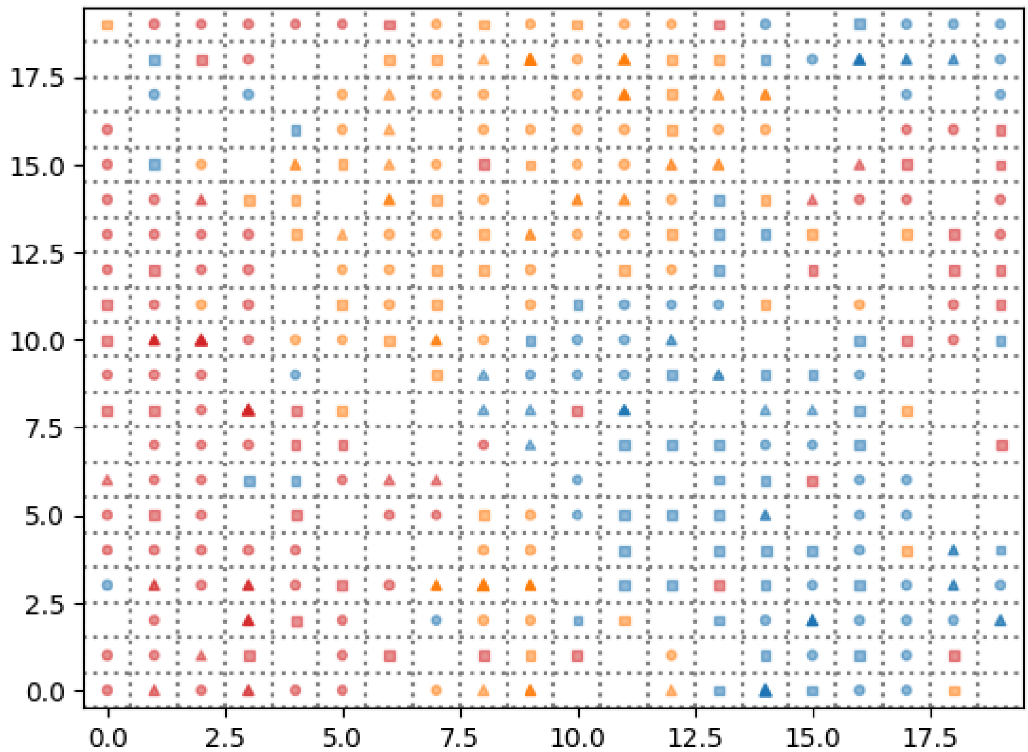
Middle State:

- As the simulation progresses, we observe the early formation of ideological clusters. Human agents begin to move towards agents with similar preferences due to homophily, while social bots reinforce content by engaging with like-minded neighbours. Recommendation algorithms demonstrate their growing influence on user visibility and behaviour.
- The influence of AI becomes more visible, as recommendations begin to adapt based on user engagement. Social bots amplify engagement signals, and content clustering becomes more pronounced.



Final State:

- By the end of our simulation, we see a good representation of echo chambers. Human users are now mostly surrounded by agents who share their content preferences, and AI agents (particularly recommendation algorithms) have adapted to maximize engagement and preference alignment.
- Social bots and recommendation agents have significantly shaped the content landscape, with clear ideological clusters visible across the grid. This final state mirrors real-world platform dynamics where exposure is heavily curated, reinforcing the users original views.



§2. Simulation Design & Implementation

2.1 System Overview

Our simulation explores the formation of echo chambers on YouTube using an agent-based modeling framework developed with the Mesa library. The model captures the dynamics between three agent types: human users, social bots, and recommendation algorithms. Its primary aim is to investigate how engagement-driven interactions and AI-guided reinforcement contribute to ideological clustering over time. Each agent functions independently, reacting to its immediate surroundings and social context, allowing complex macro-scale patterns to emerge from simple micro-level behaviors.

Agents are initialized with unique content preferences, engagement levels, and homophily thresholds that determine their comfort with ideological diversity. Deployed onto a grid, they interact with nearby agents based on simple, interpretable rules. Human users assess their local environment and relocate if their neighbors' preferences diverge too far from their own. In contrast, AI agents — both social bots and recommendation algorithms — work to influence their surroundings. Bots boost engagement metrics and amplify specific content, while recommendation algorithms learn from user activity to adaptively deliver personalized suggestions.

