Learning what cannot vary: memory as a constraint of the world

Why latent predictive architectures shift the problem of learning without crossing the threshold of biographical memory

Fourth volume of the series *Encoding, transduction and world models*. JEPA-type architectures do not learn to represent the world but the constraints that make certain transformations predictable. This shift matters, but it operates at a level where the notion of biographical memory is not defined.

1. Introduction: what the trilogy leaves open

The three previous articles in this series (Parts 1, 2 and 3) defended a thesis progressively elaborated. Every cognitive architecture, biological or artificial, operates through representational mediation. Living systems access their representations through a perception-action loop of which language models are structurally deprived. The deepest difference between human cognition and contemporary architectures lies less in the content of representations than in the nature of the relations that connect them: statistical co-occurrence on one side, biographical edge on the other.

For what follows, it matters to stabilize a functional definition of this latter notion, independently of any phenomenological commitment. A *biographical edge* designates a relation between mnesic representations that simultaneously satisfies four operational conditions:

1. Indexation on the history of a continuous agent,
2. Co-activation within the same episode,
3. Preservation of an occurrence context,
4. Possibility of situated recall from several modal perspectives.

This definition is deliberately functional. It can be held without prior commitment on the subjective nature of experience, and it is precisely this translation that will be defended in section 11 against the objection that it would abolish the earlier phenomenological position.

The question the trilogy leaves open is then formulated as follows: if the deficit of large language models lies in their secondary symbolic anchoring, and if robotic embodiment alone does not suffice to produce a memory thus defined, what do self-supervised predictive architectures, JEPA and related architectures of latent prediction, actually bring to this landscape?

The thesis defended here is the following: these architectures shift the objective of learning in an epistemically decisive manner, by passing from prediction of observations to prediction of constraints. This shift matters. It operates, however, at a level where the conditions of the biographical edge are not defined. Not that it fails to satisfy them; it does not address them.

The title assumes a rhetorical compression: it is not about learning absolute invariants, but learning what, within a given distribution, must remain stable to preserve predictability.

The domain of validity of this article is restricted: we speak of self-supervised latent predictive architectures, of which JEPA is the most discussed exemplar, and of their agentic extensions with external memory when these present themselves as crossings of the identified threshold. The conclusions do not extend mechanically to LLMs alone, to purely generative models, nor to complete composite agentic systems.

2. Three levels not to be conflated

The analysis that follows operates on three levels that must be distinguished explicitly, failing which transitions become ambiguous.

1. The **computational level** concerns the mechanics of architectures: encoders, latent spaces, prediction functions, learning objectives, regularization mechanisms.
2. The **epistemic level** concerns what is learned in the strong sense: structure of coherence, invariants, constraints captured in the representation.
3. The **phenomenological level** concerns memory as lived by a situated subject, with its properties of autonoetic consciousness, affective modulation and multi-perspective navigability.

These three levels are not substitutable.

A property at the computational level does not mechanically imply a property at the epistemic level; an epistemic property does not mechanically imply a phenomenological property.

Conflating the levels produces two symmetric errors: over-attribution ("JEPA understands the world"), and under-attribution ("JEPA has no memory").

JEPA does not fail to produce a biographical memory. It operates at a level where this notion is not defined.

This precision is not a diplomatic concession. It conditions what we are entitled to expect of these architectures in regulated environments, particularly in healthcare, where out-of-distribution robustness is not a secondary objective but a compliance requirement.

3. The framing error: learning more is not learning better

A naively empiricist reading, sometimes associated, abusively but conveniently, with Aristotelian induction, rests on the idea that from a sufficient accumulation of observations, the mind extracts by abstraction the regularities of the world. This schema has structured the majority of contemporary machine learning: more data, more parameters, more capacity, and generalization would follow.

The limits of this posture are now documented with an almost embarrassing regularity. Overfitting on superficial correlations, massive dependence on annotations, out-of-distribution fragility, sensitivity to shortcut features. These phenomena are not residual bugs that will dissolve at the next doubling of scale. They are the direct consequence of an ill-posed learning objective: learning to reconstruct what has been observed is not learning what structures the observations.

