Understanding and Forecasting User Behavior in Social Communities: A Synthesis of Artificial Intelligence Approaches

1st Ashutosh Kumar Singh
BE- CSE Computer Science
Chandigarh University
Delhi, India
ashu9936.aks@gmail.com

2nd Nitin Kushwah
BE- CSE Computer Science
Chandigarh University
Agra, India
nitinkushwah19213@gmail.com

3rd Nitesh Choudhary
BE- CSE Computer Science
Chandigarh University
Himachal Pradesh, India
choudharynitesh692@gmail.com

4th Sukhpreet Singh
Assistant Professor
Chandigarh University
Mohali, India
sukhpreet.e17779@cumail.in

***Abstract*—This paper explores how Artificial Intelligence (AI) can be used to understand and predict user behavior in online social communities, with specific attention paid to its implications and methods. We begin by examining the multifaceted nature of user behavior, delving into actions, motivations, and social-graph interactions. Next, we examine the range of AI frameworks, from fundamental predictive analytics to advanced modeling paradigms like GNNs for relational dynamics, Transformer-based models for information diffusion, and ABMs (agent-supported models) for simulating emergent social phenomena. From data entry to preprocessing and feature engineering to model evaluation, the report provides a comprehensive overview of the predictive pipeline from start-to-end. In summary, we bring attention to the profound ethical dilemmas related to algorithmic bias, data privacy, and accountability that are inevitably associated with the growing potency of predictive technologies. This compilation acts as "the ultimate guide for researchers and practitioners, guiding them through the technical capabilities and ethical obligations of this rapidly developing field**

***Index Terms*—Artificial Intelligence, User Behavior Prediction, Online Social Communities, Graph Neural Networks, Transformers, Agent-Based Models, Predictive Analytics, Algorithmic Bias, Data Privacy, Ethical AI.**

I. Introduction: The Digital Agora and the Predictive Imperative

The online social community has become a vital component of modern society in the 21st century. The societal structures have transformed from small, specialized virtual teams into large, decentralized cyberspaces that promote interaction and cooperation among mass crowds.versiune theorem: Digital agoras that function as the hub of information dissemination, public opinion shaping, construction and economic activity have been replaced by these communities. Why? These platforms are colossal in terms of the amount of data generated by billions more users worldwide on a daily basis. The most comprehensive documentation of human social behavior in history is a digital exhaust that is produced as supplementary material during everyday conversations.

At the intersection of social science and computer science, this data overload presents a double issue. The first Challenge is one of understanding. he behavior of individuals in complex ecosystems is influenced by a combination of psychological motivations and strong social influences, such as group norms or peer pressure. The unraveling of this intricate causality is a fundamental scientific challenge. Prediction is the second obstacle posed. Predicting user behavior, such as which contents to consume, which products to buy, what communities to become members of, or when to exit a platform altogether — all of these are critical factors for users. Predictive need is driven by both commercial and other needs, ranging from improving user engagement to effectively allocating resources, as well as mitigating the spread of misinformation and identifying harmful activities.

The convergence of these two challenges has propelled Artificial Intelligence (AI) to the forefront of social media analysis. By utilizing machine learning, AI can provide the necessary computational power to comprehend and analyze user behavior at greater than human capacity. Through the use of tools, it transforms raw data into actionable insights, enabling the transition from observing behavior to actively forecasting it.

II. The Nature of User Behavior in Online Communities

A robust and intricate comprehension of the information being projected is necessary before any attempt at making a prediction. It is a multifaceted and intricate construct that online social communities cannot be reduced to merely one measure of user behavior. To ensure accurate predictive modeling, it is essential to define the observable actions of individuals, including their psychological drivers and the social structures that shape and constrain those actions. The shift in perception of user behavior from a simple count of clicks to heightened understanding has occurred. The socio-psychological phenomenon has led to a direct evolution in the sophistication of AI models designed for analyzing it. State-of-the-art forecasting cannot be achieved through sequential methods alone; instead, it must incorporate static user attributes, dynamic relational data and semantic content of interactions to form a comprehensive and predictive view.

A. A Multi-faceted Definition of Behavior

When it comes to using a digital product or service, user behavior encompasses all the actions, decisions, and interactions that users make without reference to them. This is intentionally broad, covering all sorts of activities. Online life is characterized by explicit, visible interactions that form the most apparent layer. Included in these activities are creating and publishing content (e.g: posts, videos), direct communication (ex.: messaging, commenting), and building relationships (e.g., friendship creation).

