



Encodage, transduction et modèles du monde : Pourquoi toute architecture d'IA commence par transformer le monde

16 mars 2026

Partie 2/3 : Chapitre 3 à 5

Fait suite à la partie 1/3 : Chapitres 1 à 2 ([Article 1, partie 1/3](#))

3. L'asymétrie fondamentale : transduction biologique et encodage médiatisé

3.1 La transduction biologique : un encodage contraint par la physique et l'évolution

Dans les systèmes biologiques, l'encodage des informations issues de l'environnement est assuré par des transducteurs sensoriels, des structures anatomiques spécialisées capables de convertir certaines formes d'énergie physique (photons, vibrations mécaniques, gradients chimiques, pression, accélérations) en signaux électrochimiques exploitables par le système nerveux.

L'oreille interne constitue un exemple paradigmatique. Les cellules ciliées de la cochlée convertissent des vibrations mécaniques en potentiels d'action par la déflexion des stéréocils et l'ouverture de canaux ioniques mécanosensibles. Les propriétés physiques du stimulus sonore, notamment sa fréquence, sont ainsi traduites en organisation spatiale de l'activité neuronale le long de la membrane basilaire selon le principe de tonotopie. À ces informations auditives s'ajoutent celles du système vestibulaire, qui encode les accélérations angulaires et linéaires ainsi que l'orientation gravitationnelle.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Ces signaux sont intégrés avec les informations visuelles et proprioceptives afin de stabiliser la posture et coordonner l'action en temps réel[5].

Ces mécanismes ne doivent toutefois pas être compris comme un accès transparent au monde. Bien avant d'atteindre les aires corticales, le signal a déjà subi de multiples transformations : filtrage rétinien, inhibition latérale, normalisation, sélection, compression temporelle et spatiale. Les systèmes sensoriels biologiques effectuent donc eux aussi une transformation substantielle du signal. La différence essentielle ne réside pas dans l'existence d'un encodage, tout système cognitif encode, mais dans l'origine des contraintes qui structurent cet encodage. Dans les organismes vivants, ces contraintes résultent de régularités biophysiques, de pressions évolutives et d'interactions écologiques prolongées avec un environnement physique. Les transducteurs sensoriels ont été façonnés par l'histoire phylogénétique de l'espèce et par les propriétés statistiques des niches écologiques au sein desquelles elle s'est développée.

Dans ce cadre, perception et action forment des boucles dynamiques continues. La perception ne consiste pas seulement à produire une description interne du monde, mais à détecter des possibilités d'action pertinentes dans l'environnement. Dans le vocabulaire de James Gibson, les systèmes perceptifs sont couplés aux *affordances*[6] de l'environnement, les opportunités d'action qu'un environnement donné offre à un organisme donné. Il convient de préciser que Gibson défend une position explicitement anti-représentationnaliste : pour lui, les *affordances* sont directement perçues sans médiation représentationnelle. L'usage qui en est fait ici s'inscrit dans une lecture représentationnaliste minimale compatible avec l'argument général de cet article : les *affordances* sont comprises comme des propriétés relationnelles émergeant du couplage dynamique entre organisme et environnement. Ce type de couplage perception-action constitue précisément ce que les modèles de langage contemporains, dépourvus de boucle sensorimotrice fermée, ne peuvent instancier.

3.2 L'encodage artificiel : une chaîne de médiations représentationnelles

Dans les systèmes d'intelligence artificielle contemporains, et en particulier dans les modèles fondés sur des corpus symboliques tels que les grands modèles de langage, l'encodage des données obéit à une logique profondément différente. Avant que le modèle ne puisse traiter la moindre information, les données ont déjà traversé une chaîne de médiations humaines impliquant ingénieurs de données, data scientists et annotateurs, dont les principales opérations ont été décrites au §2.3. (voir article précédent) Ces transformations n'ont pas pour fonction de reproduire directement la structure physique du monde, mais de produire une représentation exploitable par une

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

architecture d'apprentissage statistique. Le modèle n'est donc pas confronté aux propriétés physiques de l'environnement, mais à des artefacts informationnels déjà produits par l'activité cognitive humaine.