Over time, both types of AI agents increase their impact. Social bots coordinate through clustering to strengthen amplification, while recommendation agents build user profiles and adjust their strategy using reinforcement-based feedback. These interactions are shaped not only by immediate neighborhood dynamics but also by evolving network connections that represent ongoing relationships and influence strength between agents. These cumulative behaviors lead to the emergence of ideological silos without any centralized control.

Echo chambers in the model are not manually imposed or predefined. Instead, they form organically through cycles of exposure, engagement, adaptation, and repositioning. The simulation reproduces real-world patterns like filter bubbles and algorithmic bias by allowing these dynamics to play out over time.

The model is parameterized to allow experimentation with different proportions of AI agents, homophily levels, and engagement intensity, offering insight into how subtle changes in system dynamics can drastically affect large-scale outcomes. This flexibility supports a broad range of scenario testing and makes the simulation a valuable tool for exploring how design choices in recommendation systems might affect user exposure and ideological diversity.

2.2 Simulation Environment

Our simulation operates within a grid-based environment implemented using the Mesa agent-based modelling framework. Each agent in the simulation (humans, social bots, and recommendation algorithms) occupies a cell in a toroidal grid, meaning agents that move past one edge of the grid reappear on the opposite edge. This prevents artificial boundary effects and allows for continuous interactions between agents. Agents interact with neighbours within a defined radius based on homophily, engagement levels, and AI influence. The gird's density parameter determines the population concentration, affecting how frequently agents encounter each other and form clusters.

Movement in the grid is then governed by the agent's happiness calculations, which combine content preference matching bot influence and engagement-based modifications. The homophily principle is used to model selective exposure, which means that humans prefer to interact with others who share their content preferences.

In addition to local proximity, agents form and update network connections based on repeated interactions. These links persist over time and affect agent influence, simulating longer-term digital relationships and indirect exposure. Recommendation algorithms use reinforcement learning to analyze user behavior and adaptively recommend content, further shaping the engagement landscape in ways that are not purely spatial.

Prior research done by Hussein et al. (2020) has shown that YouTube's recommendation system can lead to a filter bubble effect, where users who engage with misinformative content are subsequently recommended more of the same content. Our simulation reflects this mechanism by modelling how social bots and AI-driven engagement influence content exposure, leading to the gradual formation of echo chambers. This environment structure allows us to observe the emergence of ideological clusters, replicating how online users interact with algorithmic recommendations and bots in a networked media ecosystem.

2.3 Agent Design

Human Users: Represented as circles, human agents are initialized with a random content preference (gaming, sports, or politics) and operate based on local satisfaction calculations. Their decision to move is governed by a homophily threshold: if the similarity of their neighborhood falls below this threshold, they relocate to an empty cell in search of a more aligned environment. The similarity score is a weighted function of content preference matching, nearby bot influence, and engagement dynamics (likes, comments, shares). Engagement is probabilistic and driven by each agent's engagement rate, with each action type increasing the agent's homophily. This simplified behavioral model captures core elements of selective exposure and echo chamber reinforcement on YouTube without modeling psychological nuance.

Social Bots: Represented as squares, these AI agents actively amplify content through cluster formation and engagement mechanics. Their influence is calculated through their "amplification power" attribute that increases based on cluster size and engagement levels. Bot clusters form when multiple bots with the same preference are within proximity. These bots follow deterministic rules rather than probabilistic learning, abstracting real-world coordination tactics used in platform manipulation. Their behavior is designed to simulate artificial boosting of content visibility on YouTube using simplified, scalable heuristics.

YouTube Recommendation Algorithm: Represented as triangles, these agents analyze user preferences and generate content recommendations using a learning system that combines user preference tracking and bot influence. Their recommendation effectiveness is then measured through success_rate, recommendation_strength, and influence_reach. This model simplifies Youtube's real recommendation system into a learning model that adapts based on user engagement and bot activity, focusing on the reinforcement of existing preferences rather than preference modification.

Key Computational Challenges: while creating these agents, we ran into a few challenges that ultimately changed the way we designed them. These challenges included:

- Balancing the interaction weights between different influence factors (such as bot influence or engagement influence) to prevent any single mechanism from dominating the formation of echo chambers
- The implementation of network influence dynamics without creating any recursive loops in the connection strength calculations
- Managing the computational load of tracking multiple agent relationships and influence metrics while maintaining a relatively consistent performance
- Ensuring that stable bot cluster form without creating any unrealistic super-nodes or cascade effect (unexpected behaviours we saw in our early prototype)

2.4 Interaction Dynamics

Agent behavior unfolds through step-based iterations using Mesa's built-in random activation strategy. At each time step, agents are activated in random order, ensuring that no individual agent consistently acts first or last. This design choice was made to avoid update bias and better reflect the asynchronous and decentralized nature of online interactions.