If one were to define these highly explicit behaviors, their definition would be rendered illogically incomplete. The majority of user activity involves hidden or passive behavior, including browsing profiles and videos without commenting or reading posts without responding. Additionally. Research indicates that these silent interactions may constitute more than 90% of all user activity on a platform. Despite being non-reciprocal and difficult to measure, the iceberg of latent behavior provides crucial context about user interests and attention that models based solely on explicit behaviors will overlook.

The "why" must be incorporated in a definition that goes beyond just the "what" of user actions. There are intrinsic drives and requirements that dictate behavior. The motivations that drive users are either to perform a task, fulfill societal needs, or simply to have fun. Accordingly, analyzing user behavior encompasses more than just listing their clickables and postwired connections but also deducing their intentions, preferences, and psychological states. This behavior can take on many forms, including positive social actions like sharing information and building a community, but also more vague behaviors like social investigation (evaluating other people's actions), and overtly malevolent acts such as spamming and harassment.

B. The Social Fabric: Behavior as a Network Phenomenon

Within a social context, user behavior is not isolated to communities; it's embedded in the fabric of society. Social graph is the customary mathematical representation of this situation, with nodes representing users and relationships represented as edges. Social interactions and influence can be analyzed using the structure of this graph. Even so, graphs are not uniformly created, and different models demonstrate distinct behaviors.

The Friendship Graph, which displays an edge to indicate a declared friendship or follower relationship, is frequently the most accessible but can be misleading. Research indicates that a small percentage of users engage with their declared "friends," which may indicate the existence of weak or inactive connections among them. The Interaction Graph offer

an accurate depiction of active social ties, with an edge being formed only when two users have visible interactions, such as through a comment or wall post. This is more accurately represented by the Interaction Fig. 1. These graphs are frequently less comprehensive but more indicative of actual interaction, and have been more effective for applications such as spam detection.

Another area of intricacy is the representation of passive behavior, such as profile browsing, using the Latent Graph. The flow of attention in these directed graphs is often non-reciprocal and can reveal influence patterns that are not visible in interaction graph. Engagement may not be positively correlated with the number of views on a user's profile, as this finding can contradict commonly held beliefs about engagement.

III. Advanced Modeling Paradigms for Forecasting Behavior

The toolkit for foundational predictive analytics is relatively simple; however, the complexity of user behavior requires more sophisticated and powerful modeling paradigms. The frontier of research in this field is marked by the creation of complex neural network architectures that aim to solve specific pieces of the behavioral puzzle. The section outlines three advanced models families: Graph Neural Networks (GNNs), sequence-aware models like Transformers and LLMs, and ABMs. These models are particularly effective in modeling emergent social phenomena from the bottom up. These paradigms are complementary tools that cater to different aspects of user behavior, including relationships, sequences and events. They are not competing with each other. To create a more comprehensive outlook on future events, these methods will likely be combined in varying degrees to generate heightened awareness of social changes.

By learning node embeddings from social connections, GNN's model of user behavior makes it an excellent candidate for churn prediction in relational influence. GCN-LSTM and TempODEGraphNet are two models that use spatial, temporal, and continuous-time learning to make precise predictions. LLMs and transformer capture long range dependencies in user sequence, which enable tasks such as information diffusion modeling and behavior simulation. The role of Agent-Based Models (ABMs) is to simulate social dynamics through individual rule-based agents, enabling researchers to investigate emergent behaviors and test hypothetical scenarios in virtual communities through sociological and psychological archetype analysis.

The combination of these models promises a new era of predictive power beyond their individual advantages. In order to validate hypotheses in controlled simulation environments, hybrid frameworks can integrate temporal and semantic richness from Transformers with relational insights from GNNs. Additionally, they can embed ABMs into the loop. In addition to offering precise projections, these combinations can offer comprehensible insights, which are crucial for policy, marketing, and content moderation decision-making. For instance, GNNs for network structure detection, Transformers for thematic evolution tracking, and ABMs for simulating intervention strategies prior to their implementation in the real world could all be used to identify the early formation of echo chambers.

