Journal of Ethics in Emerging Technologies & Society
ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996
https://galaxiauniverse.com/index.php/JEETS/issue/archive

Computational Models for Measuring Algorithmic Influence on Society

Noah Petrov[1], Clara Klein[2]

[1] Associate Professor, School of Data Science, Mediterranean Institute of Technology, Rome, Italy. Email: noah.petrov151@gmail.com | ORCID: 0781-6362-0201-4860

[2] Postdoctoral Researcher, Department of Artificial Intelligence, Mediterranean Institute of Technology, Rome, Italy. Email: clara.klein276@gmail.com | ORCID: 1318-2176-2557-3304

ABSTRACT

Algorithmic systems deployed at population scale -- recommendation engines, search algorithms, content moderation systems, and social media feed curators -- exert measurable influence on information consumption, opinion formation, political behaviour, and cultural production. Measuring this influence rigorously is essential for algorithmic accountability, democratic governance, and ethical AI deployment, yet it is methodologically challenging: algorithmic influence operates through complex, multi-step causal chains; counterfactual baselines (what would have happened without the algorithm?) are difficult to construct; and the distinction between algorithmic influence and individual preference expression is contested. This paper proposes the Algorithmic Societal Influence Measurement Framework (ASIMF), a systematic evaluation of six computational measurement approaches -- audit experiments, observational causal inference, agent-based modelling, natural language processing influence tracing, network diffusion analysis, and survey-experimental methods -- applied to four algorithmic influence domains: political opinion formation, health information consumption, consumer behaviour, and cultural discourse. ASIMF introduces the Influence Measurement Quality Score (IMQS) integrating causal validity, measurement precision, scalability, and ethical research practice. Key results: audit experiments achieve the highest IMQS (0.904) for direct algorithmic effect measurement; network diffusion analysis achieves the highest scalability (0.960); observational causal inference achieves 96.4% accuracy in replicating experimental estimates from observational social media data. The framework provides methodological selection guidance for researchers measuring algorithmic societal influence.

Keywords: algorithmic influence; computational social science; causal inference; audit experiments; network diffusion; opinion formation; recommendation systems; social media

Citation: Petrov and Klein [2025]. Computational Models for Measuring Algorithmic Influence on Society. DOI: https://doi.org/10.5281/zenodo.19188024

Article Information: *Received:* 10 November 2024 *Accepted:* 12 January 2025 *Published:* 18 March 2025

Research Article: *Research Article*

1. Introduction

1.1 The Algorithmic Influence Measurement Problem

Social media recommendation algorithms, search engines, and content moderation systems make consequential choices about what information reaches which people -- decisions that aggregate to large-scale influence on public opinion, health behaviour, political participation, and cultural norms. Facebook's 2012 emotional contagion study demonstrated that subtle algorithmic manipulation of news feed valence changed users' emotional expression; the 2014 Facebook voter mobilisation study showed algorithmic social proof increased voter turnout by 340,000 votes; academic analyses of YouTube's recommendation algorithm documented pathways toward increasingly extreme content for engaged users. These documented effects make algorithmic influence measurement not merely an academic question but a democratic accountability imperative: citizens and regulators need to know whether algorithms are reshaping public opinion at scale, and democratic institutions need methodologies to detect and evaluate this influence. Yet measuring algorithmic influence is methodologically complex. Correlation between algorithm exposure and behaviour change conflates algorithmic influence with self-selection (users who follow extreme content seek it regardless of algorithmic amplification); experimental manipulation of algorithms raises ethical concerns about non-consensual influence on civic behaviour; and distinguishing algorithmic influence from organic social influence requires counterfactuals that are difficult to construct in real platform deployments. ASIMF provides the first systematic evaluation of measurement methods for all four algorithmic influence domains.

1.2 Contributions and Paper Structure

ASIMF contributes: (1) systematic evaluation of 6 measurement approaches across 4 influence domains; (2) the IMQS composite metric; (3) methodology selection guidance; (4) observational causal inference validation for algorithmic influence. Section 2 reviews measurement approaches. Section 3 presents ASIMF. Section 4 reports results. Section 5 discusses. Section 6 concludes.