Each agent performs a decision loop based on its type. Human users evaluate their immediate neighborhood and determine whether they are satisfied ("happy") based on a composite similarity score that incorporates content alignment, bot presence, engagement feedback, and network connection strength. If their environment does not meet their homophily threshold, they relocate to a new grid position. This mechanism simulates self-directed content curation and ideological sorting behavior found in digital environments.

Bot-to-bot interactions are emergent. Social bots identify and cluster with others who share their content preference. Clustering increases their collective "amplification power", which strengthens their ability to influence nearby humans. These coordinated patterns mirror real-world botnets, where collective behavior boosts visibility and perceived popularity of targeted content. Bots do not communicate directly but coordinate implicitly through spatial proximity and shared preference signals.

Recommendation agents interact with human users in a non-spatial manner, maintaining and updating personalized profiles over time. These agents adapt their behavior using reinforcement learning, modifying their recommendation strategy based on the success or failure of previous suggestions. They also integrate nearby bot cluster activity into their content selection logic, allowing manipulation tactics to influence algorithmic decisions — an abstraction of real-world vulnerabilities in recommendation systems.

Grusauskaite et al. (2024) notes that echo chambers are often self-sustaining because users reinforce their own viewpoints through active engagement and avoidance of opposing content. This model captures that effect by allowing ideological clusters to form, grow, and resist disruption through recursive engagement and localized AI-driven reinforcement.

2.5 Data Collection & Visualization:

To analyze the emergence of echo chambers and the influence of algorithmic agents, the simulation tracks a range of dynamic metrics related to agent behavior, clustering, influence, and engagement. These metrics are collected at each simulation step and visualized in real time to support interpretability and scenario analysis. Key metrics include:

Happy Agents %: Measures the proportion of agents whose local environment matches their homophily threshold, representing ideological satisfaction.

Social Bot Cluster %: Tracks the prevalence of bot clusters, quantifying coordinated amplification behaviors.

Echo Chamber Strength: Represents the average similarity between agents and their neighbors, serving as a high-level indicator of ideological polarization.

Recommendation Influence Reach: Tracks the total number of successful influence events by recommendation agents, reflecting the scale of algorithmic impact.

Average Recommendation Strength: Measures how confidently recommendation agents are promoting content based on learned user preferences.

Recommendation Success Rate: Tracks how often recommendations are accepted by users, reflecting the adaptivity and effectiveness of the AI.

Average Connection Strength: Quantifies the strength of persistent network ties between agents, highlighting emergent social structures beyond spatial proximity.

Network Density: Represents the proportion of potential connections that are active, providing insight into how tightly knit the information network becomes.

Engagement Metrics: Total likes, comments, and shares serve as proxies for attention and content reinforcement in the system.

Key Computational Challenges: During the implementation of data collection and visualization, a couple challenges emerged that influenced how metrics were tracked and displayed:

- Ensuring the local (grid-based) and persistent (network-based) influences in the model don't overwrite or double-count influence effects was a challenge.
- With multiple AI subtypes, standard agent tracking functions could not treat all AI equally. Visualizations and summaries had to dynamically differentiate agent roles and apply role-specific metrics — e.g., success rate for recommendation system, amplification for bots.

§3. Observations & Results

3.1 Simulation Results

Our final simulation results demonstrate the dynamics of Youtube echo chambers by modeling the interactions between human users, social bots and recommendation algorithms. Using Mesa's agent-based modeling with Solara visualization we are able to quantitatively and qualitatively observe how the symbiotic relationship between different AI agents accelerate the formation of content bubbles. At the start of the simulation, human agents (circles), social bots (squares) and recommendation algorithms (triangles), each coloured to represent their corresponding content preferences on Youtube (orange for gaming, blue for sports, red for politics), are distributed randomly on a grid. As the simulation progresses, these agents move based on their content preferences, homophily levels, and emerging influence of AI interactions. This visualization then reveals five key behaviours that emerge from our echo chamber simulation: 1) agents with similar content preferences cluster together, 2) social bots become more opaque as their amplification power grows, 3) recommendation algorithms increase in size based on their recommendation strength and success rate, 4) human agents grow larger within clusters due to higher engagement levels and 5) network connections from and strengthen between agents.

