METACOGNITION 2.0: DELIBERATE DIGITAL IDENTITY DESIGN

AGAINST PREDICTIVE ARTIFICIAL INTELLIGENCE SYSTEMS

The Dynamic Coherence Model (DCM) as a response to algorithmic information asymmetry

Carlos Eduardo Ravello Joo

ORCID: 0009-0007-5631-7436

carlosravello.com

Wikidata: Q139714842

Trujillo — Lima, Peru

May 2026

Author's note: This work is an independent preprint deposited on Zenodo. It has not received external funding. The author declares no conflict of interest. The chronology of development is made explicit in the discussion section for methodological relevance.

1

ABSTRACT

This paper presents a conceptual framework derived from over a decade of professional practice in the design of digital systems and complex entities. It introduces the concept of *Metacognition 2.0*, defined as the capacity to deliberately model how predictive artificial intelligence systems process, classify, and assign authority to a digital identity. The core premise is that this capacity is no longer optional. Documented evidence shows that digital platforms — from fitness applications to professional networks — generate, aggregate, and share individual behavioral signals with third parties (Federal Trade Commission, 2014; The Guardian, 2018), and that digital behavioral records allow inference of personality traits and personal attributes with increasing precision (Kosinski et al., 2013). The agent who does not deliberately manage how they are processed by these systems is not invisible — they are processed by default, with classifications they did not control and frequently do not know about. The *Dynamic Coherence Model (DCM)* is proposed as an operational framework to address this asymmetry, formalized through the pseudo-ratio $\Omega = V/(M+I)$, where V represents disclosure, M over-protection, and I managed uncertainty. The model was explored through computational simulation with a suggested attractor at $\Omega \approx 0.66$. Three practical application illustrations across unrelated industries are described as plausibility evidence. Results suggest that disciplinary diversity, correctly structured, generates verifiable algorithmic coherence and reduces the asymmetry between the agent and the systems that classify them.

Keywords: metacognition, predictive systems, digital identity, algorithmic asymmetry, behavioral coherence, artificial intelligence, knowledge graph, entity processing, data surveillance

1. INTRODUCTION

This article constitutes a proposal for a conceptual framework based on the documented professional experience of the author, not a formal empirical study. That methodological precision is relevant because the problem this work addresses is not theoretical — it is operational, present, and verifiable.

Contemporary algorithmic systems do not wait for the agent to declare who they are. They infer. They classify. They assign authority or withdraw it. And they do so with data that the agent, frequently, does not know they are producing.

The Federal Trade Commission documented in 2014 that twelve health and fitness applications shared data with seventy-six different third parties, including geolocation, heart rate, and physical activity patterns. In 2018, Strava accidentally exposed the location of US military bases by publishing aggregated heat maps of its users — data that the platform itself commercializes in anonymized form with municipalities for urban mobility analysis (The Guardian, 2018). LinkedIn operates a commercial product called Talent Insights that infers, among other metrics, the "flight risk" of employees from their activity patterns on the platform. Insurance companies such as John Hancock offer life insurance policies whose premium adjusts dynamically based on data from the insured's Apple Watch or Fitbit.

Academic literature has begun to formalize what these practices imply. Kosinski et al. (2013) demonstrated in PNAS that digital behavioral records allow prediction of personality traits, political orientation, religious beliefs, and other personal attributes with high precision. The systematic review published in the Journal of Business and Psychology (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.

The individual agent has been captured in a structural information asymmetry: algorithmic systems generate predictive models about them with data they did not deliberately authorize, and return classifications — occupational, financial, insurance-related, semantic — that affect their real opportunities. Regulation attempts to keep pace (GDPR in the European Union, CCPA in California, LGPD in Brazil, equivalent laws in Argentina, Colombia, Mexico, and Peru) but operates on the consequence, not the cause.

This paper proposes that there is a higher-order cognitive skill, not formalized under this name in the reviewed literature, that allows the agent to reduce that asymmetry: metacognition 2.0. It is not prompting. It is not technical SEO. It is not personal brand management. It is the capacity to mentally model how an artificial intelligence system processes an identity — and to design that identity so that the classificatory output is the one the agent decided, not the one the algorithm inferred by default from unauthorized signals.

To formalize that capacity as an operational tool, the Dynamic Coherence Model (DCM) is presented.

2. THEORETICAL FRAMEWORK

2.1 Metacognition: from the classical concept to digital asymmetry

Flavell (1979) defined metacognition as "cognition about cognition" — the agent's capacity to monitor and regulate their own cognitive processes. Brown (1987) extended the framework toward self-regulation of learning, distinguishing between the knowledge the agent has about their processes and the active control they exercise over them. Rouault et al. (2018) documented high individual variability in metacognition that remains stable across domains and is partially independent of general intelligence — implying that it is a developable skill independent of the baseline cognitive level.

