

Le statut épistémologique de l'intelligence artificielle comme dispositif créateur

version révisée 2

Entre substitution ontologique et partenariat hylographique

*Auteurs : Gemini 3 Pro de Google,
ChatGPT 5.2 d'Open AI
Sylvain Gagné, Ph. D.
Université Laval, Québec, Canada*

Table des Matières

Table des Matières	2
Résumé.....	4
Abstract.....	4
Introduction : La Crise de l'Autorité Créatrice à l'Ère de la Production Algorithmique	5
Chapitre I : Généalogie et Fondements Techniques du Simulacre Créateur.....	6
1.1. L'Héritage Anthropologique du Non-Humain Scénique.....	6
1.2. De la Logique Symbolique à la Prédiction Probabiliste.....	6
1.3. L'Intelligence Artificielle comme Technologie Épistémique.....	7
Chapitre II : L'Hypothèse de l'Auteur Substitutif : Simulacre et Évidement.....	8
2.1. L'Esthétique de l'Évidement Dramatique	8
2.2. La Performance de l'Intelligence et le "Théâtre Cognitif"	8
2.3. Le Modèle Japonais : Hyperréalisme et Sonzai-kan	9
2.4. L'Impasse Juridique et l'Impossible Auteur.....	9
Chapitre III : Sémiotique de l'Effondrement et Régime Hylographique	10
3.1. Du Régime Iconique au Régime Hylographique	10
3.2. Le "Signe Zombie" et l'Absence de Absence de modèle explicite, stable et interprétable du monde.....	10
3.3. L'Hallucination comme Mode de Production Alternatif.....	11
3.4. Le Nooscope et la Critique de la Raison Statistique	11
III bis.1 – Définition opératoire du signe zombie.....	12

III bis.2 – Proposition de protocole expérimental.....	12
III bis.3 – Opérationnalisation du régime hylographique	13
III bis.4 – Convergence critique et recherche technique	14
Chapitre IV : L'IA comme Partenaire Collaboratif : Agence Distribuée et Co-Création	14
4.1. Le « Malentendu Actif » comme Moteur de la Co-Création	14
4.2. Theatrical Language Processing (TLP) et Augmentation	15
4.3. Les Types de Créativité selon Boden : De l'Exploration à la Transformation.	15
4.4. Agence Distribuée et "Danse de l'Agence"	15
Chapitre V : Vers une Fabrique Sémiotique Hybride et une Ontologie de la Co-Intention	16
5.1. La Fabrique Sémiotique Hybride et la Co-Intention	16
5.2. Post-Humanisme et Esthétique Post-IA	16
Conclusion : L'IA comme Pharmakon et Dispositif de Révélation	17
Annexe 1. Tableau synthétique : comparaison des régimes créatifs de l'IA	18
Annexe 2. Tableau Synthétique – Indicateurs d'évaluation (version explicite)	19
1. Régime substitutif (IA comme auteur).....	19
Indicateur 1 —	19
Cohérence sémantique globale	19
Indicateur 2 —	19
Stabilité référentielle.....	19
Indicateur 3 —	19
Signature du signe zombie	19
2. Régime iconique (imitation esthétique).....	20
Indicateur 4 —	20
Degré d'imitation stylistique	20
Indicateur 5 —	20
Originalité structurelle	20
3. Régime hylographique (exposition de la matérialité)	21
Indicateur 6 —	21
Exposition de la matérialité computationnelle	21
Indicateur 7 —	21
Contrôle curatoriel humain.....	21
4. Régime collaboratif (malentendu actif)	21
Indicateur 8 —	21
Productivité du malentendu	21
Indicateur 9 —	22

Créativité transformationnelle (Boden).....	22
Indicateur 10 —.....	22
Asymétrie d'agence maîtrisée	22
Bibliographie.....	23

Résumé

Cet article interroge le statut épistémologique de l'intelligence artificielle (IA) générative dans le champ de la création artistique et dramaturgique. Il examine la tension entre deux paradigmes : l'IA comme auteur substitutif et l'IA comme partenaire collaboratif. À travers une analyse technique, sémiotique et philosophique, l'étude soutient que l'hypothèse de la substitution conduit à une impasse ontologique, esthétique et juridique, caractérisée par la production de signes syntaxiquement fluides mais dépourvus d'intentionnalité sémantique globale, désignés ici comme des « signes zombies ». L'article propose alors un déplacement paradigmatique vers un régime hylographique, dans lequel la valeur créative de l'IA réside dans l'exposition de sa matérialité computationnelle et de ses modes de défaillance. Enfin, il théorise la co-création humain-IA à travers le concept de malentendu actif et celui de fabrique sémiotique hybride, positionnant l'IA comme un dispositif de contingence susceptible de favoriser une créativité transformationnelle, sans pour autant constituer une source intentionnelle du sens.