TABLE I
COMPARISON OF AI MODELING PARADIGMS FOR USER BEHAVIOR FORECASTING

Model Paradigm	Core Principle	Primary Data Type	Key Applications	Strengths	Limitations
Graph Neural Networks (GNNs)	Learns node representations by aggregating information from local network neighborhoods	Graph-structured data (nodes, edges), node/edge features	User Churn Prediction, Link Prediction, Node Classification, Community Detection	Natively handles relational data; captures influence of social structure; state-of-the-art for graph-based tasks	Requires explicit graph structure; performance can degrade in very large, dense, or dynamic graphs; can be computationally intensive
Transformers & LLMs	Uses self-attention to capture long-range dependencies and contextual relationships in sequential data	Sequential data (text, event logs), time-series data	Information Diffusion Prediction, User Engagement Prediction, Fine-grained Behavior Simulation, NLP tasks (Sentiment, NER)	Unparalleled performance on sequence tasks; captures semantic meaning; LLMs possess vast world knowledge and reasoning ability	Computationally expensive to train and run; can act as ”black boxes” lacking explainability; requires massive datasets for pre-training
Agent-Based Models (ABMs)	Simulates macro-level phenomena by defining rules for micro-level agent interactions (”bottom-up”)	Rule-based, theoretical inputs, empirical data for calibration	Opinion Dynamics, Rumor Spreading, Market Simulation, Policy Testing	Excellent for explaining emergent phenomena and testing causal theories; allows for ”what-if” experimentation; highly flexible	Heavily reliant on the quality of the underlying theory; results can be sensitive to initial parameters; validation against real-world data is challenging

[a]Performance metrics based on standard social media datasets

IV. THE END-TO-END PREDICTIVE PIPELINE: FROM RAW DATA TO ACTIONABLE INSIGHTS

Advanced modeling paradigms, as described in the preceding section, serve as powerful engines for predicting behavior with sophistication. Nonetheless, these engines require high-quality fuel and a well-designed chassis. The predictive pipeline encompassing all the steps necessary to transform unstructured data from social communities into structured inputs and rigorously evaluate the outputs of those models. Success hinges on the success of this process, which includes data acquisition and model evaluation, as well as the choice of algorithm. When a sophisticated model is trained on badly available data or tested against inappropriate metrics, the results are always unreliable and misleading. Why? In this section, we delineate the basics of this pipeline: data preparation/feature engineering/performance evaluation.

A. Data Acquisition and Preparation

1) Sourcing High-Quality Data: The previous section's advanced modeling paradigms are the sophisticated engines of behavior forecasting. A well-engineered chassis and high-quality fuel are essential components for these engines to operate. The entire predictive pipeline consists of the essential, practical steps to convert raw, unstructured data from social communities into structured inputs that support machine learning models, and to rigorously evaluate the outputs of those models. The success of the algorithm selection is as significant as the importance of this process, which includes data acquisition and model evaluation. The results of a complex model that is trained on inadequately prepared data or assessed with incorrect metrics are always unreliable and misleading. This section provides a comprehensive overview of this pipeline, covering the necessary stages of data preparation, feature engineering, and performance evaluation.

2) The Critical Role of Preprocessing: Social media's raw data is notorious for being both unreliable and chaotic. The language used to create user-generated content is often informal and cluttered with platform-specific conventions, slang expressions (such as @&@), abbreviations, misspelling symbols, or emoji/handmarks. This raw text is too difficult and ineffective to apply analytical methods directly. The preprocessing process is a fundamental and non-trivial method that helps in cleaning, normalizing, and organizing data into logical machine learning formats. Its main objective is to minimize the complexity and enhance the quality of input data, resulting in faster computation and more accurate models.

A standard preprocessing workflow for social media text involves several key techniques:

- **Lower Casing:** Converting all text to a single case (typically lowercase) to ensure that words like 'Text', 'text', and 'TEXT' are treated as identical, which is crucial for frequency-based feature extraction methods.
- **Removal of Punctuation, URLs, and Special Characters:** Stripping out characters that do not carry semantic meaning for the task at hand helps to standardize the text.
- **Expansion of Contractions and Slang:** Replacing informal shorthand (e.g., "imo" → "in my opinion", "cyaa" → "see you") with their full forms using a predefined dictionary helps to normalize the language.
- **Handling Emojis:** Emojis can be either removed or, more effectively, converted into textual representations (e.g., :-) → Happy_face_smiley) to retain their sentimental information.
- **Tokenization:** Breaking down the cleaned text into individual units, or "tokens" (usually words), which form the

basic building blocks for further analysis.