The implicit thesis of latent predictive architectures is more precise: the problem is not to learn more data, but to change what is learned.

4. The JEPA shift: genealogy and principle

JEPA does not arrive in an empty landscape. It is part of a lineage of self-supervised architectures that has structured representation research since 2020: the contrastive approaches SimCLR [29] and MoCo [30], then BYOL [31], which shows that an asymmetric prediction between views can avoid contrastive learning without representational collapse, DINO [32] which generalizes distillation, MAE [33] which restores reconstruction of masked observations as a competitive objective, and finally I-JEPA [27] and V-JEPA [28] which formalize latent prediction over context-target pairs.

The specificity of JEPA with respect to BYOL is not the idea of a target predicted in latent space (BYOL already carried it), but the fact that the target is spatially situated via a position signal, and that the context is explicitly masked rather than defined by augmentation. This detail changes the nature of what is learned: no longer an invariance to imposed transformations, but a predictability conditioned on a localization.

The principle formalized by Yann LeCun in his 2022 program [16] on *autonomous machine intelligence* can be stated briefly. An encoder produces a representation of a context. A second encoder produces a representation of a partial target. A predictor, operating entirely within latent space and conditioned on the position of the target, predicts the representation of the target from that of the context. The learning objective is defined on the similarity between prediction and target in this latent space, and not on pixel-to-pixel reconstruction of the masked observation.

This detail is conceptually decisive. Autoencoders and diffusion models learn to reconstruct raw observations: they are constrained to represent everything in the signal, including noise, irrelevant details, surface variations that hold no structural value. A JEPA does not seek to reconstruct the raw observation. It predicts in a learned representation space, where informationally useless details have been, at least in principle, eliminated.

A crucial technical precision must be made here, otherwise the analysis transforms into ungrounded praise:

> A latent predictability objective alone converges toward the degenerate solution where all representations collapse into a single point: *collapse*.
>
> The properties of invariance and compression invoked further on are guaranteed only by explicit anti-collapse mechanisms: variance and covariance regularization (VICReg [34]), target encoder updated by exponential moving average (EMA), architectural asymmetry between context encoder and target encoder, or combinations of these mechanisms.
>
> What JEPA learns is defined by the objective combined with these structural constraints, not by the objective alone.

The predictive virtues of the architecture are therefore inseparable from inductive engineering choices that must be documented as such.

> *The model does not learn to see. It learns what is predictable.*

This formulation translates a precise architectural choice: the learning objective becomes inter-representational coherence, not observational fidelity. This is not a refinement, it is a change of target.

5. What JEPA actually learns: predictability constraints

This modification of target has a consequence often insufficiently made explicit.

The latent space learned by a JEPA is not a feature space in the classical sense, that is, a dictionary of visual or semantic patterns useful for downstream tasks. It is a *coherence space*: a geometry in which certain configurations of representations are compatible with each other and others are not.

A technical precision is required here, because it conditions everything that follows. Standard JEPA, in its I-JEPA and V-JEPA formulations, does not explicitly encode the transformations of the world. It is not a dynamical model in the sense of a system simulating state trajectories. What learning by latent prediction over context-target pairs produces is a mapping function between latent representations, constrained such that pairs corresponding to natural co-occurrences in the data have mutually predictable representations. The dynamics of the world is not represented as such; it is implicitly constrained by the predictability structure of the latent space.

> *JEPA does not encode transformations themselves. It encodes the constraints that make certain transformations predictable.*

The nuance is not cosmetic. It avoids the easy attribution of an explicit dynamical property to a system that remains, in its standard version, formally static: a mapping function from latent to latent. And it correctly situates the relation of JEPA to explicit world models such as DreamerV3 [19] or to the active inference models of Friston [20]: the latter explicitly represent iterative dynamics; JEPA, for its part, constrains a space within which certain dynamics become predictable without being simulated.