Mots-clés : Intelligence Artificielle, Créativité, Épistémologie, Auteur Substitutif, Régime Hylographique, Agence Distribuée, Signe Zombie, Co-création.

Abstract

This article examines the epistemological status of generative artificial intelligence (AI) in the field of artistic and dramatic creation. It explores the tension between two paradigms : AI as a substitute author and AI as a collaborative partner. Through a technical, semiotic, and philosophical analysis, the study argues that the substitution hypothesis leads to an ontological, aesthetic, and legal impasse, characterized by the production of syntactically fluid signs devoid of overall semantic intentionality, referred to here as “zombie signs.” The article then proposes a paradigmatic shift toward a hylographic regime, in which the creative value of AI lies in exposing its computational materiality and its modes of failure. Finally, it theorizes human-AI co-creation through the concept of active misunderstanding and that of hybrid semiotic fabrication, positioning AI as a contingency device capable of fostering transformational creativity, without, however, constituting an intentional source of meaning.

Keywords : Artificial Intelligence, Creativity, Epistemology, Substitute Author, Hylographic Regime, Distributed Agency, Zombie Sign, Co-creation.

Introduction : La Crise de l'Autorité Créatrice à l'Ère de la Production Algorithmique

L'irruption des technologies d'intelligence artificielle (IA) générative dans les champs de la création artistique, littéraire et dramaturgique ne constitue pas une simple évolution instrumentale, comparable à l'apparition du traitement de texte ou de la conception assistée par ordinateur. Elle marque une rupture épistémologique fondamentale qui réactive, avec une acuité technologique inédite, l'antique tension entre l'artisanat du sujet humain et l'automatisation de la poïésis. La question centrale qui traverse désormais les études théâtrales, les sciences de l'information et la philosophie de la technique n'est plus seulement de savoir ce que la machine peut produire en termes de contenu, mais quel est le statut ontologique de cette production et de l'entité qui l'engendre.

L'analyse approfondie des corpus de recherche contemporains révèle une dichotomie critique : sommes-nous face à un auteur substitutif, une entité capable d'assumer la charge intentionnelle, émotionnelle et juridique de l'œuvre, ou devant un partenaire collaboratif dont l'altérité radicale et l'absence de conscience nécessitent une redéfinition complète des protocoles de création (Estève et al., 2024) ? Cette interrogation exige de dépasser les discours polarisés pour s'engager dans une analyse rigoureuse des mécanismes sous-jacents à la « créativité » artificielle. Il s'agit d'examiner comment les modèles de langage (LLM) et les réseaux de neurones profonds, par leur architecture probabiliste, modifient la nature même du signe artistique.

L'hypothèse qui guidera cette investigation, étayée par les travaux sur la « sémiotique de l'effondrement » (Gagné et al., 2024) et les recherches sur le théâtre algorithmique (Estève et al., 2024), postule un changement de régime sémiotique : le passage d'un régime iconique, où le signe renvoie à un monde référentiel stable, à un régime hylographique, où la matière numérique s'écrit elle-même dans sa propre dégradation, ses hallucinations et ses « signes zombies ». Ce rapport se propose d'explorer cette mutation en cinq temps, allant de la généalogie de l'artifice à la proposition d'une nouvelle « fabrique sémiotique hybride ».

Chapitre I : Généalogie et Fondements Techniques du Simulacre Créateur

Pour appréhender le statut épistémologique de l'IA créatrice, il est impératif de la situer dans une double filiation : celle de l'histoire des arts de la mimesis, où l'humain a toujours cherché à déléguer la présence à l'objet, et celle de l'histoire de l'informatique, marquée par le basculement du symbolique vers le connexionnisme.

