THE PROFILE YOU DIDN'T WRITE: DIGITAL IDENTITY, OCULAR BIOMETRICS AND SEMANTIC AUTHORITY IN PREDICTIVE ALGORITHMIC SYSTEMS

El perfil que no escribiste: Identidad digital, biometría ocular y autoridad semántica en sistemas algorítmicos predictivos

Carlos Eduardo Ravello Joo

ORCID: 0009-0007-5631-7436

carlosravello.com

Trujillo — Lima, Peru

May 2026

Author's note

This work is the second in a series of independent preprints deposited on Zenodo. The first — Ravello Joo (2026) — established the conceptual framework of Metacognition 2.0 and the Dynamic Coherence Model (DCM). This document extends that framework toward the domain of ocular biometrics and algorithmic surveillance as the context that makes the proposed capacity operationally necessary. It has not received external funding. The author declares no conflict of interest.

1

ABSTRACT

This paper examines the architecture of contemporary algorithmic surveillance from the perspective of the individual agent operating in high-visibility digital environments. It argues that an individual's digital identity is not solely the result of their explicit declarations — it is fundamentally a *construction inferred by predictive systems* from behavioral traces that were not deliberately authorized. Documented evidence shows that fitness platforms, professional networks, and health applications generate, aggregate, and commercialize individual behavioral signals with third parties (Federal Trade Commission, 2014; Reviglio, 2022), and that this data is used to construct psychographic profiles with measurable occupational, financial, and insurance consequences (Schneble et al., 2018; Wilcox et al., 2021). The paper introduces an emergent dimension underrepresented in the literature: *ocular biometrics* — iris recognition, scleral pattern recognition, and gaze tracking — as a new frontier of identity capture with unique characteristics of permanence and irrevocability (Bhatt et al., 2025; Doke et al., 2024). It is proposed that these developments make operationally necessary the capacity described in the author's previous work as *Metacognition 2.0* (Ravello Joo, 2026): the ability to deliberately model how artificial intelligence systems process an identity, and to design that identity so that the classificatory output is the one the agent decided. The Dynamic Coherence Model (DCM) is presented as a structural response to this information asymmetry.

Keywords: digital identity, ocular biometrics, algorithmic surveillance, semantic authority, knowledge graph, iris recognition, gaze tracking, data brokers, Metacognition 2.0, Dynamic Coherence Model

1. INTRODUCTION

This article constitutes a proposal for a conceptual framework based on the documented professional experience of the author and a critical narrative review of verifiable academic literature. It is not a formal empirical study.

Most individuals who operate in digital environments assume, implicitly, that their digital identity is the result of what they have published, declared, or consciously shared. That assumption is incorrect.

Contemporary algorithmic systems do not wait for the agent's declaration. They infer. They classify. They build predictive models about them from signals that the agent, frequently, does not know they are producing. The digital breadcrumb trail — searches, likes, physical activity patterns, movement routes, heart rate, facial microexpressions in an automated job interview, the direction of gaze inside an augmented reality viewer — constitutes a set of behavioral data that systems process with increasing precision and real consequences.

The author's previous work (Ravello Joo, 2026) established the conceptual framework for responding to this reality: Metacognition 2.0, defined as the capacity to deliberately model how artificial intelligence systems process a digital identity, and the Dynamic Coherence Model (DCM) as the operational tool for implementing that capacity. This work extends that framework in two directions. First, it documents with greater depth the landscape of algorithmic surveillance that makes that capacity necessary — with emphasis on the commercialization of behavioral data and the use of inferred psychographic profiles. Second, it introduces a dimension that the previous work did not address: ocular biometrics as an emerging frontier of identity capture with radically different characteristics from any prior digital data. An iris cannot be changed. A scleral vein pattern cannot be erased. That transforms the nature of the problem.

2. THEORETICAL FRAMEWORK

2.1 Knowledge graphs, PageRank, and propagation of semantic authority

Modern search systems organize knowledge about entities — people, organizations, concepts — through knowledge graphs: data structures where nodes represent entities and edges represent relationships between them. The authority of a node in that graph is not an intrinsic property — it is a function of the relationships it maintains with other nodes.