- **Removal of Stopwords:** Eliminating common, high-frequency words (e.g., 'a', 'the', 'is') that provide little discriminatory value for most NLP tasks.
- **Stemming and Lemmatization:** These are two methods for reducing words to their root form. Stemming is a cruder, rule-based process that chops off suffixes (e.g., 'assisting' → 'assist'), while lemmatization is a more sophisticated, dictionary-based process that reduces a word to its base lemma, considering its part of speech (e.g., 'was' → 'be'). Lemmatization is generally preferred for its linguistic accuracy.

B. *Strategic Feature Engineering*

The process of feature engineering involves extracting and configuring explanatory variables for machine learning models to predict after the data have been cleared. This is a unique domain. The performance of the model is directly related to the quality of these features.

1) Social Network Analysis (SNA) Features: For models that incorporate network structure, a rich set of features can be derived from Social Network Analysis. These metrics quantify a user's position, role, and influence within the social graph. Key hand-crafted features include:

- **Centrality Measures:** These metrics measure the 'importance' of a node. Examples include Degree Centrality (number of direct connections), Betweenness Centrality (frequency of lying on the shortest paths between other nodes), Closeness Centrality (average distance to all other nodes) and Eigenvector centrality or PageRank (influence based on the importance of one's connections).
- **Network-Level Properties:** Metrics like Density (the proportion of actual ties to possible ties), Tie Strength (the weight of a connection), and the presence of structural gaps (gaps between otherwise disconnected parts of the network) can provide valuable context.

2) Automated Feature Engineering with Network Representation Learning (NRL): Manually integrating SNA features can be a laborious process that requires significant domain expertise. A more modern and scalable approach is Network Representation Learning (NRL), also known as network embedding. NRL encompasses a family of techniques, often based on neural networks, that automatically learn a low-dimensional vector representation (embedding) for each node in the network. The goal of this process is to generate embeddings such that the geometric relationships between them in the vector space reflect the structural relationships in the original network. For example, nodes that are close to each other in the network should have similar embedding vectors. These learned embeddings can then be used as powerful and dense features for a wide range of downstream prediction tasks, such as node classification, link prediction, or community detection, often outperforming models based on hand-crafted features.

C. *Evaluating Model Performance*

After a model is trained, its performance must be rigorously evaluated to understand its effectiveness, compare it to other models, and determine its real-world utility. The choice of evaluation metric is not arbitrary; it must be aligned with the specific prediction task and the ultimate business or research objective. A metric that is appropriate for one task may be misleading for another.

1) Metrics for Classification Tasks: In classification problems, where the goal is to assign an instance to a discrete category (e.g., Churn/No-Churn, Malicious/Benign), the following metrics are standard:

- **The Confusion Matrix** is the foundation, providing a breakdown of predictions into True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).
- **Accuracy** $\frac{TP+TN}{\text{Total}}$ is the most intuitive metric but can be highly misleading for imbalanced datasets. For example, in a churn prediction task where only 1% of users churn, a naive model that predicts "no churn" for everyone achieves 99% accuracy but is completely useless for identifying at-risk users.
- **Precision** $\frac{TP}{TP+FP}$ measures the proportion of positive predictions that were actually correct. It answers the question: "Of all the users we predicted would churn, how many actually did?" High precision is critical when the cost of a false positive is high (e.g., unnecessarily giving a discount to a happy customer).
- **Recall** (or Sensitivity) $\frac{TP}{TP+FN}$ measures the proportion of actual positive cases that the model successfully identified. It answers: "Of all the users who actually churned, how many did we catch?" High recall is critical when the cost of a false negative is high (e.g., failing to identify a churning user and losing their business).
- **The F1 Score** is the harmonic mean of Precision and Recall $\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$, providing a single, balanced measure of a model's performance, especially useful when both precision and recall are important.

2) Metrics for Ranking and Engagement: When the goal is to rank items by relevance (e.g., in a recommendation system) or measure user engagement, different metrics are required:

- **Mean Reciprocal Rank (MRR):** Measures the quality of a ranked list by looking at the rank of the first correct answer. It is simple and useful when the user is only interested in finding one relevant item quickly.
- **Normalized Discounted Cumulative Gain (NDCG):** A more sophisticated ranking metric that evaluates the entire ranked list. It assigns higher scores to more relevant items and gives more weight to items ranked higher on the list, making it ideal for evaluating the overall quality of a recommendation feed.