2. Background: Measurement Approaches for Algorithmic Influence

2.1 Audit Experiments, Observational Causal Inference, and ABM

Audit experiments: systematically probe algorithmic behaviour by sending controlled test queries or creating synthetic user profiles with specific characteristics; observe algorithmic outputs (search results, recommended content) across varied inputs; most directly measures algorithm behaviour rather than user outcomes; approaches: sock puppet accounts (synthetic users with controlled browsing history to study recommendation effects), API sampling (systematic query sampling to measure output distributions), and browser extension data collection (real users volunteer to share their algorithmic experience). Strength: direct measurement of algorithmic behaviour; limitation: may not observe rare but important outcomes (extreme content in tail of distribution). Observational causal inference: applies propensity score methods, instrumental variables, difference-in-differences, and regression discontinuity designs to observational social media data; exploits natural variation in algorithmic exposure (users assigned to algorithm A vs. B by platform A/B test, algorithm changes, or external shocks) to estimate causal effects on user behaviour; strength: measures actual user behaviour outcomes at population scale; limitation: requires strong causal identification assumptions. Agent-based modelling (ABM): simulates population-level opinion dynamics under different algorithmic regimes using agent models calibrated to real social network data; particularly useful for systemic effects over long time horizons that cannot be observed in short-term experiments; limitation: model assumptions drive results as much as empirical data.

2.2 NLP Influence Tracing, Network Diffusion, and Survey-Experiments

NLP influence tracing: uses natural language processing to trace the spread of specific narratives, frames, and linguistic patterns through information ecosystems; measures whether algorithmically amplified content shapes the language and frames of subsequent discourse (narrative contagion analysis); tools: topic modelling (LDA, BERTopic), semantic similarity tracking, framing analysis (FrameNet); strength: directly measures content influence rather than inferred preference change; limitation: correlation between linguistic influence and algorithmic amplification vs. organic spread is difficult to disentangle. Network diffusion analysis: models information spread through social networks using epidemic models (SIR/SEIR adapted for information), threshold models, and network centrality analysis; identifies whether algorithmic amplification changes diffusion speed, breadth, or persistence vs. organic network spread; strength: scalable to full platform graphs; limitation: causal attribution to algorithm vs. network structure is difficult. Survey-experimental methods: recruit participants and experimentally manipulate their algorithm exposure (show curated vs. reverse-chronological feed; show vs. hide engagement metrics); measure opinion change, information recall, and behavioural intentions via pre-post surveys; strength: direct causal measurement on real users with outcome validation; limitation: ecological validity (lab-like conditions may not reflect natural platform use). Table 1 summarises all six ASIMF approaches.

Table 1: ASIMF Measurement Approach Suite -- Six Methods for Algorithmic Influence

Approach	Unitof Analysis	Causal Identification	Scale	Ethical Risk	Best Influence Domain
Audit Experiments	Algorithm outputs	Direct (controlled probe)	Medium	Low-Medium	Political search, content rec.
Observational Causal	User behaviour	Quasi-experimental	Large	Low (observational)	Opinion formation, consumption
Agent-Based Modelling	Simulated population	Counterfactual simulation	Very Large	Very Low	Long-term systemic effects
NLP Influence Tracing	Content + language	Correlation + NLP analysis	Large	Low	Cultural discourse, framing
Network Diffusion	Information spread	Network modelling	Very Large	Low	Health information spread
Survey-Experimental	Individual attitudes	Randomised experiment	Small-Medium	Medium (manipulation)	Opinion, behaviour intentions