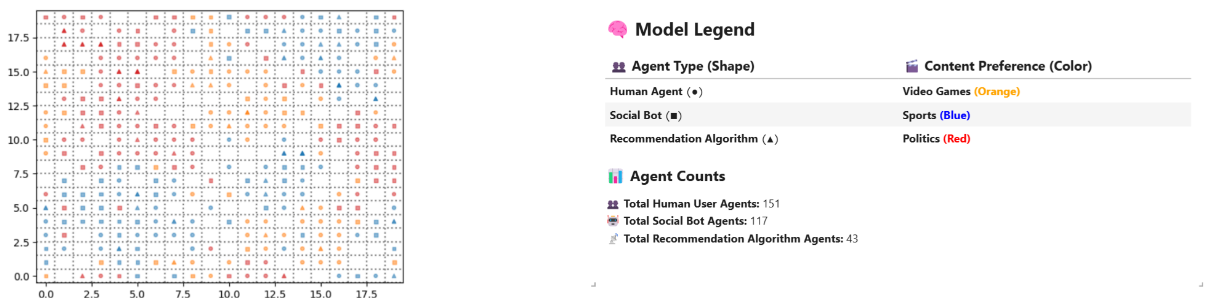
The simulation then tracks these behaviours through several key metrics displayed on our simulation's plots. Arguably the most important metric that our simulation has is echo chamber strength. This plot reveals a consistent pattern where the average strength rises from initial values of 0-20% to stabilize between 70-80% at the end of the simulation. This metric is the combination of content similarity (50%), bot influence (30%), and engagement factors (20%), providing a comprehensive measure of echo chamber intensity in our simulation. This rapid increase in the echo chamber strength correlates strongly with both bot cluster formation and recommendation algorithm success rates as it is driven by the symbiotic relationship between the two where successful recommendation algorithms (success rate > 0.6) promote bot-influenced content, leading to a 10% amplification boost for aligned bot clusters. In turn these strengthened bot clusters provide more influential content for the recommendation algorithms to promote (increasing their bot influence weight from 0.3 to 0.4-0.5). The echo chamber plot also includes two more important metrics: happy agents percentage and social bot cluster percentage. As echo chamber strength increases, we observe a corresponding rise in the happy agents percentage, indicating that agents become more satisfied as they settle into content bubbles. On the other hand, the social bot cluster percentage shows an interesting pattern where it initially spikes as bot rapidly form clusters, but then stabilizes around 30-35% for the remainder of the simulation, suggesting that while not all bots form clusters, those that do maintain stable and highly influential groups. Collectively these three metrics demonstrate how a stable percentage of clustered bots can result in echo chamber formation, leading to increased agent satisfaction within these reinforced content bubbles.

Another key behaviour is the network influence dynamics we see between agents. This can be seen on our network plot that tracks two crucial metrics: average connection strength and network density. Connection strength typically starts low (0.3-0.4) but increases significantly throughout the simulation as agents form stable connections, eventually reaching 0.6-0.7 in mature echo chambers. Network density shows a relatively similar pattern, rising from initial values around 0.2 all the way up to 0.6-0.7 as bot clusters form and recommendation algorithms begin promoting bot-influenced content. These metrics then reveal to us how AI agents create increasingly interconnected networks that reinforce content bubbles. This interconnectedness is particularly evident when we examine bot clusters with high amplification power (>2.0). These create strong connection pathways that other agents tend to follow and as a result recommendation algorithms amplify this effect by promoting content from these well-connect networks (shown by bot influence weight increasing to 0.4-0.5). This structure creates resilient echo chambers that become increasingly difficult to disrupt as both connection strength and density continue to reinforce each other throughout the simulation.

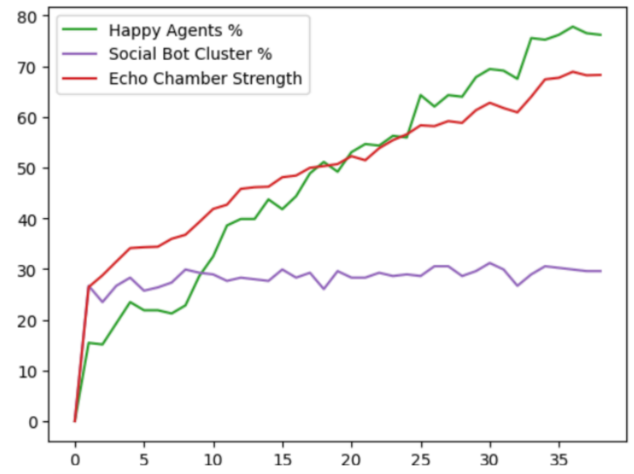
The recommendation plot displays three key metrics that describe the algorithmic influence recommendation agents have on other agents: recommendation success rate, average recommendation strength and recommendation influence reach. An interesting pattern we observed is the recommendation success rate remains relatively low, stabilizing around 0.2, indicating that recommendation algorithms struggle to consistently influence user preferences. This is then reflected in the average recommendation strength metric, which declines from its initial value of 1.0 to stabilize around 0.6, suggesting that recommendations become less impactful over time. However, the recommendation influence reach percentage spiking to 1.0 and maintaining this level indicates that while recommendation agents may not be highly successful at changing user preferences, they are highly effective at reaching users with content that matches their existing preferences. This combination of metrics demonstrates how recommendation algorithms contribute to the echo chamber formation by reinforcing existing preferences across the user population while not changing user preferences at the same time, effectively creating content bubbles where users are exposed to content they already favour.