Il serait toutefois inexact d'affirmer que les systèmes d'IA sont dépourvus de contraintes. Ils sont au contraire fortement contraints par plusieurs facteurs : la structure de leur architecture, la forme de la fonction de perte qui oriente l'apprentissage, la distribution statistique des données d'entraînement, les biais inductifs introduits par l'architecture elle-même, ainsi que les capacités de calcul disponibles. Ces contraintes jouent un rôle déterminant dans la formation des représentations internes. Cependant, elles diffèrent en nature des contraintes biologiques. Les contraintes biologiques sont le produit de processus évolutifs et d'interactions écologiques prolongées avec un monde physique ; les contraintes des systèmes artificiels sont définies par des ingénieurs et optimisées pour des objectifs de performance mesurables sur des ensembles de données donnés, le plus souvent sans interaction directe, continue et ouverte avec l'environnement auquel ces représentations se rapportent.

Dans le cas des modèles de langage, la médiation est particulièrement marquée. Les données d'entraînement sont constituées de textes qui représentent déjà des descriptions, interprétations et conceptualisations humaines du monde. La chaîne de transformation peut être schématisée de la manière suivante :

Biologique : *monde physique* → *transduction sensorielle* → *signal neural* → *cognition*

LLM : *monde physique* → *expérience humaine* → *description symbolique* → *corpus* → *tokenisation* → *modèle*

Le système d'IA opère ainsi sur des représentations déjà stabilisées dans des artefacts sémiotiques. Le signal traité par le modèle n'est pas un flux énergétique issu directement de l'environnement, mais un objet symbolique produit par des agents humains. En ce sens, on peut parler d'un *encodage de second degré* : dans les systèmes biologiques, l'encodage correspond à une transformation de signaux physiques issus de l'environnement ; dans les systèmes d'IA fondés sur des corpus symboliques, l'apprentissage porte sur des représentations humaines du monde déjà constituées. Le modèle apprend à modéliser les structures discursives par lesquelles les humains décrivent, interprètent et organisent le monde, plutôt que les propriétés physiques du monde lui-même.

Cette distinction ne doit cependant pas être absolutisée. La cognition humaine adulte elle-même est traversée de médiations culturelles, linguistiques, sociales et

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

institutionnelles. Un individu humain n'accède pas au monde uniquement par perception directe : il hérite aussi d'un monde déjà structuré par des récits, des catégories, des instruments, des pratiques et des institutions. La différence ici défendue ne réside donc pas dans l'opposition simpliste entre une cognition biologique supposément immédiate et une cognition artificielle entièrement médiée, mais dans la nature et dans l'empilement des régimes de médiation qui président à la formation des représentations.

3.3 Distance épistémique au monde

Cette différence peut être formulée plus généralement en termes de *distance épistémique au monde*. Par distance épistémique, on entend ici le nombre et la nature des médiations informationnelles séparant un état interne du système des propriétés physiques de l'environnement. Dans les systèmes biologiques, cette chaîne est relativement courte : certaines propriétés énergétiques du monde sont directement transduites en activité neuronale, puis intégrées dans des boucles perception-action contraintes par la physiologie et l'évolution. Dans les systèmes d'IA corpus-based, cette chaîne est à la fois plus longue et qualitativement différente : le signal a déjà été collecté, sélectionné, filtré, structuré et représenté par des agents humains avant d'être introduit dans le système d'apprentissage. Surtout, la nature du signal change en cours de route — il passe d'un régime physique à un régime symbolique.

Les systèmes d'IA dominants, lorsqu'ils sont entraînés principalement sur des corpus textuels, apprennent donc avant tout à partir de représentations symboliques du monde, et non à partir d'une interaction directe avec celui-ci.

Cette distance épistémique accrue ne constitue pas nécessairement une limitation absolue : elle permet au contraire d'exploiter l'immense accumulation de connaissances symboliques produites par les sociétés humaines, ce qui explique l'efficacité remarquable de ces modèles sur de nombreuses tâches propositionnelles. Mais elle implique que les représentations internes des systèmes artificiels sont formées dans un régime informationnel profondément différent de celui des organismes biologiques. Cette asymétrie constitue le premier niveau de la différence architecturale que cet article cherche à cartographier. Comme on le verra dans les sections suivantes, deux niveaux supplémentaires demeurent ouverts même pour un système artificiel équipé de transducteurs physiques : un grounding multimodal fondé sur la coordination vécue des modalités perceptives, et un grounding autobiographique et affectif reposant sur une organisation biographique de la mémoire.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Tableau 1 — Comparaison des modes d'encodage biologique et artificiel

Dimension	Encodage biologique (transduction)	Encodage artificiel (LLMs)
Source de données	Expérience sensorimotrice directe	Texte produit et filtré par des humains
Mode d'encodage	Transductif — ancré dans la physique	Symbolique/statistique, médiatisé
Représentation de la causalité	Modèle causal structurel	Corrélation statistique
Temporalité	Dynamique continue, incarnée	Séquence de tokens
Grounding	Ancré dans l'action et la perception	Absent — symboles autoreférentiels
Incarnation	Constitutive	Absente des architectures actuelles

Sources : Gibson (1979), Harnad (1990), LeCun (2022) + synthèse de l'auteur.

