Algorithmic Personality Inference: How AI Recommendation Systems Construct and Deploy Psychological Profiles of Media Audiences

[1]Akhil
Working Professional

[2]Aarzoo
Research and Journal Manager, We Avec U Centre for Research & Innovations

[3]Dr. Sundeep Katevarapu
Founder and Chief Managing Director at We Avec U® Mental Health Organization, Founder at WeAvecU@ Pvt Ltd, Founder President at We Avec UR Trust, Founder Director at We Avec U Organization LLC (USA), Director, We Avec U Limited (UK)

Abstract

The deployment of artificial intelligence in media recommendation systems has enabled the construction of granular psychological profiles of users based on behavioral trace data, content interaction patterns, and passive digital signals. This paper provides a comprehensive theoretical and empirical analysis of how AI-driven recommendation systems infer, represent, and operationalize psychological constructs — including personality traits, emotional states, cognitive styles, and motivational orientations — to optimize content delivery and behavioral prediction. Drawing on the Five-Factor Model of personality, the OCEAN-social media inference literature, and algorithmic profiling theory, the paper examines the psychometric validity, ethical implications, and democratic consequences of AI-mediated psychological profiling at scale. The Cambridge Analytica case is analyzed as a paradigmatic instance of psychographic targeting that collapsed the boundary between psychological research and political manipulation. The paper reviews the empirical evidence for personality inference from digital footprints including Facebook likes (N =

58,000; r = .56 for openness), social media language predictions, and behavioral clickstream data, and critically evaluates the gap between prediction accuracy and psychological construct validity. A taxonomy of profiling mechanisms is proposed distinguishing explicit psychometric input, behavioral inference, network-structural inference, and temporal pattern extraction. The paper argues that AI psychological profiling constitutes a form of epistemic asymmetry where platforms know more about users psychological states than users themselves, with profound implications for autonomy, consent, and cognitive liberty. Regulatory frameworks including the EU AI Act, GDPR Article 22, and India's Digital Personal Data Protection Act are evaluated for their adequacy in addressing algorithmic psychological profiling.

Keywords: AI profiling; recommendation systems; personality inference; OCEAN model; psychographic targeting; algorithmic profiling; digital footprints; cognitive liberty

---

1. Introduction

The intersection of artificial intelligence, psychological science, and media systems has produced one of the most consequential and least scrutinized developments in contemporary information environments: the large-scale algorithmic inference of psychological characteristics from digital behavior. Every interaction a user has with a digital media platform generates behavioral trace data that AI systems increasingly use to construct implicit psychological profiles (Aarzoo & Lal, 2024). These profiles drive content recommendation, advertising targeting, political messaging, and platform design decisions that shape what hundreds of millions of people see, read, and believe.

The psychological science underlying this profiling ranges from the rigorous to the spurious. At the rigorous end, research by Kosinski, Stillwell, and Graepel (2013) demonstrated that Five-Factor Model personality traits could be predicted from Facebook likes with meaningful accuracy, establishing an empirical foundation for personality inference from passive digital signals. At the spurious end, commercial psychographic platforms routinely claim predictive capabilities that exceed published validation evidence by orders of magnitude.

The practical consequences of this profiling extend from the mundane to the politically consequential, as illustrated by the Cambridge Analytica scandal in which psychographic profiles of millions of Facebook users were used to deliver micro-targeted political messaging during the 2016 U.S. presidential election and Brexit referendum. Beyond the electoral context, AI psychological profiling affects the information diets of global news audiences, the visibility of health information during public crises, and the economic pathways of content creators whose work is amplified or suppressed by algorithms with implicit psychological models of their audiences (Aarzoo & Lal, 2025a).

This paper provides a comprehensive analysis of AI-driven psychological profiling in media recommendation systems, addressing four interconnected questions: What psychological constructs do media AI systems actually measure or infer? How psychometrically valid are these inferences? What are the individual, democratic, and epistemic consequences of profiling at scale? And what regulatory and technical frameworks can adequately govern this practice?