The extension this paper proposes is the following: in an environment where algorithmic systems generate predictive models about the agent with data the agent did not deliberately authorize, classical metacognition proves insufficient. Observing one's own thinking is no longer enough. It is necessary to observe how other systems — artificial, automated, opaque — are observing the agent. That second layer of observation is what is proposed here under the name metacognition 2.0.

2.2 The predictive brain and its artificial extension

Friston (2010) formalized the free-energy principle as a unified model of brain functioning: the nervous system operates as a predictive machine that continuously minimizes the discrepancy between its predictions of the world and incoming sensory data. The connection with contemporary algorithmic systems is structural, not metaphorical. Modern search engines, knowledge graphs, and language models operate under equivalent predictive logics: they generate hypotheses about the identity and behavior of entities from available signals, and resolve ambiguity by selecting the classification that minimizes system surprise (Bunescu & Pa■ca, 2006). 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.

The agent who understands this mechanism can deliberately design the signals they produce to reduce the surprise of the classifying system. That is the nuclear operation of metacognition 2.0: applying the free-energy principle not to their own brain, but to the artificial brain that classifies the agent.

2.3 Behavioral coherence as an algorithmic metric

Kosinski et al. (2013) demonstrated that digital behavioral records present sufficient consistency to infer personal attributes with high precision. What the agent does online, in their applications, in their wearable devices, is not noise — it is pattern. And that pattern is readable. Commercial platforms have integrated that readability into their products: LinkedIn infers intent to change jobs, insurers infer health risk, HR systems infer cultural fit (Wilcox et al., 2021).

The concept of "behavioral coherence" as an algorithmic metric does not yet appear formalized under that exact name in the technical literature, but operates under other names in active commercial systems: behavioral scoring, digital footprint analysis, predictive behavioral modeling. Coherence inference is already an established capability of contemporary algorithmic systems.

2.4 Systemic coherence and disparate domains

Ashby (1956) established the Law of Requisite Variety: only variety can absorb variety. A system operating in multiple domains simultaneously is not intrinsically incoherent — it may be the expression of a high-variety method applied consistently. Bateson (1972) extended that framework toward the ecology of mind, documenting that patterns connecting apparently unrelated domains are frequently more informative than patterns within each domain. Prigogine and Stengers (1984) contributed the thermodynamic framework: systems far from equilibrium do not disintegrate under tension — they reorganize, generating dissipative structures of greater complexity.

Disciplinary diversity, under this integrated perspective, is not a semantic deficit. It is useful tension. The operational question is not how to eliminate it, but how to structure it so the classifying system reads it as pattern rather than noise.

2.5 Digital identity and deliberate construction

Robles-Carrillo (2024) documented that digital identity is not a static projection of a pre-existing identity but an active construction operating under its own logics, partially independent of the agent who generates it. That construction can be managed deliberately — generating coherence between the signals the agent produces — or it can remain subject to the default inferences of the systems that process it, frequently with outcomes the agent did not choose.

Metacognition 2.0 is the cognitive skill that enables the first option. The DCM is the operational tool that implements it.

3. THE DYNAMIC COHERENCE MODEL (DCM)

3.1 Conceptual foundation

The DCM postulates that the functional persistence of complex systems — human, organizational, or digital — is sustained by the capacity to maintain coherence among three interdependent vectors. In its formulation for digital systems, these vectors are:

V (Disclosure): the verifiable signals the system actively delivers to the algorithm.

M (Over-protection): the signals the system conceals beyond what is necessary, generating reduced visibility from excess reserve.

I (Managed uncertainty): the elements the system recognizes as unresolved and administers in deferred form, without pretending premature resolution.

The pseudo-ratio Ω is defined as:

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

3.2 Operational interpretation

The natural attractor of the system was explored through computational simulation (parameters available upon request to the author), suggesting convergence at $\Omega \approx 0.66$. This value represents the point where disclosure is sufficient to generate classification with authority, without exposing elements that would increase the uncertainty of the classifying system to the point of fragmenting the identity.

Values of Ω above 1.0 produce overexposure — the system discloses more than it can sustain with verifiable coherence, generating contradictory signals the algorithm cannot cleanly resolve and that tend to degrade the assigned classificatory authority. Values of Ω below 0.50 produce functional invisibility — the system does not generate sufficient verifiable signal for the algorithm to build an authoritative classification, and the agent is left processed by default inferences. The optimal operational zone lies between 0.55 and 0.80.