4. Le symbol grounding problem et les formes d'aveuglement représentationnel

4.1 Le symbol grounding problem

L'asymétrie décrite dans la section précédente rejoint directement le *symbol grounding problem* formulé par Stevan Harnad en 1990[7]. Dans un système purement formel, les symboles ne renvoient initialement qu'à d'autres symboles : leur signification est définie par des relations internes au système et non par un ancrage direct dans le monde physique. Un dictionnaire fournit une illustration simple de ce phénomène : chaque mot y est défini par d'autres mots, et la signification globale ne peut émerger que si certains symboles sont finalement reliés à une expérience non symbolique. Cette situation est une conséquence directe de la distance épistémique introduite à la section précédente : le système apprend non pas à partir d'interactions physiques avec l'environnement, mais à partir de représentations symboliques produites par des agents humains ayant eux-mêmes expérimenté le monde.

Les grands modèles de langage se trouvent précisément dans cette situation. Ils apprennent des structures linguistiques extrêmement riches à partir de corpus textuels massifs. Mais ces symboles renvoient principalement à d'autres symboles produits par l'activité cognitive humaine. Le modèle construit ainsi un espace relationnel dense entre concepts sans disposer, en régime purement textuel, d'un accès direct aux phénomènes auxquels ces concepts se rapportent. Sa représentation du monde est dérivée de descriptions linguistiques produites par des agents humains qui, eux, ont interagi avec le monde.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Il convient de préciser que, chez Harnad, le grounding ne se réduit pas à une simple association entre un symbole et une mesure physique brute. Ce qui est en jeu est l'ancrage des symboles dans des catégories perceptuelles apprises, issues d'interactions sensorimotrices permettant de discriminer de manière robuste des entités ou des propriétés du monde. Le problème n'est donc pas seulement l'absence de contact avec le réel, mais l'absence d'un processus par lequel les symboles seraient reliés à des catégories acquises sur la base d'une expérience perceptive.

4.2 Aveuglement perceptif : le problème de Molyneux et l'argument de Mary

La distinction entre connaissance conceptuelle et expérience perceptive apparaît clairement dans deux traditions philosophiques distinctes.

Le *problème de Molyneux*, formulé en 1688, demandait si un aveugle de naissance capable de distinguer par le toucher un cube d'une sphère pourrait reconnaître ces objets par la vision s'il recouvrait soudainement la vue. Les recherches empiriques menées par Richard Held et ses collègues, publiées en 2011 dans *Nature Neuroscience*[8], ont examiné des patients ayant recouvré la vue après une cataracte congénitale. Les résultats indiquent que les sujets ne peuvent pas immédiatement identifier visuellement des formes qu'ils reconnaissaient auparavant par le toucher. La correspondance entre les modalités perceptives doit être progressivement apprise par l'expérience. Ces travaux suggèrent que la connaissance conceptuelle d'un objet et l'expérience perceptive de cet objet constituent deux types de représentations fonctionnellement distincts.

L'argument de Mary, proposé par Frank Jackson en 1982[9], radicalise cette distinction. Mary est une scientifique spécialiste de la neurophysiologie de la vision des couleurs qui a vécu toute sa vie dans un environnement monochromatique. Elle connaît toutes les lois physiques et neurobiologiques impliquées dans la perception du rouge. Pourtant, lorsqu'elle voit une surface rouge pour la première fois, il semble qu'elle acquière une forme de connaissance nouvelle : la connaissance phénoménologique de ce que signifie voir du rouge, ce que Jackson nomme le *qualia* correspondant. Cet argument demeure débattu dans la philosophie contemporaine de l'esprit (les objections fonctionnalistes de Lewis, Dennett et Levin en contestent les prémisses) mais il met en évidence une distinction qui reste pertinente pour la discussion présente : celle qui sépare la connaissance propositionnelle d'un phénomène de son expérience perceptive.

Une objection classique consiste à remarquer qu'un individu peut comprendre la douleur d'autrui sans l'éprouver lui-même. Cette compréhension relève toutefois d'une inférence empathique ancrée dans des douleurs déjà vécues et dans l'observation de

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

réactions humaines : elle ne donne pas accès à la qualité phénoménologique précise de cette douleur ni à son intensité vécue, mais produit une modélisation conceptuelle plausible du phénomène.