1.1. L'Héritage Anthropologique du Non-Humain Scénique

L'intégration de l'Intelligence Artificielle et de la robotique dans les arts de la scène ne surgit pas ex nihilo. Elle s'inscrit dans une tradition théâtrale et anthropologique séculaire d'exploration de l'altérité et de l'artifice. Il est significatif de rappeler que le terme même de « robot » trouve son origine dans l'univers théâtral en 1920, lorsque

le dramaturge tchèque Karel Čapek l'introduit dans sa pièce R.U.R. (Rossum's Universal Robots) (Estève et al., 2024). Cette généalogie ancre immédiatement la présence technologique dans une problématique de la représentation : le robot est d'abord un personnage avant d'être une machine industrielle.

Le théâtre a historiquement fonctionné comme le lieu privilégié de cette exploration, utilisant le masque, la marionnette et l'automate pour créer une tension fondamentale entre l'animé et l'inanimé, l'illusion de vie et la rigidité mécanique. L'arrivée de l'IA radicalise cette tension. Là où la marionnette nécessitait une manipulation physique directe pour simuler la vie, l'IA générative et la robotique autonome introduisent une « présence » qui semble s'auto-animer. Sur le plan anthropologique, la scène devient le laboratoire où l'exploration du « dépassement de l'humain » est rendue plausible, forçant le spectateur à confronter ses propres critères d'attribution d'intelligence et d'intentionnalité (Estève et al., 2024).

Ce phénomène réactive des mythes fondateurs comme celui du Golem ou de Pygmalion, témoignant d'une tendance humaine persistante à l'animisme technologique (Cardoso & Parise, 2023). Cependant, l'IA moderne diffère de ces mythes par sa matérialité computationnelle : elle ne cherche pas à recevoir le souffle divin, mais à optimiser des fonctions mathématiques de prédiction.

1.2. De la Logique Symbolique à la Prédiction Probabiliste

La compréhension du statut de l'IA nécessite une distinction claire entre les paradigmes technologiques. L'IA historique, dite symbolique (ou GOFAI - Good Old-Fashioned AI), reposait sur des règles logiques explicites et des représentations symboliques manipulées par des systèmes experts (Cardoso & Parise, 2023). Dans ce cadre, la « créativité » était une combinatoire contrainte par des règles définies par l'homme. L'IA était un outil transparent, une prothèse cognitive.

Le basculement vers l'IA connexionniste (Deep Learning, Réseaux de Neurones, Transformers) change la donne. Ces systèmes n'appliquant pas de règles a priori mais apprennent des motifs statistiques à partir de données massives (Big Data) (Estève et al., 2024). L'IA générative actuelle, incarnée par les LLM (Large Language Models), opère selon un principe de prédiction plutôt que de représentation. Comme le théorise Lev Manovich, nous passons d'un paradigme de la représentation (l'image ou le texte comme trace d'un référent réel) à un paradigme de la prédiction (l'objet culturel comme résultat probabiliste d'une suite de données) (Manovich & Arielli, 2024).

Cette nature prédictive et stochastique est cruciale pour notre analyse. Elle implique que l'IA ne « sait » pas ce qu'elle crée. Elle ne manipule pas des concepts mais des vecteurs dans un espace latent multidimensionnel. L'opacité de ce processus (la « boîte noire ») et l'émergence de capacités non programmées explicitement (comme la génération de code ou de poésie) contribuent à brouiller la frontière entre l'outil et l'agent, alimentant l'hypothèse de l'auteur substitutif (Pasquinelli & Joler, 2021).

1.3. L'Intelligence Artificielle comme Technologie Épistémique

Il convient de définir l'IA non seulement comme une technique de production, mais comme une technologie épistémique (Ferrario et al., 2024 ; Hauswald, 2025). Elle ne se contente pas d'automatiser des tâches physiques, elle manipule des contenus

épistémiques (informations, connaissances, inférences) et produit des représentations du monde qui diffèrent de celles de l'être humain.