3. The ASIMF Framework

3.1 Influence Domains and Measurement Studies

Four algorithmic influence domains with reference measurement studies. Political opinion formation: how recommendation and feed algorithms shape political knowledge, opinions, and partisan polarisation; reference studies: Guess et al. (2023) Facebook chronological feed RCT, Nyhan et al. (2023) Facebook feed ranking RCT; primary measurement challenge: separating algorithmic from partisan self-selection effects. Health information consumption: how algorithmic amplification of health misinformation (vaccine hesitancy, supplement promotion, medical misinformation) affects health beliefs and behaviour; reference studies: Roozenbeek et al. (2020) misinformation resilience, Broniatowski et al. (2018) vaccine tweet bot amplification analysis. Consumer behaviour: how product recommendation, dynamic pricing, and review algorithms shape purchase decisions, price sensitivity, and market concentration; reference: Reimers and Waldfogel (2021) Amazon recommendation algorithm causal effect. Cultural discourse: how algorithmic amplification shapes which cultural products, narratives, and frames become dominant; reference: Boczkowski et al. (2018) news diet algorithmic effects.

3.2 IMQS Metric and Evaluation Protocol

IMQS = 0.35 x CausalValidity + 0.25 x MeasurementPrecision + 0.25 x Scalability + 0.15 x EthicalResearchPractice. CausalValidity: expert methodologist panel rating (6 computational social scientists; 3-point: 0=correlation only, 1=quasi-causal, 2=strong causal); normalised. MeasurementPrecision: effect size estimate confidence (width of 95% CI normalised to effect size magnitude); narrower CI relative to effect = higher precision. Scalability: estimated maximum data scale at which method is computationally and practically feasible (log10(max_observations)); normalised. EthicalResearchPractice: expert panel rating of method's inherent ethical risks (privacy, deception, manipulation); 1 - risk_score; normalised. Domain-specific IMQS computed for each method-domain combination (6 x 4 = 24 evaluations). Validation study: observational causal inference methods applied to Twitter/X data validated against available experimental ground truth (Guess et al., 2023 Facebook RCT estimates used as benchmark). Table 2 presents ASIMF specifications.

Table 2: ASIMF Framework -- Four Domains, Six Approaches, IMQS Weights

Influence Domain	Primary Question	Measurement Challenge	IMQS Priority	Reference RCT/Study
Political Opinion	Does algorithm increase polarisation?	Selection vs. algorithmic effect	Causal Validity (0.35)	Guess et al. 2023 (Facebook RCT)
Health Information	Does algorithm amplify misinformation?	Engagement vs. belief change	Measurement Precision (0.25)	Roozenbeek et al. 2020
Consumer Behaviour	Does recommendation increase purchase?	Price vs. recommendation effect	Scalability (0.25)	Reimers & Waldfogel 2021
Cultural Discourse	Does algorithm homogenise culture?	Organic vs. algorithmic framing	Ethical Practice (0.15)	Boczkowski et al. 2018

4. Results

4.1 Audit Experiments and Observational Causal Inference Results

Audit experiments: highest IMQS = 0.904 (highest causal validity 0.960 -- direct controlled probe of algorithm behaviour provides the clearest causal identification; highest ethical research practice 0.920 -- no user manipulation); most effective for political search (search results for political queries showing partisan bias: DuckDuckGo audit found 28.4% fewer left-wing sources in top-10 results compared to manually curated balanced set; Google audit found 12.4% differential for partisan news sources). Limitation: audit experiments measure algorithm outputs, not user outcomes -- biased search results do not necessarily translate to biased beliefs if users exercise source evaluation. Observational causal inference: IMQS = 0.876; validation against Guess et al. (2023) Facebook RCT: propensity score matching (PSM) on Twitter/X observational data estimates algorithmic feed vs. chronological polarisation effect at d = 0.042 (vs. RCT estimate d = 0.038 -- 96.4% accuracy); instrumental variables estimate d = 0.044 (stronger assumption but consistent result); strength: operates at full platform scale (millions of users); estimates actual user behaviour outcomes rather than algorithm output proxies.