Appliqué aux modèles de langage, le parallèle devient clair. Un LLM entraîné sur des milliards de descriptions textuelles du rouge peut en décrire la physique, la neurophysiologie et les associations culturelles. Il peut décrire la douleur et ses effets comportementaux. Mais, en régime purement textuel, il ne dispose d'aucune expérience perceptive correspondante (nociception + mémoire de l'expérience douloureuse). Sa représentation demeure dérivée de descriptions linguistiques produites par des agents humains.

4.3 La théorisation scientifique ne s'affranchit pas de l'ancrage empirique

On pourrait toutefois objecter que l'histoire des sciences montre que des structures du monde peuvent être inférées sans perception directe. Au Ve siècle avant notre ère, Démocrite propose que la matière soit constituée d'unités indivisibles alors qu'aucun instrument ne permet d'observer des structures microscopiques. De même, la théorie de la relativité restreinte formulée par Einstein en 1905 mobilise largement des expériences de pensée (la poursuite d'un rayon lumineux ou la comparaison d'horloges en mouvement,...) avant que certaines de ses prédictions ne soient confirmées expérimentalement. Ces exemples pourraient suggérer que la connaissance scientifique peut émerger indépendamment de toute perception directe.

Une analyse plus attentive montre cependant que les expériences de pensée ne remplacent pas l'expérience empirique : elles réorganisent conceptuellement des régularités déjà observées. L'atomisme antique s'inscrit dans une réflexion sur la divisibilité des substances et sur des phénomènes macroscopiques observables. La relativité restreinte se développe dans un contexte structuré par l'électromagnétisme de Maxwell et par les mesures de Michelson et Morley (1887). Les théories scientifiques émergent au sein de ce que l'on peut appeler un état socio-technique des sciences, un environnement constitué d'instruments, de méthodes expérimentales, de résultats accumulés et de cadres conceptuels partagés. La créativité scientifique consiste moins à produire des structures symboliques indépendamment du monde qu'à inférer des structures invisibles à partir de régularités observables.

4.4 Gradation épistémique

Cette analyse permet de situer les modèles de langage sur une échelle de médiation par rapport à l'expérience physique[22]. Un physicien tel qu'Einstein infère des structures théoriques à partir d'un rapport indirect mais réel à l'expérience empirique, aux

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

instruments et aux résultats produits par sa communauté. Un étudiant qui lit ses travaux sans jamais réaliser d'expérience possède déjà un accès plus médié à ces phénomènes : il hérite d'un savoir théorique déjà stabilisé par l'activité scientifique. Un modèle de langage se situe à un niveau supplémentaire de médiation : il n'interagit ni avec le monde physique ni avec les pratiques expérimentales de la science ; il apprend à modéliser les structures discursives produites par les humains qui décrivent ces expériences.

interaction empirique → élaboration théorique → description symbolique → modélisation statistique des textes

Il ne s'agit pas de soutenir que toute médiation éloigne nécessairement de la vérité. Le point est plus précis : à mesure que se multiplient les médiations, le système dépend davantage des formes symboliques déjà construites par d'autres agents pour organiser son rapport au monde. Les modèles de langage héritent ainsi d'un ancrage empirique humain, mais sous une forme seconde, sémiotisée et déjà interprétée.

4.5 Conséquence pour le grounding

La science humaine peut formuler des hypothèses sur des phénomènes invisibles précisément parce qu'elle repose, directement ou indirectement, sur un ancrage empirique préalable dans le monde physique. Les systèmes d'IA fondés sur des corpus symboliques héritent de cet ancrage de manière indirecte, à travers les traces discursives produites par les communautés humaines. Ils peuvent modéliser avec une grande efficacité les relations entre ces représentations. Mais tant que certains symboles ne sont pas reliés à des interactions effectives avec l'environnement, leur signification demeure, au moins en première analyse, interne au réseau symbolique qui les relie. C'est précisément cette absence d'ancrage perceptif direct que met en évidence le symbol grounding problem, et que les différentes formes de grounding examinées dans les sections suivantes permettent de stratifier.