This characterization applies to standard JEPA. LeCun's original roadmap proposes explicitly dynamical extensions, hierarchical (H-JEPA) or action-conditioned (A-JEPA), which would shift part of the analysis presented here. These variants remain to date largely at the program stage, with a few partial implementations. They are not the subject of the present article, but their existence forbids freezing the *static* characterization of JEPA as a definitional property of the entire architectural family.

The position these precisions allow us to hold can be formulated in a single sentence, which serves as the gravitational center of the entire article. *JEPA is neither a memory, nor a simulator, nor an agent: it is an architecture that learns a geometry of predictability.*

6. The long-memory analogy: critique, and positive reconstruction

An analogy circulates regularly in the literature and in presentations of these architectures. The latent space of a JEPA is said to function as a form of long-term memory, even as an approximate analogue of proprioception, in that it maintains an internal coherence stable across transformations of the input signal.

The analogy captures a partial intuition. There is indeed, in a well-trained JEPA, a form of representational continuity: transformations of the input signal that do not modify the underlying structure (partial occlusion, noise, minor geometric transformations) do not perturb the representation. This internal stability, independent of surface fluctuations of the signal, presents an obvious structural kinship with the perceptual invariances described by cognitive psychology.

But the biological analogy misses what is essential. Biological proprioception is not mere stability of representation; it is the continuous emission, by the body, of an internal signal that informs the central nervous system of the effective state of the motor system. It is embodied in the strong sense: there is a body, real receptors, a closed sensorimotor loop. No current JEPA possesses anything of the kind. Its *coherence* is purely representational; it is anchored in no physiological substrate and in no motor action.

The critique of the analogy does not, however, suffice. What remains is to say what JEPA does positively, without the biological crutch. Three properties deserve to be named in their own right.

1. First, a **structural invariance under partial transformation**. The learned representations are stable under a class of perturbations of the input signal corresponding to transformations that preserve the predictable structure of the data. This invariance is not posited a priori; it emerges from the learning objective, modulo the anti-collapse constraints recalled in section 4.

2. Then, a **compression oriented toward predictability**. The latent space privileges information that contributes to inter-representational prediction, and marginalizes information that does not. This is a form of informational filtering by predictive utility, distinct from filtering by information loss in classical autoencoders.

3. Finally, a **selection of information constrained by predictability**. Dimensions of the input signal that bring no constraint to the context-target pairs tend not to be preserved in the representation, not by attentional decision of an agent, but by construction of the learning objective. What is not constrained by predictability has no reason to be stabilized in the representation.

These three properties constitute a distinctive epistemic profile, which has no need of the biological analogy to be characterized.

7. From observation to constrained latent continuation

The shift performed by JEPA can then be formulated in paradigmatic terms. The classical paradigm of supervised learning articulates three operations: observe, abstract, classify. A sample enters, a category exits. The JEPA paradigm articulates differently: observe, constrain, anticipate. A context enters, a space of plausible targets is defined.

The consequence is precise: the model constrains a space of plausible latent continuations, without making them explicit. It does not generate complete trajectories, it does not produce photorealistic images, it does not unfold iterative simulations. But its latent space structure makes certain transitions predictable and others not, which amounts to implicitly delimiting a space of admissible continuations. The distinction with an explicit simulator is crucial: a dynamical world model generates trajectories; JEPA delimits the space within which these trajectories should be generable by an appropriate system.

This characterization joins, by an entirely different path, the literature on predictive processing (Friston, Clark): a predictive brain is not a brain that generates hallucinations, it is a brain that anticipates prediction error and minimizes its free surprise. The analogy has its limits, active inference presupposes a sensorimotor loop that a standard JEPA does not have, but it correctly situates the type of computational object being constructed. Not a classifier, not a generator, not an agent. A system that constrains a space.

8. Labels: an arbitrary projection into a space optimized on other criteria

A practical consequence of this shift concerns the status of supervised annotations. The dominant paradigm presents them as ground truth. This presentation contains an equivocation worth dispelling.