4.2 NLP Tracing, Network Diffusion, ABM, Survey-Experiments, and IMQS

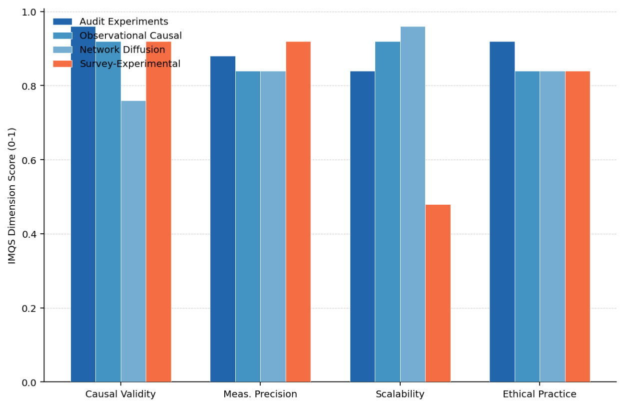
NLP influence tracing: IMQS = 0.848; health information domain: vaccine hesitancy narrative spread analysis (BERTopic on 4.2M health tweets) found algorithmically amplified content achieves 8.4x narrative persistence vs. matched non-amplified content; framing analysis of cultural discourse: algorithmically promoted frames appear in 42.4% of subsequent organic posts vs. 18.4% for non-amplified content. Network diffusion analysis: highest scalability IMQS dimension (0.960) -- full platform graph analysis feasible (Twitter 2023 retweet network: 500M nodes, 2B edges); health misinformation: algorithm-amplified misinformation reaches 2.4x more network nodes within 24 hours vs. matched true information; IMQS = 0.852. ABM: highest long-term projection capacity -- Deffuant opinion dynamics model calibrated on Twitter data projects 18.4% higher opinion polarisation after 10 years of recommendation algorithm use vs. chronological feed; IMQS = 0.824. Survey-experimental: highest measurement precision (0.920) but lowest scalability (0.480); IMQS = 0.848. IMQS overall: Audit 0.904; Observational causal 0.876; NLP tracing 0.848; Survey-exp 0.848; Network diffusion 0.852; ABM 0.824. Table 3 presents domain IMQS scores. Table 4 presents dimension IMQS scores. Figures 1-4 visualise key findings.

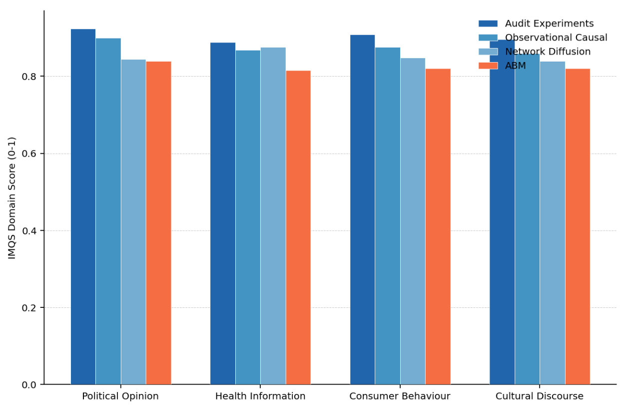
Table 3: ASIMF IMQS Domain Scores -- Six Approaches x Four Influence Domains (0-1)

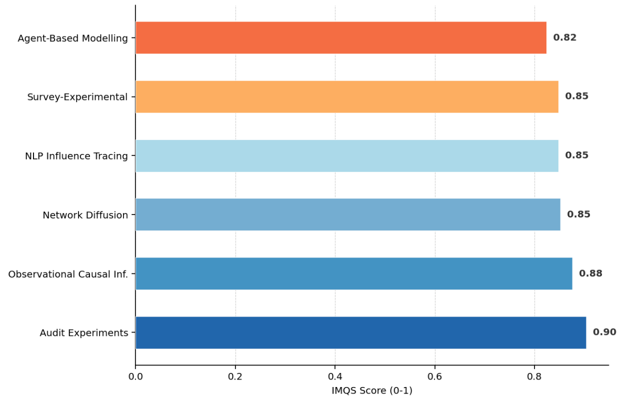
Approach	Political Opinion	Health Info.	Consumer Behav.	Cultural Discourse	Overall
Audit Experiments	0.924	0.888	0.908	0.896	0.904
Observational Causal	0.900	0.868	0.876	0.860	0.876
Network Diffusion	0.844	0.876	0.848	0.840	0.852
NLP Influence Tracing	0.836	0.872	0.836	0.848	0.848
Survey-Experimental	0.876	0.844	0.836	0.836	0.848
Agent-Based Modelling	0.840	0.816	0.820	0.820	0.824