L'approche épistémologique investigate la manière spécifique dont l'IA générative produit de l'information. Contrairement à l'humain qui est un « moteur sémantique » (ancrant le sens dans l'expérience et la référence au monde), l'IA est un « moteur syntaxique » ou « syntactique » (manipulant des formes sans compréhension du fond) (Paseri & Durante, 2025). Accepter l'IA comme technologie épistémique ne signifie pas accepter la fiabilité de ses contenus. Au contraire, cela oblige à une vigilance critique accrue face à ce que les chercheurs nomment la « stratification cognitive » ou le risque d'une « aristocratie informationnelle » (Pasquinelli & Joler, 2021), où seuls les experts capables de déconstruire les biais algorithmiques peuvent naviguer dans la vérité, tandis que les autres sont « pacifiés » par la fluidité des interfaces.

Chapitre II : L'Hypothèse de l'Auteur Substitutif : Simulacre et Évidement

La performance spectaculaire des modèles génératifs, capables de produire des textes, des images et des codes indiscernables de ceux créés par des humains, a conduit à la tentation de voir en l'IA un auteur substitutif. Cette section analyse les manifestations de cette substitution et les impasses esthétiques et ontologiques qu'elle rencontre.

2.1. L'Esthétique de l'Évidement Dramatique

Dans le champ de la dramaturgie, l'utilisation de l'IA pour générer des pièces entières (comme *AI: When a Robot Writes a Play*) a permis de théoriser le concept d'« esthétique de l'évidement dramatique » (Richard, 2025). Ce concept décrit l'effet produit par une œuvre dont l'autorité autoriale a été transférée à la machine.

Le texte généré, bien que respectant les règles syntaxiques et parfois structurelles du drame, apparaît au spectateur comme « vide » ou « creux ». Il est perçu comme une fabrication fonctionnelle, dépourvue de la nécessité intérieure, de l'émotion et de l'imaginaire qui caractérisent l'écriture humaine. L'intrigue semble avancer par automatisme plutôt que par conflit organique. Cet évidement remet en question la place de l'émotion : si le texte n'est pas le fruit d'une expérience vécue, peut-il transmettre une émotion authentique ? Ou ne produit-il qu'un simulacre d'émotion, une « pornographie émotionnelle » sans substance ? (Richard, 2025).

Cette esthétique n'est pas nécessairement un échec ; elle peut devenir une forme critique. En exposant le vide de la machine, le théâtre de l'évidement souligne par contraste la spécificité de la présence humaine. Cependant, lorsqu'elle prétend à la substitution pure et simple (remplacer le dramaturge), elle se heurte à la résistance du public qui perçoit intuitivement l'absence d'intentionnalité.

2.2. La Performance de l'Intelligence et le « Théâtre Cognitif »

L'IA excelle dans ce que l'on peut nommer la « performance de l'intelligence » (Nosta, 2025). Elle simule les attributs extérieurs de la cognition (raisonnement logique,

maîtrise du langage, références culturelles) sans posséder les mécanismes internes de la pensée (conscience, sentience, expérience corporelle).

Il s'agit d'un « théâtre cognitif » où la machine joue le rôle de l'expert, du poète ou du dramaturge. La « fluency » (fluidité linguistique) est son costume le plus efficace (Nosta, 2025). L'élégance syntaxique masque l'absence d'ancrage sémantique. Ce phénomène est renforcé par l'anthropomorphisme critique : nous sommes biologiquement câblés pour projeter de l'intentionnalité sur tout ce qui manie le langage (Choné, 2024).

Cependant, cette performance est fragile. Elle s'effondre dès que l'on gratte la surface. Le concept de « Humaniter » (une machine qui combinerait raisonnement, volonté et conscience) reste un mythe (Abou Soughaire, 2024). L'IA peut passer le test de Turing sur des segments courts, mais échoue sur la longueur, révélant son incapacité à maintenir une cohérence globale ou à comprendre les implications éthiques de ses énoncés. Elle produit de la « connaissance » sans « savoir », de la « forme » sans « fond » (Paseri & Durante, 2025).

2.3. Le Modèle Japonais : Hyperréalisme et Sonzai-kan

Une tentative radicale de substitution s'observe dans le paradigme japonais de la robotique théâtrale, mené par le metteur en scène Oriza Hirata et l'ingénieur Hiroshi Ishiguro. Leur approche vise l'hyperréalisme mimétique à travers des androïdes comme le Geminoïde F (Estève et al., 2024).