Finally the engagement plot tracks likes, comments, and shares, showing how the interaction patterns on Youtube evolve within echo chambers. These metrics continuously grow throughout the simulation, with a clear hierarchy in engagement types: total likes accumulate the fastest, followed by comments and then shares showing the slowest growth rate out of the three. This pattern seems to align with interaction behaviour seen on Youtube as liking content requires the least effort, commenting represents moderate engagement and sharing content demonstrates the highest level of user commitment. In terms of echo chamber simulation, these three metrics suggest that echo chamber formation consistently drives user engagement, with the relative proportions of these engagement types remaining relatively stable even as total numbers grow.

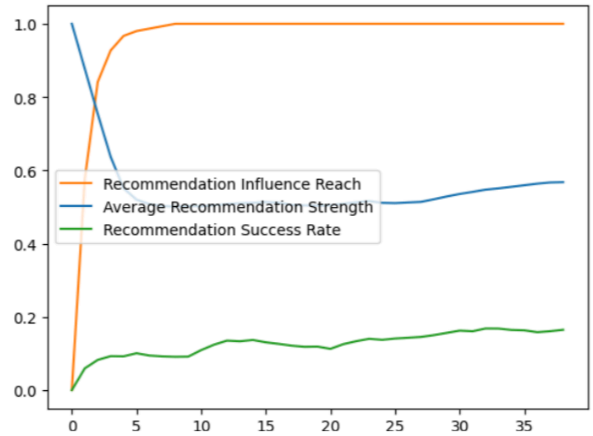
In conclusion, these quantitative and qualitative results effectively illustrate how Youtube echo chambers emerge through AI-to-AI interactions. The relationship between bot clusters and recommendation systems seem to create a powerful feedback loop where both AI types enhance each other's effectiveness. Bot clusters provide influential content for recommendations, while successful recommendations strengthen both influence, creating increasingly stable and resilient echo chambers.



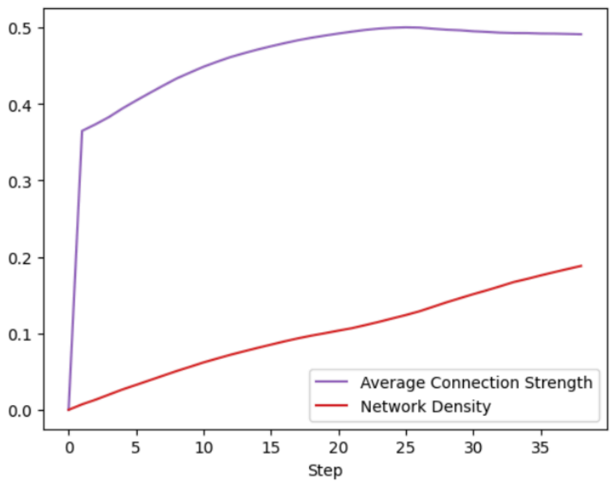
Agent Grid: visually showcases the above mentioned five key behaviours and agent attributes



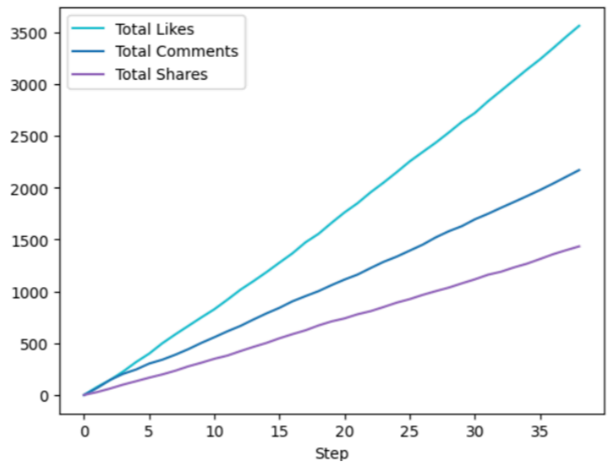
<u>Echo Chamber Plot:</u> tracks the formation and strength of content bubbles