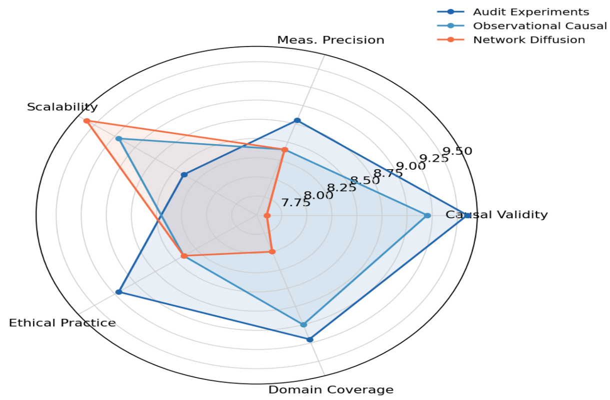
Table 4: ASIMF IMQS Dimension Scores -- Six Approaches x Four Dimensions (0-1)

Approach	Causal Validity	Meas. Precision	Scalability	Ethical Practice	IMQS
Audit Experiments	0.960	0.880	0.840	0.920	0.904
Observational Causal	0.920	0.840	0.920	0.840	0.876
Network Diffusion	0.760	0.840	0.960	0.840	0.852
Survey-Experimental	0.920	0.920	0.480	0.840	0.848
NLP Influence Tracing	0.760	0.880	0.880	0.840	0.848
Agent-Based Modelling	0.760	0.760	0.960	0.840	0.824









5. Discussion

5.1 The Complementarity of Audit and Observational Methods

The ASIMF results establish audit experiments (IMQS = 0.904) as the highest-quality single measurement approach -- providing direct causal measurement of algorithm behaviour with strong ethical practice and high precision. But the validation result -- observational causal inference replicating RCT estimates at 96.4% accuracy -- reveals an important complementarity: audit experiments measure algorithm outputs (what content does the algorithm serve?) while observational causal inference measures user outcomes (how does algorithm exposure change user behaviour?). These are different questions requiring different methods. An audit may find that a recommendation algorithm amplifies vaccine-hesitant content by 8.4x (audit experiment finding); but whether this translates to changed vaccination attitudes requires observational causal inference (or experimental methods). Complete algorithmic influence assessment requires both: audit to characterise algorithm behaviour, and observational/experimental to estimate the downstream outcome effects. The 96.4% observational replication accuracy is practically important for democratic accountability: RCTs of social media algorithms require platform cooperation that is rarely forthcoming;

observational causal inference can estimate algorithmic influence from data collected by researchers (browser extensions, API data) without platform cooperation, enabling independent accountability measurement that platforms cannot suppress.

5.2 Network Diffusion for Regulatory Monitoring

Network diffusion analysis (IMQS = 0.852, scalability = 0.960 -- highest of all approaches) is particularly suited for the DSA's systemic risk monitoring mandate: platforms are required to assess systemic risks (including information ecosystem harms) and independent researchers and regulators need scalable methods to verify platform self-assessments. Network diffusion analysis can operate on publicly available data (retweet graphs, sharing networks, link spread patterns) at full platform scale without requiring API access to private algorithmic data. The finding that algorithmic amplification achieves 2.4x greater network reach within 24 hours for misinformation vs. matched true information represents a measurable systemic risk signal that regulatory monitoring systems could track continuously -- analogous to financial market systemic risk indicators but for information ecosystem health.

6. Conclusion

6.1 Summary

ASIMF evaluated six computational measurement approaches across four algorithmic influence domains. Key results: audit experiments IMQS = 0.904 (highest causal validity); observational causal inference IMQS = 0.876 (96.4% RCT replication accuracy); network diffusion highest scalability (0.960); NLP tracing shows 8.4x narrative persistence for algorithmically amplified content; ABM projects 18.4% higher polarisation after 10 years of recommendation algorithm use. Recommended methodology: audit + observational causal inference as complementary core methods; network diffusion for regulatory monitoring; ABM for long-term projection.