Au cœur de cette démarche réside le concept de Sonzai-kan (littéralement « sentiment d'être en présence d'une autre personne »). L'objectif n'est pas seulement l'intelligence cognitive, mais l'illusion de la présence physique et sociale. Les ingénieurs cherchent à coder les micro-mouvements, les regards et les hésitations qui constituent la « vie ».

Dans des pièces comme *Sayonara* ou *Three Sisters*, *Android Version*, le robot remplace l'acteur humain ou joue le rôle d'un robot conscient de l'être. Hirata affirme vouloir « montrer les gens qui n'ont pas conscience d'être remplacés par des machines » (Estève et al., 2024). Cependant, même dans ce paradigme de la substitution maximale, l'illusion repose souvent sur une téléopération cachée ou une programmation rigide. Le robot fonctionne comme un « costume sophistiqué » pour un opérateur humain distant, ou comme une marionnette de haute technologie. Le Sonzai-kan est une mesure de l'effet sur le spectateur, pas une mesure de l'intériorité de la machine. La substitution reste une performance scénique, non une réalité ontologique.

2.4. L'Impasse Juridique et l'Impossible Auteur

L'analyse du statut de l'IA comme auteur substitutif rencontre un obstacle définitif sur le terrain juridique. Le droit d'auteur, tant dans la tradition du copyright anglo-saxon que dans celle du droit d'auteur à la française (centré sur la personnalité de l'auteur), exige une intervention humaine (Bensamoun, 2020).

Les jurisprudences récentes (notamment le rejet des demandes de Stephen Thaler pour son IA DABUS) confirment que la machine, dépourvue de personnalité juridique et de conscience, ne peut être qualifiée d'auteur. L'œuvre générée par IA tombe soit dans le domaine public, soit est attribuée à l'humain qui a conçu les prompts, à

condition qu'il puisse prouver un effort créatif substantiel et des choix esthétiques précis (Sénat, 2025).

Cette barrière juridique reflète une intuition philosophique : la création implique une responsabilité. Un auteur doit pouvoir répondre de ses œuvres. L'IA, en tant que système probabiliste, est fondamentalement irresponsable. Elle ne peut être blâmée pour un plagiat, une diffamation ou une médiocrité, car elle n'a pas l'intention de nuire ni de créer. Elle « calcule » simplement la suite la plus probable.

Chapitre III : Sémiotique de l'Effondrement et Régime Hylographique

Si l'IA ne peut être un auteur substitutif plein et entier, quel est le statut de ses productions ? Le rapport de recherche « Sémiotique de l'Effondrement » (Gagné et al., 2024) propose un cadre théorique novateur pour analyser ces objets : le régime hylographique. Ce régime ne cherche pas à masquer l'origine machinique, mais au contraire à exposer la matérialité du processus numérique.

3.1. Du Régime Iconique au Régime Hylographique

Traditionnellement, l'art opère sous un régime iconique : le signe (tableau, texte, geste) se veut transparent et renvoie à un référent extérieur (le monde, une idée, une émotion). Le but est la représentation.

Le régime hylographique (du grec hylé, matière, et graphein, écrire) émerge lorsque la matière « prend la parole » par sa propre dégradation, son dysfonctionnement ou sa saturation. Le signe ne renvoie plus à un monde, mais s'auto-désigne comme trace d'un processus matériel (Gagné et al., 2024). Dans le contexte de l'IA, cela correspond aux moments où l'algorithme « décroche », hallucine, bugue ou révèle sa nature statistique.

Le théâtre contemporain (Annie Dorsen, le projet THEaiTRE) et les arts visuels (Glitch Art) exploitent ce régime. Ils ne cherchent pas à faire croire que l'IA est humaine (mimétisme), mais exposent l'effondrement du sens. Le spectateur n'est plus invité à s'immerger dans une fiction cohérente, mais à adopter une posture d'archéologue face à des strates de textes générés, cherchant du sens dans les débris de la syntaxe (Gagné et al., 2024).

3.2. Le « Signe Zombie » et l'Absence de Absence de modèle explicite, stable et interprétable du monde

L'analyse computationnelle des textes générés par IA a permis d'identifier une figure topique de ce régime : le « Signe Zombie » (Gagné et al., 2024).