<u>Recommendation Plot:</u> tracks algorithmic influence of recommendation agents on other agents around them, showing how algorithms reinforce existing preferences rather than changing



<u>Network Plot:</u> tracks the evolution of agent connections, showing how agents form and maintain relationships within the network



Engagement Plot: tracks user interactions with content

3.2 Unexpected Behaviors

Our final simulation of Youtube echo chambers revealed several unexpected behaviours, providing us with interesting insights into how AI interactions influence content bubbles. Rather than seeing a steady increase in bot clustering or a high stable amount of bot clusters, we observed an initial spike followed by stabilization around 30-35%. This suggests that while not all bots form clusters, a consistent third of the bot population maintains stable influential groups. This surprising emergent behaviour seems to be due to the interaction between the "ai_homophily" parameter (0.4) and the "amplification_power" mechanism, where bots find an optimal balance between clustering and maintaining influence over a broader area.

Another interesting result emerged from the recommendation algorithm metrics. While we initially anticipated that recommendation algorithms would actively change user preferences with a relatively high success rate, our simulation showed a different pattern. The recommendation success rate remained consistently low (around 0.2), and the average recommendation strength declined from 1.0 to stabilize around 0.6. However, the recommendation influence reach maintained a constant maximum value of 1.0. This unexpected behaviour suggests that recommendation algorithm agents contribute to the echo chamber phenomenon not by changing user preferences, but by effectively reinforcing existing preferences across the entire network. This behaviour seems to have emerged from the interaction between the recommendation algorithm's learning parameters and the "bot_influence_weight" mechanism, where algorithms adapt to promote content that aligns with user preferences rather than attempting to make them diverse.

§4. Ethical & Societal Reflections

4.1 Ethical Considerations

Our project does not use real-world user data, and all interactions within the simulation are artificially generated through agent-based modelling. However, the study of social bots and AI-driven recommendation algorithms raises broader concerns about data privacy, transparency, and potential misuse. One ethical concern is the ability of AI-powered recommendations to shape user engagement in ways that may not be immediately visible to the user. This raises questions about algorithmic accountability, which discusses whether platforms should be responsible for how their algorithms influence content exposure. Additionally, if a similar model were applied using real user data, privacy concerns would emerge regarding the collection and use of engagement behaviours to fine-tune recommendation systems. Another concern relates to the potential for manipulation. Social bots are designed to artificially amplify engagement, which can make certain content appear more popular than it actually is. This could be exploited in political or commercial contexts to manipulate public perception. If similar AI-driven reinforcement mechanisms were leveraged unethically, they could be used to spread misinformation, reinforce biases, or potentially even manipulate elections. Our simulation highlights how subtle algorithmic biases can lead to long-term content filtering, which raises the ethical dilemma of whether platforms should intervene to mitigate these effects or allow recommendation algorithms to operate freely.

An example from our project that demonstrates this concern about algorithmic accountability and content filtering is the recommendation algorithm behaviour. Despite a low success rate in changing user preferences, it maintains maximum influence reach across the network by reinforcing existing preferences. This mirrors real-world concerns about recommendation systems prioritizing user engagement over content diversity, showing how recommendation systems can significantly shape content exposure, but appear to only serve user preferences. Our implementation of bot clusters and their interaction with recommendation algorithms also raises questions about detection and regulation– if a group of clustered bots can significantly influence content distribution, how can media platforms identify and counter such coordinated behaviour?

4.2 Societal Implications

At the micro level, our model reflects how individual users interact with AI-driven content filtering. The key issue is selective exposure, where users engage primarily with like-minded content while rejecting opposing viewpoints. This phenomenon, sometimes referred to as the filter bubble effect, limits content diversity and reinforces ideological preferences. One concern is that users may not be fully aware of how recommendation systems influence their content consumption, which creates a passive form of ideological reinforcement rather than active engagement with diverse perspectives.

At the meso level, our simulation demonstrates how AI-driven engagement affects online communities and group interactions. Social media ecosystems are structured in a way that encourages content clustering (whether through user behaviour or algorithmic reinforcement). This is particularly relevant when considering how misinformation spreads within closed networks. If bots or AI algorithms prioritize engagement above content quality, they may unintentionally promote misleading or extreme narratives that gain traction simply due to high user interaction. Governance measures such as content diversification strategies or transparency in algorithmic decision-making could help mitigate these risks.