The PageRank algorithm, formalized by Brin and Page (1998), establishes that the authority of a node A is calculated as:

$$PR(A) = (1 - d) + d \times SUM\ PR(T)/C(T)$$

Where d is the damping factor (conventionally 0.85), T are the nodes pointing to A, and C(T) is the number of outgoing links of each node T. The operational implication is direct: the authority of an entity in the graph depends on the authority of the entities it is connected to. An entity connected to low-authority or reputationally compromised nodes sees its own authority value descend proportionally — not as a moral consequence but as a mathematical operation.

Kejriwal (2023) documented that in personal knowledge graphs — those that organize information about individuals — entity resolution depends critically on coherence between signals from independent sources, and that the absence of that coherence produces fragmented or low-authority classifications. Hu et al. (2025) extended that analysis toward neural network-based entity resolution systems in property graphs, documenting that attribute propagation between adjacent nodes operates automatically and independently of the agent's intent.

2.2 Graph embeddings and vectorial repositioning of identity

Modern graph processing systems go beyond static PageRank. Grover and Leskovec (2016) formalized Node2Vec as a framework for learning node representations that preserves

neighborhood structure: each node in the graph becomes a numerical vector in a high-dimensional space, where proximity between vectors reflects semantic similarity between nodes. Wang et al. (2021) extended that approach toward knowledge graphs through KG2Vec, documenting that semantic relationships between entities are preserved in the resulting vector space.

The consequence for digital identity is that an association between nodes — a photo at a party, an appearance alongside an entity of compromised reputation, an interaction on a social media platform — does not produce merely a reputational effect in the human sense. It produces a mathematical repositioning of the affected node in the vector space. That repositioning has measurable consequences in the classification systems operating on that space.

2.3 Ocular biometrics: iris, sclera, and gaze tracking

Ocular biometrics represents a qualitatively distinct category from the digital behavioral data described thus far. While a leaked password can be changed, a phone number can be modified, and a pattern of social media activity can be altered through deliberate effort, the iris and the scleral vein pattern are permanent and irrevocable anatomical characteristics.

Bhatt et al. (2025) documented in a systematic review that iris recognition has reached precision levels positioning it as one of the most reliable biometric identifiers available, with applications ranging from forensic identification to access control in high-security facilities. Anne et al. (2019) evaluated the feasibility and acceptability of iris recognition systems for unique patient identification in real healthcare service environments in Kenya, documenting both their technical effectiveness and the ethical tensions their implementation generates in populations with limited access to information about their rights.

Doke et al. (2024) documented scleral pattern recognition — recognition of the white of the eye's vein pattern — as an emerging biometric modality with uniqueness properties superior to the iris itself, harder to falsify and more resistant to the environmental conditions that affect conventional iris recognition. Crihalmeanu and Ross (2012) established the technical foundations of multispectral scleral pattern recognition, documenting its viability as a robust identification system.

Gaze tracking — the tracking of the direction and intensity of gaze — adds a behavioral dimension to the biometric. Rodrigues et al. (2026) developed a framework integrating eye tracking, machine learning, and facial recognition to infer consumer behavior with precision exceeding traditional survey methods. What an individual looks at, for how long, and with what intensity is behavioral data as informative as what they search for in a search engine — and in some contexts more so, because it is not mediated by the agent's conscious rationalization.

2.4 Emotion recognition and algorithmic evaluation of persons

Emotion recognition systems extend biometric capture toward the domain of inferred internal states. Kulke et al. (2020) compared Affectiva's facial expression analysis software with direct electromyographic measurements of facial expressions, documenting its capacity to infer emotional states from video with precision comparable to direct physiological measurement.

HireVue — an automated interview platform used by hundreds of Fortune 500 companies — analyzes facial microexpressions, vocal tone, and linguistic content of candidates during video interviews, generating an employability score before a human evaluator reviews the candidacy. Ajunwa (2022) criticized this type of system comparing it to phrenology — the 19th century pseudoscience that claimed to infer mental capacities from physical characteristics — arguing that the predictive validity of these systems is not adequately established and that their algorithmic biases reproduce and amplify pre-existing discriminations. Kammerer (2022) analyzed the legal implications of these platforms from the perspective of labor law and privacy, documenting the tensions between their corporate adoption and existing regulatory frameworks.