2. Literature Review

The scientific foundation for personality inference from digital behavior rests on a well-established body of research demonstrating systematic relationships between Big Five personality traits and patterns of online self-expression, social network characteristics, and content preferences.

Kosinski, Stillwell, and Graepel (2013) established that automated analysis of Facebook likes predicted personality with meaningful statistical accuracy: Openness to experience at $r = .56$, conscientiousness at $r = .44$, extraversion at $r = .40$, agreeableness at $r = .35$, and neuroticism at $r = .31$. The study demonstrated that 10 Facebook likes allowed personality prediction at the accuracy level of an acquaintance; 70 likes approached the accuracy of a friend; 300 likes matched the accuracy of a spouse (Aarzoo & Lal, 2025b). While these effect sizes were notable, correlation coefficients in the .35 to .56 range explain 12 to 31 percent of variance in personality scores, leaving 69 to 88 percent unexplained.

Youyou, Kosinski, and Stillwell (2015) found that a computer model using Facebook likes surpassed average human judge accuracy in personality prediction, receiving widespread attention often reported without the critical observation that both the AI model and human judges explained only a fraction of personality variance.

Language-based personality inference represents a complementary approach. Mairesse et al. (2007) demonstrated that LIWC-based features predicted all Big Five traits from written text samples. Schwartz et al. (2013) open-vocabulary Facebook language analysis identified lexical markers of personality, gender, age, and mental health status, enabling granular psychological inference from social media text at scale.

Beyond personality, AI profiling systems extend to emotional state detection, political orientation inference, and mental health prediction. De Choudhury et al. (2013) used machine learning on Twitter data to predict post-partum depression with sensitivity and specificity exceeding 70 percent, establishing algorithmic mental health prediction from passive social media data (Aarzoo & Lal, 2026). These findings illustrate the dual-use potential: the same capabilities enabling public health surveillance can enable commercial exploitation of mental vulnerability.

The Cambridge Analytica case provided the starkest illustration of psychographic profiling deployed for political manipulation, drawing on personality profiles derived from the myPersonality data and supplemented by Facebook engagement data from approximately 87 million users. While independent assessments of the campaign's effectiveness have been mixed, the case demonstrated that psychographic micro-targeting at electoral scale was technically feasible and institutionally poorly governed (Cadwalladr and Graham-Harrison, 2018).

Media recommendation systems employ psychological profiling through mechanisms less visible than explicit psychometric surveys but potentially more pervasive in their effects. Netflix's recommendation algorithm implicitly models viewer preferences in high-dimensional latent spaces capturing psychologically interpretable dimensions. YouTube's watch-time optimization algorithm has been shown to model anxiety-susceptibility and outrage reactivity by discovering that content triggering these states maximizes watch time, a finding with profound implications for radicalization dynamics.

3. Theoretical Framework

The Algorithmic Psychological Profiling Framework (APPF) proposed here integrates four theoretical traditions to provide a comprehensive account of how AI systems construct, validate, and deploy psychological profiles of media users.

The first pillar is psychometric theory in the tradition of the Five-Factor Model. Valid psychological profiling requires that inferred constructs correspond to established psychological constructs with validated measurement properties, demonstrate convergent validity against ground-truth psychometric measures, demonstrate discriminant validity from related but distinct constructs, and demonstrate predictive validity for theoretically relevant behavioral outcomes. Most commercial AI profiling systems are developed and deployed without any of these validity checks, constituting what the paper terms psychometric laundering: the appropriation of psychological terminology to legitimize algorithmic categorization lacking scientific validity.

The second pillar is information asymmetry theory from economics and organizational behavior. Applied to psychological profiling, the information asymmetry is between platforms which hold detailed behavioral data enabling psychological inference, and users who do not know what psychological conclusions platforms have drawn about them, what content manipulations those conclusions have justified, or what third parties have received profiling data. This asymmetry constitutes a form of epistemic coercion when profiling is used to deliver content that exploits psychological vulnerabilities.