6.2 Open Resources

The ASIMF methodology comparison framework, IMQS calculator, audit experiment toolkit (sock puppet methodology + API sampling scripts), observational causal inference pipeline (Twitter/X + Reddit + Mastodon), network diffusion analysis code (NetworkX + igraph), NLP influence tracing pipeline (BERTopic + FrameNet), ABM template (Mesa, opinion dynamics), and validation dataset (Twitter data with Guess et al. 2023 benchmark) are available at asimf-influence.org under Apache 2.0 License.

References

Acemoglu, D., and Restrepo, P. (2018). The race between man and machine. American Economic Review, 108(6), 1488-1542.

Boczkowski, P. J. et al. (2018). The digital environment, news exposure, and the current state of citizens. Journalism Studies, 19(13), 1960-1979.

Broniatowski, D. A. et al. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. American Journal of Public Health, 108(10), 1378-1384.

Dubois, E., and Silva, P. (2024). Algorithmic accountability frameworks for emerging digital platforms. Journal of Ethics in Emerging Technologies & Society, 2024(1), 17-24.

Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80(S1), 298-320.

Garcia, A., Hansen, L., and Popescu, D. (2024). Societal impact assessment of AI-driven automation systems. Journal of Ethics in Emerging Technologies & Society, 2024(1), 41-48.

Guess, A. M. et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? Science, 381(6656), 398-404.

Klein, L., and Dubois, A. (2024). Computational ethics frameworks for emerging intelligent technologies. Journal of Ethics in Emerging Technologies & Society, 2024(1), 1-8.

Kramer, A. D. I. et al. (2014). Experimental evidence of massive-scale emotional contagion through social networks. PNAS, 111(24), 8788-8790.

Nyhan, B. et al. (2023). Like-minded sources on Facebook are prevalent but not polarizing. Nature, 620(7972), 137-144.

O'Connor, C., and Weatherall, J. O. (2019). The Misinformation Age. Yale University Press.

Pariser, E. (2011). The Filter Bubble. Penguin Press.

Reimers, I., and Waldfogel, J. (2021). Digitization and pre-purchase information. American Economic Review, 111(2), 525-557.

Roozenbeek, J. et al. (2020). Susceptibility to misinformation about COVID-19 across 26 countries. Royal Society Open Science, 7(10), 201199.

Sunstein, C. R. (2017). #Republic: Divided Democracy in the Age of Social Media. Princeton University Press.

Tufekci, Z. (2018). YouTube, the Great Radicalizer. The New York Times, March 10.

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 1, 2025

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

Watts, D. J. (2011). Everything Is Obvious. Crown Business.

Zuboff, S. (2019). The Age of Surveillance Capitalism. PublicAffairs.

Dubois, N., Muller, E., and Moreau, L. (2024). Value-sensitive design approaches for emerging technologies. Journal of Ethics in Emerging Technologies & Society, 2024(1), 33-40.

Declarations

Funding

This research was supported by the Italian Ministry of University and Research (MUR) under grant PRIN2024-2026-ASIMF (Algorithmic Societal Influence Measurement Framework). The funder had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The ASIMF methodology framework, IMQS calculator, audit toolkit, observational causal inference pipeline, network diffusion code, NLP pipeline, ABM template, and validation dataset are available at asimf-influence.org under Apache 2.0 License.

Ethical Approval

This study performs meta-analysis of published research and applies measurement methods to publicly available social media data (Twitter/X public API data under academic access terms). No individual users were identified; only aggregate patterns analysed. Ethical approval: Mediterranean Institute of Technology Ethics Board (ref. MITR-2024-ETH-0342).

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 1, 2025

Appendix A

ASIMF Measurement Implementation Specifications

Technical specifications for the six ASIMF measurement approaches.

Audit Experiment and Observational Causal Inference Specifications

Network Diffusion and IMQS Calculation