> *Labels do not describe the world. They constrain its use.*

The label *malignant tumor* affixed to a radiological image does not describe an intrinsic property of the image. It indicates the expected clinical use of this image within a given decision framework. The disease is not in the pixel; it is in the articulation between the pixel, the patient's history, the diagnostic protocol and the therapeutic decision. The label compresses this articulation into a binary signal useful for supervised learning, but this compression is a domain-specific projection, not an ontological description.

The tension with JEPA is then explicit and too rarely formulated. A JEPA learns a geometry of coherence independently of any domain-specific partition. The latent space it constructs is optimized on an internal criterion, inter-representational predictability, which has no structural reason to align with the taxonomic boundaries of a particular clinical use. When a supervised head is added to a pre-trained SSL encoder, learning is not being completed: an arbitrary projection dictated by the needs of a downstream task is being reinjected into a space optimized on other criteria.

This reinjection is legitimate. It is even indispensable for operational uses. But it is not neutral, and it must not be confused with a revelation of the structures the encoder would have discovered: the encoder discovered its structures, and the supervised head projects these structures onto the boundaries of a given task. Hence a strategic partition: self-supervision is a mechanism of structure discovery; supervision is a mechanism of use projection. Both are necessary. They do not do the same thing.

9. Implications in regulated environments: healthcare

This distinction is not speculative. It has concrete technical consequences for AI systems in regulated environments, and particularly in healthcare. Three properties take on particular importance there, which must be formulated as engineering hypotheses rather than as intrinsic properties of the architecture, but which now have a substantial body of empirical evidence.

1. **First hypothesis: out-of-distribution robustness**. A supervised classifier trained on a European hospital cohort, deployed on a North American cohort, sees its performance degrade all the more as its representations were optimized for correlations specific to the training site. The operational hypothesis is that an SSL encoder pre-trained on a broad distribution, then projected by supervised fine-tuning, generally degrades less. This hypothesis has received empirical support in medical imaging notably through the work of Azizi and colleagues on SSL for medical image classification [38], CheXzero-type architectures [41] on chest radiography, and more recently RETFound [39] for ophthalmological imaging. It nonetheless remains to be established case by case: no theoretical guarantee imposes it, and the performance gap depends strongly on the distributional distance between training site and deployment site.

2. **Second hypothesis: dependence on annotated datasets**. The HDLSS situation (High Dimension, Low Sample Size), endemic in biomedicine, severely penalizes pure supervision. Self-supervised pretraining on unannotated corpora, when technically feasible, can displace part of the learning complexity outside the phase costly in medical annotation. This possibility depends on the availability of a pretraining corpus structurally comparable to the application domain. The review by Krishnan and colleagues [40] on SSL in healthcare documents both the promise and the restrictive conditions of this approach: for rare specialized cohorts (for example precision oncology on molecular subpopulations), the pretraining corpus often does not exist in the target domain, and the use of transfers from RadImageNet [42] or other generic references reintroduces domain biases that final supervision must address.

3. **Third hypothesis: trajectory modeling**. Medicine is essentially temporal. An architecture that learns predictability constraints over temporal pairs can produce representations useful for modeling disease or treatment trajectories. Here again, this is an engineering hypothesis, whose empirical verification for specialized cohorts remains to be completed.

None of the three guarantees its own success. All three orient a reasonable engineering strategy when the conditions are met.

> ***Illustrative box.*** *In the TweenMe / OCTOPUS program on mNSCLC patients carrying the BRAF V600E mutation (n=184, 5 European countries), the work on trajectories led to mobilizing a combination of representation learning and modeling by SurvTRACE [43], a transformer architecture for survival analysis in the presence of competing events, with a TSTR fidelity measured at 95.2 percent on the validation cohort, evaluated against a classifier baseline trained on real data. This metric is not a proof of general statistical indistinguishability, nor a demonstration of the intrinsic superiority of the learned representations. It is an indicator, within the considered evaluation framework, that the generated trajectory preserves the operational properties useful for downstream tasks. Implementation terrain, not universal demonstration. The methodological detail is the subject of a separate publication.*

10. Limits: not turning JEPA into a religion

Several authentic limits must be firmly held, otherwise we slide into the evangelical register that haunts every new architecture.