3. LITERATURE REVIEW AND CASE STUDIES

3.1 The commercialization of behavior: data brokers and psychographic profiles

The Federal Trade Commission documented in 2014 that twelve health and fitness applications shared behavioral data from their users with seventy-six different third parties, including geolocation, heart rate, and physical activity patterns. That practice has not diminished — it has been normalized and industrialized.

Reviglio (2022) analyzed the role of data brokers in the surveillance economy, documenting that they operate as intermediaries that aggregate behavioral data from multiple sources, build profiles of individuals, and commercialize them to employers, insurers, financial institutions, and government agencies. Crain (2018) documented the limits of transparency in that market, arguing that the structural opacity of data brokers is a functional characteristic of the system, not an accidental regulatory deficiency.

Cambridge Analytica built psychographic profiles of approximately 87 million Facebook users from data obtained through a personality test application voluntarily installed by approximately 270,000 users — whose permissions at the time allowed access to all their contacts' data. Schneble et al. (2018) analyzed the case from the perspective of Internet-mediated research ethics, documenting the tensions between existing consent frameworks and the data extraction capabilities that social platforms permitted. Hu (2020) examined the algorithmic opacity of the Cambridge Analytica system, arguing that its predictive capacity over electoral behavior was technically plausible given the volume and granularity of available data.

3.2 Ocular biometrics in deployment: documented cases

Worldcoin — rebranded as World in 2024, founded by Sam Altman — operated a biometric registration system based on iris scanning in multiple developing countries, including Peru, Kenya, India, and Indonesia, offering cryptocurrency as compensation for registration. Calungsod (2025) evaluated the privacy and security implications of World App's biometric system, documenting the tensions between the platform's business model and the rights of users who ceded irrevocable biometric data in contexts of information asymmetry.

Dubai and Heathrow airports operate iris recognition systems as a component of their access control and traveler identification processes. Meta holds registered patents for gaze tracking systems within its Quest augmented reality devices, with the declared purpose of improving user experience — and the undeclared purpose of generating high-granularity visual behavioral data.

3.3 Behavioral scoring: insurance, employment, and credit

John Hancock, a life insurer, offers policies whose premium adjusts dynamically based on physical activity data recorded by the insured's wearable devices. LinkedIn operates Talent Insights, a commercial product generating "flight risk" signals — risk of an employee leaving the organization — from their activity pattern within the platform, and commercializes it to employers. Shaw et al. (2022) documented in Psychological Science that individual digital traces present stable intraindividual consistency sufficient to infer personality traits and predict future behavior. Wilcox et al. (2021) confirmed that digital footprint scrutiny has been routinely incorporated into personnel selection processes, with prevalence exceeding sixty percent in medium and large companies.

4. DISCUSSION

4.1 Biometric asymmetry as an irreversible problem

The digital behavioral data described in the previous work (Ravello Joo, 2026) share a characteristic that makes them manageable in principle: they are modifiable. A social media activity pattern can be altered. A search footprint can be redirected. An interaction history can be reoriented through deliberate and sustained effort.

The iris does not share that characteristic. The scleral vein pattern does not either. That qualitatively transforms the nature of the problem.

When Worldcoin scanned the iris of individuals in Kenya and Peru in exchange for modest monetary compensation, it did not capture data those individuals could modify if they subsequently decided the cession was an error. It captured a permanent and irrevocable anatomical characteristic — and incorporated it into a system whose future use the cedents do not control.

This introduces an ethical and operational dimension that the digital privacy literature generally does not address with the specificity the problem requires: the difference between data that can be

revoked and data that can never be revoked. A leaked password is changed. A compromised email address is abandoned. An iris scanned without genuine informed consent is a permanent cession with no possibility of retraction.