The third pillar is Surveillance Capitalism theory (Zuboff, 2019), which frames behavioral surplus as the raw material for psychological prediction products. Zuboff argues that the extraction, analysis, and monetization of behavioral surplus constitutes a novel economic logic that treats human experience as a free input for production, instrumentalizing psychological knowledge for commercial ends without meaningful consent.

The fourth pillar is cognitive liberty theory (Ienca and Andorno, 2017). Cognitive liberty represents the foundational normative objection to AI psychological profiling deployed for behavioral modification. When recommendation systems use psychological profiles to optimize content delivery for engagement rather than user welfare, they engage in cognitive manipulation that undermines the epistemic autonomy necessary for democratic participation.

4. Methodology

A systematic mixed-methods research design is proposed for empirically investigating AI psychological profiling in media systems. Three integrated studies are recommended.

Study 1 is a Validity Audit: A concurrent validity study with N = 500 in which participants consent to data collection from one or more social media platforms for a 60-day observation period, complete validated Big Five assessments using the BFI-2, and authorize the research team to reconstruct the implicit psychological model derived from their behavioral data. Pearson correlations between AI-inferred and psychometrically measured personality scores will provide validity coefficients. IRT analysis will assess whether behavioral inference captures the same latent dimension structure as validated measures.

Study 2 is a Content Audit: N = 200 participants with diverse personality profiles document their recommendation feed content across three platforms over 30 days using a structured daily diary and screen recording protocol. Multi-level modeling tests whether platform-recommended content systematically differs by personality profile in theoretically predicted directions, such as high-neuroticism profiles receiving more anxiety-inducing content.

Study 3 is a Regulatory Compliance Audit: Legal analysis of data access requests under GDPR Article 15 submitted by 50 participants to major media platforms, examining whether platforms disclose AI-inferred psychological attributes in their subject access responses as required under GDPR Article 22 automated decision-making provisions.

5. Results

Based on the existing empirical literature and the proposed research design, expected findings are as follows. Study 1 validity correlations are expected to range from r = .25 to r = .45 for core Big Five dimensions, below the published Kosinski et al. (2013) estimates, reflecting that most platforms provide less discriminating behavioral signal than Facebook's explicit like architecture. Discriminant validity is expected to be substantially lower than convergent validity, indicating that most AI profiling conflates psychologically distinct constructs.

Study 2 content audit findings are expected to confirm that recommendation feeds are significantly differentiated by neuroticism profile: high-neuroticism users are predicted to receive significantly more negative-valence, conflict-heavy, and threat-oriented content, a prediction consistent with platforms' watch-time optimization incentives.

Study 3 regulatory audit is expected to reveal substantial non-disclosure of AI-inferred psychological attributes, consistent with prior civil society investigations indicating that most platforms' subject access responses omit inferred categories from their disclosed data inventories.

6. Discussion

The findings of the proposed research program would establish an empirical foundation for policy-relevant conclusions about AI psychological profiling. The validity gap between published accuracy claims and empirical validity evidence suggests that commercial profiling products routinely misrepresent their psychometric foundations. The predicted neuroticism-content correlation would provide field-experimental evidence of psychological profile exploitation in recommendation systems.

The epistemic asymmetry concern warrants specific elaboration. When AI systems construct psychological profiles that users cannot inspect, contest, or opt out of, and use those profiles to determine what information reaches users, they implement a system of cognitive management that substitutes algorithmic judgment about users' psychological needs for users' own epistemic agency. This is not hypothetical: it is the current operational logic of every major media platform's recommendation architecture.

The regulatory implications extend beyond disclosure to substantive intervention. Platform governance frameworks relying exclusively on transparency disclosure are insufficient to address welfare harms from engagement-optimizing recommendation. Mandatory welfare-oriented metrics in recommendation optimization objectives, required psychological impact assessments for algorithm updates, and established user rights to welfare-aligned recommendation are each indicated by the theoretical analysis.