1. First, there exists to date no proof that a JEPA learns a *complete physics* of the world. The existing demonstrations (I-JEPA on images [27], V-JEPA on videos [28]) show invariances learned on natural distributions, but these invariances cover only a fraction of real physical constraints, and generalization to regimes very far from the training distribution remains to be demonstrated.

2. Then, the learned latent space is, in the great majority of configurations, non-interpretable. This limit is not specific to JEPA; it is shared by the whole of SSL. But it weighs particularly in regulated contexts where feature traceability is required: a software medical device falling under the MDR Regulation (EU) 2017/745 and, where applicable, under the high-risk AI systems regime of the European AI Regulation, must satisfy cumulative requirements of transparency, robustness and human oversight. The non-interpretability of the latent space must then be compensated by other guarantees: post-hoc explainability, drift monitoring, independent validation by external cohort, extensive technical documentation.

3. Third, the evaluation of these architectures is technically delicate. Classical metrics (precision, AUC) do not directly measure what JEPA is supposed to learn. Linear probes on downstream tasks give an indication, not a direct measure of the quality of the modeled world.

4. Fourth, performance depends strongly on the design of masks and on context-target pair generation strategies. What sometimes resembles an automatic discovery is in reality partially encoded in inductive engineering choices. Legitimate choices, but ones that must be documented as such.

5. Fifth, SSL pretraining requires substantial compute. The argument of reducing dependence on annotations in section 9 must be read alongside the fact that this reduction is paid for in GPU cycles on massive pretraining corpora, which transfers part of the cost rather than eliminating it. The economic balance depends on the ratio between medical annotation cost and compute cost, which evolves rapidly.

6. Finally, temporal drift. An SSL encoder pretrained in 2025 on a corpus distributionally characteristic of that period has no guarantee of remaining valid in 2030, when imaging protocols, cohort demographics or acquisition modalities will have evolved. This drift is documentable and requires a surveillance apparatus; it is not absent from the landscape simply because the encoder was pre-trained once.

JEPA shifts the problem of learning. It does not entirely resolve it.

11. The threshold not crossed: three operational conditions, and the architectures that claim to satisfy them

There remains the question articulated in Part 3/3, and which constitutes the most important point of vigilance. What does JEPA bring, or more generally world modeling architectures and agentic architectures with external memory, to the problem of biographical memory as defined in section 1?

Let us hold to the strictly functional formulation. Three operational conditions distinguish a biographical memory from a latent coherence.

1. **First condition: contextual reindexing**. A biographical memory permits access to a mnesic content along several entry paths (modal, temporal, affective) and reactivation, from each of them, of a coherent configuration of the entire episode. A JEPA produces representations stable under transformation, but this stability is defined on a single axis: inter-representational predictability. There is no multi-perspective indexation structure in the latent space.

2. **Second condition: multi-episode integration**. A biographical memory articulates distinct episodes between themselves through relations that are neither purely statistical nor purely temporal, but structured by an agent's history. A JEPA learns regularities across the entire pretraining corpus, without

differentiated preservation of individual episodes.

3. **Third condition: agent persistence**. This condition demands an operational definition, otherwise it remains a formula. By *agent persistence*, we mean the continuity in time of a unique referent to which mnesic edges are indexed, with the additional property that this agent can treat past episodes as episodes *lived by itself*, and not as external data available for consultation. The distinction between indexation and ownership is crucial. A journal indexes events to an identifier; it does not make them belong to a subject.

A serious objection deserves to be examined head-on. Several families of contemporary agentic architectures claim precisely what the three conditions appear to describe.