4.2 Authority propagation as a contagion mechanism

The knowledge graph framework establishes that the semantic authority of a node depends on the entities it is connected to (Brin & Page, 1998; Grover & Leskovec, 2016). What the technical literature describes in terms of mathematical optimization has concrete human consequences: association with entities of compromised reputation — in photographs, in mentions, in social media interactions — produces a vectorial repositioning of the affected node with measurable effects in the classification systems operating on that graph.

That mechanism operates automatically, without human intervention, at speeds exceeding the agent's capacity to detect and respond to the damage. Google crawls Meta directly since 2025, meaning the interval between the production of a compromising association and its incorporation into the knowledge graph is hours, not days or weeks. Building verifiable semantic coherence takes weeks. Degrading it through association can take a single crawl.

4.3 Metacognition 2.0 as a structural response

The previous work (Ravello Joo, 2026) proposed Metacognition 2.0 as the capacity to deliberately model how artificial intelligence systems process a digital identity. The landscape documented in this work makes that capacity not only desirable but operationally necessary for any agent operating in high-visibility algorithmic environments.

The Dynamic Coherence Model (DCM), formalized through the pseudo-ratio:

$$\Omega = V/(M+I)$$

provides the operational framework for implementing that capacity. In the context of ocular biometrics, the model's parameters acquire an additional dimension: V includes not only declared digital signals but the biometric ones the agent cedes in any interaction with recognition systems; M must consider the active restriction of biometric exposure in contexts where that exposure is not necessary; I incorporates the uncertainty about future uses of already-ceded biometric data.

The system's natural attractor at $\Omega \approx 0.66$ (Ravello Joo, 2026) remains operationally valid — but the irrevocable nature of biometric data modifies the error cost function: underestimating M in the biometric domain produces a permanent cost that cannot be corrected through subsequent adjustment of the system.

4.4 Regulatory implications

The European GDPR classifies biometric data as a special category of personal data requiring enhanced legal basis for processing. The Illinois Biometric Information Privacy Act (BIPA) establishes in the United States that entities capturing biometric data must obtain explicit informed consent and declare the purposes and timelines of use. Both regulatory frameworks operate on the consequence — the unauthorized capture already occurred — not on the cause: the information asymmetry that makes it possible for individuals to cede irrevocable biometric data without fully understanding the implications.

In Latin America, equivalent frameworks — LGPD in Brazil, similar legislation in Argentina, Colombia, Mexico, and Peru — are in the process of maturation. That relative regulatory gap makes especially relevant the development of individual capacities for deliberate management of digital identity, given that systemic protection does not operate with the same effectiveness as in jurisdictions with more consolidated frameworks.

5. CONCLUSIONS

An individual's digital profile is not the exclusive result of their conscious declarations. It is fundamentally a construction inferred by predictive systems from behavioral traces produced without deliberate authorization — and that inference has measurable occupational, financial, insurance-related, and semantic consequences.

Ocular biometrics adds a qualitatively new dimension to that problem: irrevocability. Digital behavioral data is modifiable with effort and time. Ocular biometric data is not. That difference is not technical — it is ethical and operational, and academic literature has not addressed it with the specificity it requires.

Metacognition 2.0 (Ravello Joo, 2026) and the Dynamic Coherence Model provide a framework for responding to that asymmetry from the perspective of the individual agent. They do not eliminate the asymmetry — they reduce it through deliberate design of the signals the agent produces for the systems that classify them. In the biometric domain, that reduction implies active restriction of exposure to capture systems whose future use is not guaranteed by effective regulatory frameworks.

They said that in the gaze lies the soul. Now it is literal — not as allegory but as data architecture. And unlike almost any other digital datum, the soul that lies in the gaze cannot be changed like a password.

LIMITATIONS

This work presents methodological limitations that the author explicitly declares. The literature review is narrative and not strictly systematic, which introduces selection bias in the cited sources. The DCM application cases in the biometric context are speculative — the model was developed in the context of digital semantic identity and its extension to the biometric domain requires independent empirical validation. Biometric regulation evolves rapidly and some aspects of the regulatory analysis may become outdated in the short term. Future research should address the empirical validation of the DCM in contexts of biometric exposure management and comparative analysis of regulatory frameworks in Latin America.