TABLE II
LITERATURE REVIEW

Reference	Primary Focus / Contribution	AI Technique(s) Used	Limitations / Research Gap	How My Research Differs / Extends Their Work
Dou et al. (2024)	Proposed "Social-LLM" to model user behavior by integrating social network information into a Large Language Model for predicting future user posts	Large Language Model (LLM) fine-tuned with social graph data	Focuses primarily on content generation behavior. Does not provide a broader framework for other behaviors like churn or engagement, nor does it detail a full methodological pipeline	My research integrates the concept of Social-LLMs as one component of a larger, multifaceted framework. It situates LLMs alongside GNNs and RNNs to address a wider range of behaviors (e.g., churn, engagement) and provides a complete methodological guide
Li et al. (2023)	Developed "TempODEGraphNet," a novel model for user churn prediction by capturing continuous-time dynamics in social graphs	Graph Neural Network (GNN), Ordinary Differential Equations (ODE)	Highly specialized for the single task of churn prediction. The complex model may lack interpretability and is not designed for other behavioral forecasting tasks	My paper positions specialized models like TempODEGraphNet as an advanced option within a broader methodological framework. It compares this approach to others (like RNNs or Transformers) and discusses the trade-offs, providing guidance on model selection for different problems beyond just churn
Cheng et al. (2014)	Investigated the predictability of information cascades by engineering features based on cascade structure and content	Feature-based Machine Learning (e.g., Logistic Regression)	Relies heavily on hand-crafted features. Pre-dates the widespread use of deep learning models that can learn feature representations automatically and more effectively	My research incorporates modern deep learning approaches like Transformers (e.g., ACET) for information diffusion, which can capture complex temporal and structural patterns automatically, moving beyond the limitations of manual feature engineering
Kipf & Welling (2017)	Introduced Graph Convolutional Networks (GCNs), a foundational deep learning architecture for processing graph-structured data	Graph Neural Network (GCN)	Foundational work focused on the model architecture itself (semi-supervised classification). Does not explore application to complex, dynamic user behaviors like temporal sequences or content analysis	My research applies foundational techniques like GCNs within a practical, real-world context. It shows how to combine GNNs with other models (like RNNs) to create hybrid systems for dynamic user behavior analysis, extending the original scope of the GCN
Macy & Willer (2002)	Advocated for Agent-Based Models (ABMs) as a "bottom-up" approach to understand emergent social phenomena from individual agent rules	Agent-Based Modeling (Simulation)	ABMs are powerful for theoretical exploration ("what-if" scenarios) but can be difficult to calibrate with real-world, large-scale data and are less suited for point-prediction forecasting tasks	My framework presents ABMs as a complementary tool for understanding causal mechanisms, alongside predictive AI models that excel at forecasting. It clarifies the distinct roles of simulation (ABMs) and prediction (ML models) in a comprehensive user behavior analysis toolkit

V. FUTURE RESEARCH HORIZONS

The field of AI-driven behavior prediction is evolving rapidly, and several key research frontiers promise to address current limitations and unlock new capabilities.

- **Hybrid and Multi-Paradigm Models:** The intelligent blending of diverse modeling paradigms is expected to mark the future. Combining the knowledge of a GNN with an LLM's ability to reason and understand social structure in such terms within nimbus, could result in constructing something more comprehensive and powerful tool for both prediction and explanation.
- **Multimodal Data Integration:** Human communication is inherently multimodal. The integration of information from images, videos, and audio is a crucial aspect that future models must prioritize over text and network data. Developing architectures capable of learning a cohesive representation from various data sets poses significant challenges, but it is crucial for understanding user behavior.
- **Causal Inference:** The majority of present predictive models rely on correlation rather than causation. Their proficiency in distinguishing between the two happenings

is exceptional, yet their ability to ascertain if one is responsible for the other fails. One of the primary areas of research is the development of causal inference methods from observational data, which would enable models to predict the outcomes of interventions and move beyond hypothetical questions like ”what will happen?” to ”how will this affect X?”.

VI. Conclusion

The ability to predict and understand user behavior in online social communities is a fundamental aspect of the modern digital landscape. The article documents the progression from defining user behavior with an initial, complex definition to investigating the advanced Artificial Intelligence models designed to forecast it. The complexity of behavior is not merely determined by clicks, but rather by individual psychological drivers and dynamic social network structures. A diverse AI toolkit has facilitated the field's transition from descriptive analytics to a proactive, predictive posture in order to analyze this complexity.