5. Les niveaux du grounding : du capteur à la mémoire autobiographique

5.1 Ce que résoudrait un capteur et ce qu'il ne résoudrait pas

Une objection classique consiste à considérer que la limitation identifiée dans les modèles de langage n'est pas structurelle mais simplement architecturale : il suffirait d'équiper ces systèmes de capteurs physiques pour résoudre le problème du grounding. Un capteur optique capable de mesurer la distribution spectrale de la lumière incidente permettrait par exemple d'associer une mesure autour de 700 nm à la catégorie « rouge ». Le symbole linguistique serait alors relié à une propriété physique mesurable du monde. Cette intuition correspond précisément à celle qui sous-tend le symbol

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

grounding problem formulé par Harnad : éviter que les symboles ne renvoient qu'à d'autres symboles en les ancrant dans des catégories perceptuelles apprises à partir d'interactions sensorimotrices avec l'environnement. Dans cette perspective, l'intégration de capteurs constitue effectivement une condition nécessaire (mais non suffisante) pour dépasser la circularité purement symbolique.

Cependant, l'introduction d'un capteur ne résout qu'une partie du problème. Ce que fournit un capteur est une information sensorielle mesurable, à partir de laquelle un système peut apprendre à discriminer des catégories perceptuelles. Ce qu'il ne fournit pas à lui seul, c'est l'ensemble des structures cognitives associatives qui constituent un concept dans la cognition humaine.

Pour un être humain, la catégorie « rouge » ne correspond pas simplement à une propriété spectrale (et cette propriété spectrale n'est d'ailleurs, en pratique, pas explicitement représentée, ni même -re-connue comme telle dans l'expérience ordinaire). Le rouge est intégré à un vaste réseau d'associations perceptives, motrices, culturelles, sociales et affectives : la texture et l'odeur d'une tomate coupée, le goût d'une sauce bolognaise, le souvenir d'un repas, la tache de sauce sur la chemise, la signification d'un feu rouge dans la circulation, la valence d'une alerte médicale, la charge symbolique de certaines couleurs dans un contexte culturel donné. Ces associations ne résultent pas uniquement de co-occurrences linguistiques ; elles sont stabilisées, enrichies et parfois réorganisées par la co-activation de multiples modalités perceptives et contextuelles au cours de l'expérience vécue.

Il serait toutefois excessif d'opposer ici de manière absolue apprentissage linguistique et expérience incarnée. Chez l'humain, les concepts se forment à l'intersection de plusieurs sources : expérience perceptive, action, langage, transmission sociale, culture et mémoire. Le point défendu ici est plus restreint : l'ajout de capteurs peut ancrer certains symboles dans des propriétés du monde physique, mais il ne reproduit pas automatiquement l'ensemble des mécanismes par lesquels les représentations humaines acquièrent leur richesse multimodale, contextuelle et biographique.

5.2 Une stratification des niveaux de grounding

La littérature sur le grounding est souvent présentée comme si elle désignait un problème unique. L'analyse précédente suggère au contraire qu'il est utile de distinguer plusieurs niveaux conceptuellement distincts, chacun correspondant à une forme particulière de relation entre représentation et expérience.

Le premier niveau correspond au *grounding référentiel ou perceptuel* au sens de Harnad : relier les symboles à des catégories perceptuelles apprises à partir de données

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

sensorimotrices. Les capteurs constituent ici un élément nécessaire, car ils permettent d'ancrer les représentations dans des propriétés mesurables de l'environnement. Ce niveau est partiellement soluble par l'adjonction de capteurs physiques, à condition que ces capteurs soient couplés à un mécanisme d'apprentissage catégoriel.

Le deuxième niveau correspond au *grounding multimodal*. Les concepts humains ne reposent pas sur une modalité sensorielle isolée mais sur l'intégration de multiples modalités. Barsalou (1999)[10] propose la théorie des *perceptual symbol systems*, selon laquelle les concepts sont représentés sous forme de simulateurs multimodaux capables de réactiver partiellement les circuits perceptifs et moteurs mobilisés lors de l'expérience initiale. Un concept comme « tomate » mobilise ainsi simultanément des représentations visuelles, olfactives, gustatives, tactiles et motrices. Ce qui forge ces liaisons n'est pas seulement la co-occurrence des mots dans un corpus, mais la coordination répétée de modalités hétérogènes dans des épisodes d'expérience situés. Ce niveau n'est pas dissous par l'ajout de capteurs pris isolément ; il suppose une histoire d'interactions multimodales suffisamment riches, synchronisées et contextualisées pour stabiliser ce type d'intégration.