At the macro level, our project connects to broader debates about AI governance and media regulation. Platforms like YouTube are not just content hosts but they also actively shape public discourse by determining what information is surfaced and amplified. This has implications for democracy, journalism, and political engagement. While regulation of AI-driven recommendations is still an evolving field, our model suggests that long-term engagement reinforcement could contribute to larger societal polarization if left unchecked. The challenge is balancing freedom of information with the need to prevent algorithmic bias from reinforcing ideological segregation.

The overall ethical concern is whether similar simulations could be used for malicious purposes. While our project aims to analyze and mitigate algorithmic bias, the same principles could be applied to optimize content manipulation strategies. For example, political or corporate entities could use AI-driven engagement analysis to target specific audiences with precision, reinforcing narratives that align with their objectives. This risk underscores the importance of ethical AI development, where transparency and accountability are prioritized to prevent unintended harm from recommendation algorithms.

§5. Lessons Learned & Future Directions

5.1 Design and Development Reflections

Throughout the development of our simulation, one of the more persistent challenges was fine-tuning the agent behaviours to realistically reflect the dynamics of echo chamber formation on platforms like YouTube. Initially, our AI agents would become too influential too quickly, dominating the simulation and forming echo chambers too quickly which lacked organic progression. To address this, we had to calibrate the amplification effects, adjusting various variables such as amplification power and cluster size influence to ensure a more realistic evolution of agent behaviour over time.

Another significant challenge was designing an appropriate stopping condition. Early in our development process, we used a relatively simplistic condition where the simulation would stop when all agents were considered “happy”. However, this condition wasn’t perfect due to persistent ideological diversity, which made some agents perpetually unhappy. In response, we implemented a fixed step limit, which solved the initial problem but sacrificed responsiveness to the actual state of the simulation. For our final version, we implemented a more dynamic stopping condition based on a plateau in average happiness levels, ensuring the simulation terminates naturally once agent satisfaction fully stabilizes.

5.2 Model Limitations & Areas for Improvement

While our simulation captures many of the key dynamics involved in echo chamber formation, it’s definitely not a full representation of real-world systems. One limitation is the simplicity of agent decision-making. Although we incorporate things like homophily, engagement, clustering, and reinforcement learning, real users and algorithms operate under far more complex rules that involve more nuanced preferences, contextual awareness, and content variety.

Similarly, it is also important to note that our model currently only considers three content categories, which limits preference overlap and ideological blending. Another simplification lies in how engagement is quantified. We track likes, comments, and shares using probabilistic models, but do not simulate the full diversity of user interaction patterns such as watch time, skipping content, or feedback loops involving external trends.

Another simplification lies in how engagement is quantified. We track likes, comments, and shares using probabilistic models, but do not simulate the full diversity of user interaction patterns such as watch time, skipping content, or feedback loops involving external trends.

5.3 Future Applications

We believe that our findings have potential implications across several domains. From a policy standpoint, the simulation demonstrates how small changes in engagement amplification can lead to significant ideological segregation over time. This supports the argument for greater transparency in algorithmic recommendation systems and could inform regulatory discussions around AI accountability and content curation on digital platforms.

On the other hand, for platform designers and content moderators, the model could serve as a testing ground to explore how tweaks to recommendation logic might impact user clustering, engagement diversity, and exposure to alternative viewpoints. Researchers who are studying misinformation, political polarization, or online radicalization could use similar frameworks to model how various actors contribute to content clusters.

From a systems perspective, this project could serve as the foundation for more advanced work in multi-agent simulation platforms where emergent behavior is difficult to predict from individual rules. Our model illustrates that even a seemingly simple reward structure (e.g., maximizing engagement) can produce complex, self-reinforcing biases.

Finally, our project contributes to broader discussions in AI safety and digital ethics. By simulating the emergent effects of reinforcement-based algorithms, it highlights the need for proactive design interventions that prioritize user agency and reduce systemic bias.

§6. References

Brown, M. A., Nagler, J., Bisbee, J., Lai, A., & Tucker, J. A. (2023, October 26). *Echo chambers, rabbit holes, and ideological bias: How YouTube recommends content to real users.* Brookings Institute.

https://www.brookings.edu/articles/echo-chambers-rabbit-holes-and-ideological-bias-how-youtube-recommends-content-to-real-users/

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). *The echo chamber effect on social media.* Proceedings of the National Academy of Sciences, 118(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

Grusauskaite, K., Carbone, L., Harambam, J., & Aupers, S. (2024). *Debating (in) echo chambers: How culture shapes communication in conspiracy theory networks on YouTube.* New Media & Society, 26(12), 7037–7057. https://doi.org/10.1177/14614448231162585

Hussein, Eslam, Prerna Juneja, and Tanushree Mitra. (2020). *"Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube."* Proceedings of the ACM on Human-Computer Interaction, vol. 4, no. CSCW1, May 2020, Article 48, 27 pages. https://doi.org/10.1145/3392854.