The *generative agents* of Park and colleagues [35] equip LLMs with a memory stream (flux of indexed observations), a reflexion mechanism (periodic synthesis into meta-memories) and a retrieval system combining similarity, recency and importance.

Voyager [36] endows a Minecraft agent with a persistent skill library and an automatic curriculum.

ReAct [37] and its extensions interleave reasoning, action and simple episodic memory.

These architectures are serious and cannot be dismissed by a formula. Let us examine them in light of the three conditions, without complacency:

On **contextual reindexing**, generative agents do offer a multi-criteria retrieval (semantic similarity, temporal recency, weighted importance). It is a form of multi-axis indexation, but one that operates on homogeneous textual entries; it does not reactivate a multimodal configuration of an episode, it composes a prompt from selected textual fragments. The distinction is operational: contextual reindexing in the strong sense reactivates the episode; the memory stream re-articulates it. Voyager has no reindexation in the episodic sense, its skills are indexed by functionality.

On **multi-episode integration**, generative agents have a dedicated mechanism, reflexion, which periodically synthesizes the memory stream into higher-level propositions. It is an integration, but it is compressive and lossy: it produces summaries, not relations preserving the individuality of episodes. Voyager integrates acquired skills, not episodes. ReAct integrates nothing beyond the current window.

On **agent persistence**, this is the condition that distinguishes them most sharply from biographical memory. These architectures all have a persistent identifier and a journal of episodes attached to this identifier. They do not, however, satisfy the ownership condition: the agent does not treat the memory stream as episodes lived by itself, it consults it as a database indexed to its identifier.

Retrieval is an indexing operation, not a situated reactivation.

This distinction is not philosophical coquetry: it has verifiable functional consequences, notably on the capacity to modify prior commitments in coherence with a personal trajectory rather than by selection of fragments compatible with the current query.

None of these architectures therefore satisfies the three conditions simultaneously in the strict sense. They approach them, sometimes strikingly, but they operate by juxtaposition: an encoder, a journal, a retrieval mechanism, an identified agent. The question is not whether one can represent a biography by juxtaposing these elements, but whether the juxtaposition produces the property of ownership, or only its useful functional simulacrum.

A philosophical precision is required at this point, so as not to leave ambiguity with the position defended in Part 3/3 of the series. The latter argued that human biographical memory is distinguished by irreducibly phenomenological properties: autonoetic consciousness in Tulving's sense, affective modulation, multi-perspective navigability. The translation into three operational conditions performed here does not

abolish this defense. It isolates an operational minimum below which one can affirm that the threshold is not crossed, independently of any phenomenological commitment. If an architecture does not satisfy contextual reindexing, multi-episode integration and agent persistence in the strong sense, it does not cross the threshold, whatever the metaphysical arbitrations on consciousness. If it does satisfy them, the phenomenological question of autonoetic consciousness remains open, as a structural surplus above the operational minimum. This stratification allows holding a defensible position without engaging the phenomenological quarrel at every architectural evaluation. It has a cost: the strong phenomenological position of Part 3/3 ceases to be necessary in order to distinguish JEPA from biographical memory. It becomes necessary again only in the zone where the operational minimum would be satisfied, a zone that current architectures do not reach.

To date, and to my knowledge, no published architecture simultaneously satisfies these three operational conditions in the strict sense defined here. This finding is neither a strong phenomenological thesis, nor an argument of inaccessibility in principle. It is an architectural description, susceptible to revision by the next publication that explicitly demonstrates satisfaction of the three conditions, and not their simulacrum by juxtaposition.

12. Conclusion: intelligence and invariance

What latent predictive architectures change is not the nature of artificial intelligence. It is the target of learning. Before them: learn answers, classify, reconstruct. After them: learn the constraints that make certain transformations predictable, encode the coherence of a space rather than fidelity to a signal.

This shift is neither a revolution, nor a detail. It is a precise epistemic movement, whose scope must be assessed at the level of what it does, namely reduce dependence on annotations in certain regimes, improve out-of-distribution robustness under certain conditions, structure trajectory modeling through predictability constraints, and of what it does not do: satisfy the functional conditions of a biographical memory, neither by itself, nor by simple addition of an episode journal.