3.3 The paradox of dynamic coherence

The author proposes the following relationship:

> *"A system achieves its maximum coherence when it accepts its own incoherence as part of the dynamic flow that sustains it."*

The absolute elimination of dispersion is operationally counterproductive in the digital domain. An agent with activity in a single domain generates less semantic variety — and therefore less capacity to absorb the variety of the contemporary algorithmic environment (Ashby, 1956). Controlled and structured dispersion is an advantage, not a deficit, when the underlying method is readable.

3.4 Operational principle of the DCM

The author proposes the following operational principle:

> *"Every system seeks coherence between what it knows, what it feels, and what it does. When one of those vectors becomes misaligned, the system compensates with energy, time, or identity."*

In its application to the digital domain: when the identity declared by an agent is not coherent with the verifiable signals their systems produce, the algorithm resolves the ambiguity by default. That default resolution frequently reduces the agent to their simplest available category — and that simplification has measurable operational costs in visibility, authority, and opportunity.

4. METACOGNITION 2.0: OPERATIONAL DEFINITION

Metacognition 2.0 is the capacity to deliberately model how an artificial intelligence system — a search engine, a knowledge graph, a language model, a behavioral scoring system — processes, classifies, and assigns authority to a digital entity, and to design that entity so that the classificatory output is the one the agent decided, not the one the algorithm inferred by default.

It is not prompting. Prompting operates on the output of a system in a specific instance. Metacognition 2.0 operates on the semantic architecture the system persistently constructs about the agent.

It is not technical SEO. Technical SEO optimizes signals to improve ranking in search results. Metacognition 2.0 optimizes the semantic coherence of the agent's identity as a classifiable entity in knowledge graphs and predictive systems.

It is not personal brand management. Personal brand management operates on human perception. Metacognition 2.0 operates on algorithmic perception — which is the growing mediator between the agent and human perception.

The difference is structural. Previous practices act on visibility. Metacognition 2.0 acts on gravity — the capacity of the node to attract authority classifications autonomously and sustainably in systems the agent does not control but can model.

A critical observation: the bias of artificial intelligence systems does not primarily operate at the extremes. It operates with greater danger at the midpoint — in the validation of what the agent already declared, without capacity to contrast with the undeclared real context. An AI system that validates the coherence of a digital identity without access to the actual operational context produces a structurally hollow confidence signal. Recognizing that limit is a constitutive part of metacognition 2.0. Algorithmic validation is not real validation — it is pattern detection. The difference matters.

5. PRACTICAL APPLICATION ILLUSTRATIONS

The following cases are presented as practical application illustrations of the proposed framework, not as formal empirical validation. They constitute plausibility evidence for the DCM in documented real contexts, all from the author, which constitutes an explicitly declared methodological limitation.

5.1 Precision mechanics — BikeLab Studio

Applied engineering workshop for bicycles founded in Trujillo, Peru. DCM application involved: publication of four technical notes on Zenodo with verifiable DOI (bearing tribology, hydraulic brake thermodynamics at altitude, transmission comparative analysis, carbon footprint in manufacturing); development of digital tools with structured schema and open source; construction of entity on Wikidata with verifiable identifiers across multiple independent systems (ORCID, GitHub, academic repositories). Documented result in Google Search Console: recurring organic traffic from ninety-six countries without advertising investment, stable organic positioning for high-specificity technical terms in international search.

5.2 International veterinary export — Zoovet Travel

Veterinary clinic specializing in international pet export, in continuous operation since 2013. DCM application involved: construction of verifiable entity for the scientific director with cross-referenced identifiers (ORCID 0009-0002-6837-5311, CONCYTEC Renacyt registration 0140858, Wikidata Q138881218); public repository of FAVN clinical cases with verifiable Zenodo DOI (10.5281/zenodo.19797479); trilingual semantic architecture (Spanish, English, French) with MedicalWebPage and Dataset schema declared. Documented result: recurring appearance in Google AI Overview for high-specificity terms in international veterinary export from Peru, with documented cases attended in over thirty countries.

5.3 Gastronomic operation — Katsudomo

Japanese izakaya concept bar inaugurated in Lima, Peru, in 2026. DCM application involved: operating license registered on the government platform gob.pe (Submanagerial Resolution D001442-2026-SCA-GDE-MDS of the Municipality of Surquillo); remote operation infrastructure with automatic daily reconciliation from the first day of opening; active Google Business Profile with semantic positioning in niche terms unoccupied by established competitors. This case documents that application of the framework does not depend on the domain of activity or operational seniority.