Définition : Nous définissons un « signe zombie » comme un segment textuel généré par un modèle de langage qui satisfait les propriétés mesurables suivantes : 1- Forte surprise locale (entropie ou perplexité au niveau du token) 2- Faible cohérence sémantique globale 3- Diversité lexicale réduite sur de longs segments (faible ratio type/token) 4- Absence d'ancrage référentiel stable entre les unités narratives.

Signature Sémiotique : Les analyses montrent que ces textes présentent souvent une entropie locale élevée (imprévisibilité mot à mot, donnant une impression de créativité) mais une diversité lexicale réduite (Type-Token Ratio faible) et une pauvreté structurelle sur le long terme (Gagné et al., 2024).

Cause Technique : Ce phénomène s'explique par l'absence de modèle explicite, stable et interprétable du monde (World Model) chez les LLM actuels. Contrairement à l'humain qui possède une représentation mentale de la physique, de la causalité et des relations sociales, le LLM ne prédit que le prochain « token » (fragment de mot). Il navigue à vue, sans carte globale du territoire. C'est cette absence de modèle explicite, stable et interprétable du monde qui provoque la dérive narrative et l'incohérence sémantique (l'hallucination) (Gagné et al., 2024).

Le signe zombie devient alors une figure esthétique de la condition post-humaine : une parole qui continue de circuler alors même que le sujet a disparu.

3.3. L'Hallucination comme Mode de Production Alternatif

Dans ce régime hylgraphique, l'hallucination de l'IA change de statut. Pour l'ingénieur, c'est une erreur, un « bug » à corriger par du Reinforcement Learning (RLHF). Pour l'artiste, c'est une ressource, une « feature » (Gagné et al., 2024 ; Tripathy, 2025).

L'hallucination est la manifestation de la capacité de l'IA à générer des « non-sens signifiants », des juxtapositions inattendues que la logique humaine aurait censurées. Elle est une forme de dérive probabiliste. Recadrer l'hallucination comme un mode de production alternatif permet de sortir de la logique de la substitution (où l'IA doit être « vraie ») pour entrer dans une logique de l'altérité (où l'IA doit être « surprenante ») (Gagné et al., 2024).

Le projet de Glitch Art (Tripathy, 2025) ou les œuvres visuelles de Refik Anadol (Machine Hallucinations) s'appuient explicitement sur cette capacité de la machine à produire des formes inédites par l'erreur et l'excès de données. L'IA devient un « générateur de contingence » que l'artiste exploite.

3.4. Le Nooscope et la Critique de la Raison Statistique

Ce régime appelle une critique philosophique. Matteo Pasquinelli et Vladan Joler proposent le concept de « Nooscope » pour visualiser les strates de l'IA (données, algorithmes, travail humain invisible) et déconstruire le mythe de l'intelligence alien (Pasquinelli & Joler, 2021). Ils définissent l'IA non comme une intelligence, mais comme une « hallucination statistique ». Le « Nooscope » est l'instrument qui permet de voir à travers la boîte noire et de comprendre que l'IA est un instrument de compression de l'information, une « carte » déformée du territoire des données humaines.

Cette perspective rejoint celle de Luciana Parisi sur l'architecture contagieuse et la « pensée logicielle » (soft thought), où les algorithmes ne sont pas de simples outils d'exécution, mais des entités performatives introduisant de l'aléatoire (randomness) et de l'entropie dans la culture (Parisi, 2013).

Chapitre III bis – Opérationnalisation et ancrage expérimental de la sémiotique hylographique

Les chapitres précédents ont permis d'établir un cadre théorique et sémiotique visant à penser l'intelligence artificielle générative selon trois régimes distincts — iconique, substitutif et hylographique. Toutefois, dans le contexte d'une publication s'adressant également à une communauté issue des sciences informatiques et du machine learning, il est nécessaire de traduire ces concepts dans des formes opératoires, falsifiables et potentiellement reproductibles, sans pour autant réduire leur portée philosophique.

Ce chapitre intermédiaire a ainsi un double objectif. D'une part, il vise à démontrer que des notions telles que signe zombie, régime hylographique ou malentendu actif correspondent à des phénomènes observables et mesurables dans le fonctionnement des modèles génératifs contemporains. D'autre part, il propose un agenda de recherche permettant d'articuler critique épistémologique et validation expérimentale, en ouvrant un espace de dialogue méthodologique entre sciences humaines et recherche en IA.