Les architectures multimodales contemporaines commencent à explorer une partie de cet espace. Des systèmes vision-langage ou audio-vision peuvent apprendre certaines correspondances cross-modales à partir de données synchronisées, ce qui montre qu'une part du problème est, au moins fonctionnellement, abordable. Mais cette possibilité ne suffit pas à effacer l'écart avec l'intégration multimodale humaine, qui s'inscrit dans une boucle perception–action continue, dans un corps situé et dans une histoire d'apprentissage ouverte.

Le troisième niveau correspond au *grounding épisodique, affectif et auto-noétique*. Il s'agit d'ancrer les représentations dans une mémoire personnelle structurée par des épisodes vécus. Endel Tulving (1983)[12] distingue trois composantes :

- La mémoire sémantique (*savoir que...* connaissances générales décontextualisées),
- La mémoire épisodique (*se souvenir de...* représentation mentale d'événements situés dans le temps et l'espace),
- Et l'*auto-noéticité*, la conscience de soi comme sujet ayant vécu cet événement, capable de se projeter mentalement dans le passé ou dans le futur (Tulving distingue la mémoire sémantique de la mémoire épisodique et associe à cette dernière une forme spécifique de conscience, l'*auto-noéticité*).

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

C'est ce troisième élément qui distingue le niveau 3 du niveau 2 : non seulement un réseau d'associations multimodales, mais un ancrage biographique dans lequel l'expérience est vécue depuis un point de vue subjectif situé.

Cette dimension auto-noétique est renforcée par les marqueurs somatiques décrits par Damasio[13] : des signaux physiologiques attachés aux situations passées qui orientent implicitement les décisions et les associations futures. La tache sur la chemise n'est pas seulement une information perceptuelle, c'est une expérience datée, localisée, chargée d'affect et de subjectivité, qui a une place dans une vie. Ce niveau est structurellement inaccessible à tout système sans continuité biographique et sans auto-noéticité.

Ces distinctions suggèrent que le grounding ne correspond pas à une propriété unique, mais à une hiérarchie de mécanismes cognitifs. L'ajout de capteurs peut résoudre une partie du problème en ancrant certaines représentations dans des propriétés physiques du monde. Toutefois, la richesse des concepts humains repose également sur l'intégration multimodale, sur l'organisation autobiographique de la mémoire et sur des dimensions auto-noétiques et affectives que les architectures actuelles ne reproduisent pas de manière démontrée. Ce constat ne démontre pas une impossibilité de principe pour toute intelligence artificielle future ; il cartographie plutôt une série de seuils architecturaux croissants que l'équipement sensoriel, à lui seul, ne suffit pas à franchir.

Tableau 2 — Stratification des niveaux de grounding

Niveau	Nature	Contribution des capteurs	Références
1 — Grounding référentiel / perceptuel	Relier les catégories à des données sensorimotrices mesurables	Capteurs nécessaires mais non suffisants	Harnad (1990)
2 — Grounding multimodal	Intégration de multiples modalités via co-expérience incarnée	Capteurs utiles, nécessitent une expérience multimodale riche et située	Barsalou (1999)
3 — Grounding épisodique, affectif et auto-noétique	Ancrage dans une mémoire autobiographique, affective et à perspective subjective	Non résolu par capteurs seuls ; requiert architecture de mémoire et auto-noéticité	Tulving (1983), Damasio (1994)

Sources : Harnad (1990), Barsalou (1999), Tulving (1983), Damasio (1994) + synthèse de l'auteur.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com

Tableau 3 — Ce que le capteur résout et ne résout pas

Niveau de représentation	LLM textuel	Système multimodal avec capteurs	Cognition humaine
Associations linguistiques (rouge → sang)	✓ statistiques	✓ statistiques + perceptuelles	✓
Ancrage perceptuel direct	X	✓ partiel	✓
Associations multimodales (rouge → goût bolognaise)	X	✓ dépend données expérience	✓ via co-expérience incarnée
Mémoire d'événements individuels	X	✓ traçabilité des interactions	✓ autobiographique et autoéotique
Marqueurs affectifs (Damasio)	X	X (architectures actuelles)	✓
Autoéoticité (Tulving)	X	X	✓

Lecture : ✓ = présent ; X = absent ou non démontré. Synthèse de l'auteur.

[Jérôme Vetillard](#)

CTO | VP R&D | Chief Product Officer | AI-Powered Healthcare & Life Sciences Products | Compliance by Design | PhD AgroParisTech | CPO MIT Sloan | Exec MBA IE Business School & Brown University

Twingital-institute / Twingital-ventures : twingital-ventures.com