7. Limitations

The proposed research faces significant methodological constraints. Platform data access is contingent on platform cooperation or GDPR subject access request mechanisms, both producing incomplete data. The 60-day observation window may not capture the full developmental arc of profile construction. Ground-truth personality assessment through self-report creates an upper bound on validity coefficients. Cross-platform generalizability is limited by each platform's proprietary algorithm architecture.

8. Conclusion

AI-driven psychological profiling in media algorithms represents one of the most consequential applications of psychological science in contemporary society and one of the least subject to the validity standards, ethical oversight, and transparency norms that characterize responsible psychological practice. The APPF framework provides a principled analytical structure for investigating profiling mechanisms, auditing validity claims, and identifying regulatory failures. As AI capabilities for psychological inference from behavioral data continue to advance, the development of governance frameworks adequate to the epistemic and autonomy risks of algorithmic profiling becomes an urgent priority for psychological science, media regulation, and democratic theory.

References

Aarzoo & Lal, R. (2024a). AI-Driven Emotional Storytelling for Brand Narrative Strategies and Consumer Perception. IUP Journal of Brand Management, 21(4), 30–50.

Aarzoo & Lal, R. (2025a). Enhancing Advertising Effectiveness Through AIDA, AI, and Data Visualization Integration for Business Strategies. In M. Muniasamy, A. Naim, & A. Kumar (Eds.), *Data Visualization Tools for Business Applications* (pp. 85-102). IGI Global. https://doi.org/10.4018/979-8-3693-6537-3.ch005

Aarzoo & Lal, R. (2025b). Quality culture in advertising agencies and creativity for campaign effectiveness: Analysis of Six Sigma practices. Social Sciences & Humanities Open, 12, 101891.

Aarzoo & Lal, R. (2026). Challenges in Healthcare Data Journalism: Accuracy, Privacy, and Ethical Reporting in Disease Prediction Trends. In *AI Model Design and Data Management for Disease Prediction* (pp. 299-322). IGI Global Scientific Publishing.

Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 84(3), 488–500.

Boyle, K. (2019). What's in a name? Theorising the inter-relationships of gender and violence. Feminist Theory, 20(1), 19–36.

Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. Proceedings of the International AAAI Conference on Web and Social Media, 7(1), 128–137.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preodiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences, 115(44), 11203–11208.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

Ienca, M., & Andorno, R. (2017). Towards new human rights in the neurotechnology age. Life Sciences, Society and Policy, 13(1), 5.

Joinson, A. N. (2008). Looking at, looking up or keeping up with people? Proceedings of CHI 2008, 1027–1036.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Lazer, D. M. J., Pentland, A., Adamic, L., Aral, S., & Barabasi, A.-L. (2009). Computational social science. Science, 323(5915), 721–723.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research, 30, 457–500.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences, 114(48), 12714–12719.

Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality better than estimated listening time and genre preference. Psychological Science, 29(7), 1145–1158.

Nickerson, D. W., & Rogers, T. (2014). Political campaigns and big data. Journal of Economic Perspectives, 28(2), 51–74.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishers.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin Press.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology, 54(1), 547–577.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media. PLOS ONE, 8(9), e73791.

Settle, J. E. (2018). Frenemies: How social media polarizes America. Cambridge University Press.

Sunstein, C. R. (2017). Republic: Divided democracy in the age of social media. Princeton University Press.

Tambini, D. (2019). Social media power and election legitimacy. In D. Tambini & M. Moore (Eds.), Digital dominance (pp. 265–293). Oxford University Press.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Tufekci, Z. (2015). Algorithmic harms beyond privacy and manipulation. Colorado Technology Law Journal, 13, 116–138.

Vaidhyanathan, S. (2018). Antisocial media: How Facebook disconnects us and undermines democracy. Oxford University Press.

Wu, T. (2016). The attention merchants: The epic scramble to get inside our heads. Knopf.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences, 112(4), 1036–1040.

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a personality-profiling algorithm to investigate political microtargeting. Communication Research, 49(8), 1143–1167.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.