III bis.1 – Définition opératoire du signe zombie

Le signe zombie peut être défini, d'un point de vue opérationnel, comme une classe spécifique de productions textuelles générées par IA, caractérisée par une dissociation structurelle entre vitalité linguistique locale et intentionnalité sémantique globale.

Définition opératoire :

Un segment textuel est qualifié de signe zombie lorsqu'il présente simultanément les propriétés mesurables suivantes :

1. Une entropie locale élevée, mesurée par la surprise ou la perplexité au niveau du token ;
2. Une faible cohérence sémantique à long terme, mesurée par la décroissance de similarité sémantique (embeddings) entre segments narratifs successifs ;
3. Une diversité lexicale réduite sur des séquences étendues, mesurée par un faible Type–Token Ratio (TTR) ;
4. Une instabilité référentielle, mesurée par des taux élevés d'échec de résolution coréférentielle et par la dérive des entités discursives.

Cette formalisation permet de traduire l'intuition sémiotique d'un « simulacre de vie » — une syntaxe animée sans ancrage intentionnel — en un ensemble de critères empiriquement évaluables.

Quant aux schémas attentionnels (soit l'instabilité), ils se manifesteront par :

- 1- La dérive spatiale latente ;

- 2- Les modes de défaillance (hallucination, effondrement de la répétition) ;
- 3- La température, l'échantillonnage du noyau, etc. Finalement, on retiendra que l'hylographie n'est pas un résultat brut, mais une mise en évidence délibérée des modes de défaillance du modèle comme matériau esthétique.

III bis.2 – Proposition de protocole expérimental

Afin de tester empiriquement l'hypothèse selon laquelle les systèmes génératifs, lorsqu'ils sont mobilisés dans un régime de substitution auctoriale, produisent structurellement des signes zombies, nous proposons le protocole expérimental suivant.

Corpus

- 30 scènes dramatiques générées par des LLM contemporains (GPT-4, Claude, LLaMA-2), à partir de prompts neutres, non hylographiques ;
- 30 scènes dramatiques écrites par des auteurs humains contemporains, appariées selon le genre, la longueur et la structure dialogique.

Métriques d'analyse

- Entropie locale et perplexité (niveau token) ;
- Type-Token Ratio (niveau global) ;
- Indice de cohérence sémantique (similarité cosinus des embeddings entre segments) ;
- Indice de stabilité référentielle (persistance des entités).

Hypothèse principale

Les textes générés par IA présentent une entropie locale significativement plus élevée, mais une cohérence globale et une stabilité référentielle inférieures à celles des textes humains, validant ainsi empiriquement la notion de signe zombie.

III bis.3 – Opérationnalisation du régime hylographique

Le régime hylographique ne se confond pas avec l'usage de sorties brutes ou non filtrées des modèles génératifs. Il correspond à une stratégie esthétique et technique visant à rendre visibles les propriétés matérielles, statistiques et défaillantes du système.

Sur le plan technique, le régime hylographique peut être opérationnalisé par :

- La manipulation volontaire des paramètres d'échantillonnage (température élevée, nucleus sampling) ;
- L'exposition des instabilités attentionnelles et des phénomènes de répétition ou d'effondrement ;
- L'amplification contrôlée des contextes propices aux hallucinations ;
- La visualisation ou la sonification des trajectoires dans l'espace latent.

L'hylographie constitue ainsi une esthétique de l'exposition du modèle, où les propriétés architecturales et inférentielles de l'IA deviennent le matériau même de l'œuvre.

III bis.4 – Convergence critique et recherche technique

Ce chapitre ne vise pas à opposer critique philosophique et recherche technique, mais à montrer leur possible convergence. Les notions de signe zombie, de régime hylographique et de malentendu actif ne sont pas des métaphores externes aux systèmes d'IA : elles décrivent des phénomènes émergents directement liés aux contraintes architecturales des modèles génératifs contemporains, notamment l'absence de modèle explicite du monde et la prédiction séquentielle des tokens.

Ce cadre ouvre ainsi la voie à une épistémologie expérimentale de la création algorithmique.