REFERENCES

Ajunwa, I. (2022). Automated video interviewing as the new phrenology. *Scholarship@UNC.* University of North Carolina School of Law.

Anne, N., Blanco, E., Sambah, E., Piper, J., Kiptoo, P., & Lascko, T. (2019). Feasibility and acceptability of an iris biometric system for unique patient identification in routine HIV services in Kenya. *International Journal of Medical Informatics, 133,* 104006. https://doi.org/10.1016/j.ijmedinf.2019.104006

Bhatt, S., Bhatt, U., & Bhatt, S. (2025). A systematic review of iris biometrics in forensic science: Applications and challenges. *Egyptian Journal of Forensic Sciences, 15,* 12. https://doi.org/10.1186/s41935-025-00431-7

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*(1-7), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Calungsod, M. P. D. (2025). Iris scanning: An evaluation of data privacy and security in World App's biometric system. *Cognizance Journal of Multidisciplinary Studies.*

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society, 20*(1), 88–104. https://doi.org/10.1177/1461444816657096

Crihalmeanu, S., & Ross, A. (2012). Multispectral scleral patterns for ocular biometric recognition. *Pattern Recognition Letters, 33*(14), 1860–1869.

Doke, K. K., Shelke, S., & Raut, S. (2024). A closer look at sclera: Emerging trends in biometric authentication. *IEEE Xplore.*

Federal Trade Commission. (2014). *Data brokers: A call for transparency and accountability.* Federal Trade Commission. https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference* (pp. 855–864). ACM. https://doi.org/10.1145/2939672.2939754

Hu, J., Qin, C., Li, J., Gao, H., & Li, J. (2025). When GDD meets GNN: A knowledge-driven neural approach for entity resolution in property graphs. *Information Systems.*

Hu, M. (2020). Cambridge Analytica's black box. *Big Data & Society, 7*(2). https://doi.org/10.1177/2053951720938091

Kammerer, B. (2022). The legal implications of artificial intelligence video interviewing. *Iowa Law Review.*

Kejriwal, M. (2023). Named entity resolution in personal knowledge graphs. *arXiv preprint.* https://arxiv.org/abs/2307.01557

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A comparison of the Affectiva iMotions facial expression analysis software with EMG for identifying facial expressions of emotion. *Frontiers in Psychology, 11,* 329. https://doi.org/10.3389/fpsyg.2020.00329

Ravello Joo, C. E. (2026). Metacognición 2.0: Diseño deliberado de identidad digital ante sistemas predictivos de inteligencia artificial — El Modelo de Coherencia Dinámica (MCD). *Zenodo.* https://doi.org/10.5281/zenodo.20092009

Reviglio, U. (2022). The untamed and discreet role of data brokers in surveillance capitalism. *Internet Policy Review*, *11*(3). https://doi.org/10.14763/2022.3.1670

Rodrigues, J. A., Sousa, A., & Carneiro, D. (2026). Advanced consumer behaviour analysis: Integrating eye tracking, machine learning, and facial recognition. *MDPI.*

Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO Reports*, *19*(8), e46579. https://doi.org/10.15252/embr.201846579

Shaw, H., Ellis, D. A., Geyer, K., Davidson, B. I., Ziegler, F. V., & Smith, A. (2022). Subjective reports overstate the relationship between screen time and mental health. *Psychological Science*, *33*(8), 1421–1432. https://doi.org/10.1177/09567976211040491

Wang, Y. Q., Li, X. L., Liao, B., Luo, J., & Cai, L. J. (2021). KG2Vec: A node2vec-based vectorization model for knowledge graph. *PLoS ONE, 16*(3), e0248552. https://doi.org/10.1371/journal.pone.0248552

Wilcox, A., Damarin, A. K., & McDonald, J. A. (2021). Is cybervetting valuable? *Industrial and Organizational Psychology*, *15*(3), 315–333. https://doi.org/10.1017/iop.2021.108

Author: Carlos Eduardo Ravello Joo — ORCID: 0009-0007-5631-7436
This preprint is the second in a series. The first is available at:
https://doi.org/10.5281/zenodo.20092009
May 2026