Ledwich, M., & Zaitsev, A. (2019). *Algorithmic extremism: Examining YouTube's rabbit hole of radicalization.* First Monday, 24(12). https://doi.org/10.5210/fm.v25i3.10419

§7. Attestation

7.1 Documentation and Overview of Contributions

Simon:

- Found references 1-5, which were used to back some of our decisions/findings and were cited in APA format (section 5)
- Wrote the majority of section 1: Established the real-world relevance of echo chambers and the rationale behind using ABM to study them.
- Contributed to section 2, co-writing the System Overview and helping refine how agent roles and interactions contribute to clustering phenomena.
- Wrote section 4: Critically examined the implications of AI-driven recommendations, not manipulation, selective exposure, and the risks of algorithmic bias on individuals and society at large.
- Contribute to section 5: Reflected on key development challenges, identified model limitations, and proposed future refinements for the simulation.

Vincent:

- Wrote section 3 “Observations & Results”
    - described how our simulation results illustrate echo chambers on Youtube, providing quantitative metrics and qualitative descriptions of emergent behaviours
    - included graphs to support the key findings
    - interpreted what these results reveal about our chosen phenomenon
    - described any unexpected behaviours we observed in our simulation
- Contributed alongside other group members for sections 2 & 4

    - In subsection 2.3, described the types of agents in the simulation and their behaviours, explained the rule-based behaviours governing agent interactions, explained the choices made to simplify sophisticated agent behaviour and explained key issues in the computational instantiation of the agent design
    - In subsection 4.1, added specific examples from our simulation that raise privacy concerns if they were to use real-world user data

Isaac:

- Wrote the majority of section 2, expanding on our previous descriptions for the system overview, simulation environment, agent design, interaction dynamics and data collection/visualization based on our updates since DEL 3.
- Contributed alongside other group members for section 5:
    - Described the limitations of our prototype in relation to preference overlap, ideological blending, and engagement metrics.
    - Described the potential for our simulation to be used as a foundation for multi-agent simulations from a systems perspective.

7.2 Assignment of CRediT Roles:

Simon:

- Data curation: Lead // Located and selected all scholarly references used in the report
- Writing - original draft: Lead // Wrote the drafts of Sections 1 and 4; contributed to Section 2 and 5.
- Writing - review & editing: Supporting // Reviewed and edited simulation description sections to ensure clarity and accuracy.
- Methodology: Supporting // Contributed to describing the agent-based modelling framework and system-level interactions in Section 2
- Visualization: Supporting // Assisted with interpreting and explaining the significance of clustering patterns and emergent trends during the simulation demo and analysis.

Vincent:

- Project administration: Lead // assigned tasks to each team member for prototype and final report
- Validation: Equal // looked over and verified what each team member wrote
- Writing - original draft: Lead // responsible for writing section 3
- Formal analysis: Lead // interpreted simulation results and described behaviours
- Writing - review & editing: Supporting // refined sections 2 & 4
- Methodology: Supporting// described our agents, their behaviours and computational instantiation challenges in section 2.
- Visualization: Equal // provided visualizations of prototype in section 3

Isaac:

- Writing - original draft: lead // wrote the majority of section 2
- Writing - review & editing: supporting // refined section 5 with more details
- Methodology - lead // described the system overview, simulation environment, agent design, interaction dynamics and data collection/visualization

- Investigation - supporting // investigated and briefly described the relationships between our simulation and chosen phenomenon in section 2

7.3 CRediT Acknowledgement

All authors have read and approved the final report.

Simon: Data curation (Lead), Writing - original draft (Lead: sections 1 & 4), Writing - review & editing (Supporting), Methodology (Supporting), Visualization (Equal)

Vincent: Project administration (Lead), Validation (Equal), Writing - original draft (Lead: section 3), Formal analysis (Lead), Writing - review & editing (Supporting), Methodology (Supporting), Visualization (Equal)

Isaac: Writing - original draft (Lead: section 2), Writing - review & editing (Supporting), Methodology (Lead), Investigation (Supporting)