JEPA is neither a memory, nor a simulator, nor an agent: it is an architecture that learns a geometry of predictability. Holding this formula means accepting not to project onto these systems properties they do not have, and recognizing those they effectively do.

The strategic question for industrial architects deploying these systems in regulated environments is therefore not *should we adopt JEPA?* The question is poorly posed. It is: which property are we trying to instantiate, at what level, and does the chosen architecture instantiate it, or does it merely simulate its surface? Both answers are valid depending on context, but they are not equivalent, and their confusion produces systems that appear intelligent up until the exact moment one moves them out of their training distribution.

Intelligence does not reside in what is observed, but in what cannot vary. It remains to be seen whether what cannot vary is enough to constitute a subject who remembers.

Notes and additional references

Numbering [1] to [26] follows that of Parts 1/3, 2/3 and 3/3. See: article 1, article 2, article 3.

[27] Assran, M. et al. (2023). Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture (I-JEPA). CVPR 2023.

[28] Bardes, A., Garrido, Q., Ponce, J., Chen, X., Rabbat, M., LeCun, Y., Assran, M., Ballas, N. (2024). Revisiting Feature Prediction for Learning Visual Representations from Video (V-JEPA). arXiv:2404.08471.

[29] Chen, T. et al. (2020). A Simple Framework for Contrastive Learning of Visual Representations (SimCLR). ICML 2020.

[30] He, K. et al. (2020). Momentum Contrast for Unsupervised Visual Representation Learning (MoCo). CVPR 2020.

[31] Grill, J.-B. et al. (2020). Bootstrap Your Own Latent (BYOL). NeurIPS 2020.

[32] Caron, M. et al. (2021). Emerging Properties in Self-Supervised Vision Transformers (DINO). ICCV 2021.

[33] He, K. et al. (2022). Masked Autoencoders Are Scalable Vision Learners (MAE). CVPR 2022.

[34] Bardes, A., Ponce, J., LeCun, Y. (2022). VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. ICLR 2022.

[35] Park, J. S. et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. UIST 2023.

[36] Wang, G. et al. (2023). Voyager: An Open-Ended Embodied Agent with Large Language Models. arXiv:2305.16291.

[37] Yao, S. et al. (2023). ReAct: Synergizing Reasoning and Acting in Language Models. ICLR 2023.

[38] Azizi, S. et al. (2022). Big Self-Supervised Models Advance Medical Image Classification. Nature Biomedical Engineering.

[39] Zhou, Y., Chia, M. A., Wagner, S. K. et al. (2023). A foundation model for generalizable disease detection from retinal images (RETFound). Nature, 622(7981), 156-163.

[40] Krishnan, R., Rajpurkar, P., Topol, E. J. (2022). Self-Supervised Learning in Medicine and Healthcare. Nature Biomedical Engineering, 6(12).

[41] Tiu, E. et al. (2022). Expert-Level Detection of Pathologies from Unannotated Chest X-Ray Images via Self-Supervised Learning (CheXzero). Nature Biomedical Engineering.

[42] Mei, X. et al. (2022). RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. Radiology: Artificial Intelligence.

[43] Wang, Z., Sun, J. (2022). SurvTRACE: Transformers for Survival Analysis with Competing Events. ACM-BCB 2022 (arXiv:2110.00855).

This article constitutes the fourth volume of the series Encoding, transduction and world models. The three previous volumes (Parts 1/3, 2/3 and 3/3, March 2026) are available on twingital-ventures.com.

https://twingital-ventures.com/fr/publications/encodage-transduction-modeles-du-monde-1/3 (FR, summarized English version available)

https://twingital-ventures.com/fr/publications/encodage-transduction-modeles-du-monde-2/ (FR, summarized English version available)

https://twingital-ventures.com/fr/publications/encodage-transduction-modeles-du-monde-3/ (FR, summarized English version available)