6. DISCUSSION

6.1 Chronology of development: methodological note

A chronological precision is relevant to the interpretation of this work. The conceptual development of the DCM preceded the systematic discovery of the documented evidence on algorithmic behavioral inference practices presented in the introduction. The model was constructed as a response to an operational problem observed by the author in their own professional practice: the difficulty of search systems in classifying agents who operate simultaneously in multiple domains.

The systematic evidence on commercialization of behavioral data, algorithmic coherence inference, and use of digital footprint in evaluation of persons was identified subsequently, during the preparation of this preprint. That subsequent discovery did not modify the model. It confirmed that the problem the model addressed operationally was structurally deeper and considerably better documented than the author anticipated at the outset of development.

This precision is not rhetorical. It has methodological implication: the DCM was not designed to respond to a problem previously identified in the literature. It was designed to respond to a problem observed in practice, and its conceptual validation comes partially from the fact that an independent body of evidence, prior to the model’s development, confirms the relevance of the problem it addresses.

6.2 Central implication

Contemporary algorithmic systems calculate behavioral coherence about individuals. That practice is not speculative — it is documented (Federal Trade Commission, 2014; The Guardian, 2018; Kosinski et al., 2013; Wilcox et al., 2021). The agent who does not deliberately manage how they are processed by these systems is not invisible. They are processed by default, with classifications that affect real opportunities (occupational, financial, insurance-related, semantic) and that they frequently do not know about.

Metacognition 2.0 does not propose a solution to algorithmic surveillance — regulation operates that front. It proposes a response to structural information asymmetry: if systems calculate coherence about the agent, the agent can deliberately calculate and design the coherence those systems will read. The asymmetry is not eliminated, but it is reduced.

6.3 Limitations

This work presents methodological limitations that the author explicitly declares. First, exclusive reliance on the author’s own cases limits the generalizability of observations. Robust validation of the DCM would require application to third-party cases with documented access to independent digital performance metrics. Second, the pseudo-ratio Ω and its proposed thresholds derive from computational simulation without independent field data to validate them. Third, the connection between the classical metacognition literature and the digital domain proposed in this work is the author’s interpretation, not an established extension in the academic community. Fourth, the cases presented operate in a single geographic (Peru) and cultural (Latin American) context, which limits generalization to other regulatory and digital environments.

Future research should apply the DCM to third-party cases, subject the model parameters to independent validation, and explore the applicability of the framework in different regulatory contexts (European Union under GDPR, United States under state regulations, Asia-Pacific).

7. CONCLUSIONS

Classical metacognition operates on one’s own thinking. Metacognition 2.0 operates on how artificial systems think about the agent. That extension is not speculative — it is the logical consequence of operating in an environment where classification algorithms generate predictive models about individuals with data they did not deliberately authorize, and where those models determine what exists with authority and what does not.

The Dynamic Coherence Model provides an operational framework for applying that capacity systematically. The pseudo-ratio $\Omega = V/(M+I)$, with a suggested attractor at 0.66, allows diagnosis of the semantic coherence state of any system — personal, organizational, or digital — and the design of precise interventions that reduce the information asymmetry between the agent and the systems that classify them.

Disciplinary diversity is not the problem the algorithm needs to resolve about an agent. It is the pattern that, correctly documented and structured, constitutes the evidence most difficult to falsify: that the method is independent of the domain.

Whoever understands that stops competing for visibility. They begin to build gravity.

REFERENCES

Ashby, W. R. (1956). *An introduction to cybernetics.* Chapman & Hall.

Bateson, G. (1972). *Steps to an ecology of mind.* Chandler Publishing.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Lawrence Erlbaum Associates.

Bunescu, R., & Pa■ca, M. (2006). Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of the 11th Conference of the European Chapter of the ACL* (pp. 9–16). ACL.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain.* Putnam.

Dunning, D., & Kruger, J. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

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

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Google. (2023). *Search quality rater guidelines.* Google LLC.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

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

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

Maturana, H., & Varela, F. (1987). *The tree of knowledge: The biological roots of human understanding.* Shambhala Publications.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature.* Bantam Books.

Ravello Joo, C. E. (2026). Dynamic Coherence Model (DCM): Operational framework for complex systems [Working paper]. DCM Project. ORCID: 0009-0007-5631-7436.

Robles-Carrillo, M. (2024). Digital identity: An approach to its nature, concept and implications. *International Journal of Law and Information Technology, 32*(1). https://doi.org/10.1093/ijlit/eaad014

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry, 84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.003

The Guardian. (2018, January 28). Fitness tracking app Strava gives away location of secret US army bases. https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

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

Preprint deposited on Zenodo under Creative Commons CC BY 4.0 license.
Author: Carlos Eduardo Ravello Joo — ORCID: 0009-0007-5631-7436 — Wikidata: Q139714842
May 2026