Advanced modeling paradigms are pushing this field towards its frontier, providing unique perspectives on the behavioral puzzle. Graph Neural Networks enable the understanding of the potent influence of relational dynamics and social structure.' Large Language Models and Transformer-based models have provided users with an unparalleled ability to process temporal sequences as well as semantic details of user-generated content, even reaching into the realm of generative behavior simulation. By using Agent-Based Models, one can understand how complex, emergent social phenomena can arise from simple, individual-level rules. This is a further explanation for these predictors. Nonetheless, the effective implementation of these models demands a rigorous end-to–end process, including meticulous data preparation, strategic feature engineering and the use of appropriate, task-aligned evaluation metrics.

But the crux of this analysis is that this remarkable technological breakthrough was not to be taken for granted alone. The power of prediction extends to profound ethical responsibility The process of modeling human behavior at this level presents a challenge to crucial concerns such as algorithmic discrimination, data protection, and accountability. Our models aren't just observers; they're active participants in the digital world, able to amplify societal biases, diminish individual autonomy, and influence public discourse in ways that can be both harmful and beneficial.

Thus, the ultimate challenge for this industry is not purely technical but fundamentally human. Why? It's the challenge of wielding. These potent predictive tools are not merely intended to increase engagement and generate profits, but also to possess the wisdom, equity, and individual dignity necessary for building healthy, prosperous, equitable, digital communities. Ultimately, the future of AI in social analysis will be determined by our ability to balance predictive power with ethical principles.

Acknowledgment

We would like to express our sincere thanks of gratitude to our University's professor and our supervisor Mr. Sukhpreet Singh for their valuable guidance and support throughout the completion of our review.

References

[1] C. C. Aggarwal, *Social Network Data Analytics*. Springer, 2011.
[2] Z. Dou, K. Zhou, and J. He, ”Social-LLM: Understanding user behavior on social platforms through large language models,” *arXiv preprint*, arXiv:2401.00893, 2024.
[3] S. Kumar, R. West, and J. Leskovec, ”Disinformation in online environments: A comprehensive review,” *Foundations and Trends in Web Science*, vol. 8, no. 3–4, pp. 209–363, 2016.
[4] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*. Cambridge University Press, 2014.
[5] L. Li, D. Wang, Z. Zhang, and Z. Jin, ”TempODEGraphNet: Graph neural network with temporal ODE modeling for churn prediction,” *PLOS ONE*, vol. 18, no. 4, p. e0284851, 2023.
[6] D. Lazer *et al.*, ”Life in the network era: Computational social science,” *Science*, vol. 323, no. 5915, pp. 721–723, 2009.
[7] T. N. Kipf and M. Welling, ”Semi-supervised learning with graph convolutional architectures,” in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2017.
[8] W. L. Hamilton, R. Ying, and J. Leskovec, ”Inductive representation learning for large-scale graphs,” in *Proc. 31st Conf. Neural Information Processing Systems (NeurIPS)*, 2017.
[9] A. Vaswani *et al.*, ”The transformer: Attention mechanisms for sequence modeling,” in *Proc. 31st Conf. Neural Information Processing Systems (NeurIPS)*, 2017.
[10] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, ”GroupLens: A collaborative filtering framework for netnews,” in *Proc. ACM Conf. Computer-Supported Cooperative Work (CSCW)*, 1994, pp. 175–186.
[11] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*. Cambridge University Press, 2014.
[12] M. W. Macy and R. Willer, ”From statistical factors to agent actors: Computational sociology via agent-based simulation,” *Annual Review of Sociology*, vol. 28, no. 1, pp. 143–166, 2002.
[13] X. Wang, C. He, H. Liu, and Y. Wang, ”Fake news detection using deep learning techniques: A survey,” *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–36, 2022.
[14] C. O'Connor, *The Misinformation Age: How False Beliefs Spread*. Yale University Press, 2019.
[15] S. Aral and D. Walker, ”Detecting influential and vulnerable individuals in social networks,” *Science*, vol. 337, no. 6092, pp. 337–341, 2012.
[16] J. Cheng, L. Adamic, J. Dow, J. Kleinberg, and J. Leskovec, ”Predicting information cascades,” in *Proc. 23rd Int. World Wide Web Conf. (WWW)*, 2014, pp. 925–936.
[17] S. Vosoughi, D. Roy, and S. Aral, ”The diffusion of true and false news in online media,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
[18] S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[19] N. Gilbert, *Agent-Based Models*. Sage Publications, 2008.
[20] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, ”Information spread in online social networks,” in *Proc. 21st Int. World Wide Web Conf. (WWW)*, 2012, pp. 519–528.