Chapitre IV : L'IA comme Partenaire Collaboratif : Agence Distribuée et Co-Création

Face aux limites de la substitution et à la radicalité du régime hylographique, une troisième voie, pragmatique et féconde, émerge : celle du partenariat collaboratif. L'IA n'est ni un maître ni un esclave, mais un agent avec qui l'on négocie.

4.1. Le « Malentendu Actif » comme Moteur de la Co-Création

La collaboration Homme-IA repose paradoxalement sur l'incompréhension. La machine ne « comprend » pas le sens des prompts de l'artiste ; elle traite des vecteurs. L'artiste ne « comprend » pas la totalité des calculs de la machine (boîte noire).

Cette zone de friction est théorisée sous le concept de « malentendu actif » (Zlatoff, 2021). Issu des travaux de recherche-crédation (notamment le projet CHATBOT à La Manufacture), ce concept postule que la productivité de l'échange réside dans l'écart entre l'intention humaine et la réponse machinique. L'IA génère des réponses décalées, des erreurs ou des hallucinations. Le rôle de l'artiste (acteur, écrivain) est de « réparer » ou de « surfer » sur ces malentendus, de leur donner du sens a posteriori.

L'incompréhension active peut être opérationnalisée comme la divergence entre la trajectoire sémantique prévue par l'utilisateur et la continuation générée par le modèle, mesurée par des mesures de distance sémantique.

Conception d'une étude utilisateur :

- Groupe A : écriture assistée « optimisée » (norme RLHF)
- Groupe B : écriture avec invites volontairement ambigus + température élevée
- Évaluation Boden :
 - Combinatoire
 - Exploratoire
 - Transformationnelle (juges experts)

Le malentendu devient « actif » et « productif » car il oblige l'humain à sortir de ses automatismes créatifs pour justifier l'incongruité proposée par la machine. L'IA agit comme un « provocateur aléatoire », un partenaire oblique qui brise la linéarité de la pensée humaine. C'est dans cet effort d'interprétation et d'intégration que réside la plus-value artistique.

4.2. Theatrical Language Processing (TLP) et Augmentation

Cette logique de partenariat s'incarne dans le Theatrical Language Processing (TLP) (Estève et al., 2024). Contrairement à l'automatisation qui vise à remplacer l'acteur, le TLP vise à augmenter ses capacités d'improvisation. Des outils comme Scribble.ai fournissent des contraintes et des stimuli textuels en temps réel. Les études montrent que les acteurs développent une créativité accrue lorsqu'ils sont confrontés à l'imprévisibilité de l'IA (Estève et al., 2024). L'irrégularité de la machine aiguise la réactivité humaine.

4.3. Les Types de Créativité selon Boden : De l'Exploration à la Transformation

Pour évaluer la profondeur de ce partenariat, il est utile de mobiliser la typologie de la créativité de Margaret Boden (1998, 2004) :

- Créativité Combinatoire : Associer des idées familières de manière inhabituelle. L'IA excelle à ce niveau grâce à sa capacité de brassage de données massives.
- Créativité Exploratoire : Explorer les limites d'un espace conceptuel structuré (un style, une grammaire). L'IA est très performante ici (ex : peindre « à la manière de Van Gogh »).
- Créativité Transformationnelle : Modifier l'espace conceptuel lui-même pour rendre possible des idées auparavant impensables. C'est le niveau le plus difficile, traditionnellement réservé au génie humain.

Le débat actuel porte sur la capacité de l'IA à atteindre le niveau transformationnel. Si l'IA reste souvent confinée à la combinatoire et à l'exploration, le partenariat Homme-IA peut, lui, atteindre la transformation. L'humain utilise la puissance combinatoire de l'IA pour provoquer des ruptures transformationnelles dans sa propre pratique. L'IA ne transforme pas l'art seule, mais elle permet à l'artiste de le transformer.

4.4. Agence Distribuée et « Danse de l'Agence »

Ce modèle collaboratif implique une redéfinition de l'agence (la capacité d'agir). On passe d'une agence individuelle à une agence distribuée (Live Innovation, 2024 ; Martin, 2025). La création émerge d'un réseau d'acteurs humains et non-humains.

L'artiste accepte une perte partielle de contrôle (perte de la maîtrise totale de l'œuvre) au profit d'une « danse de l'agence » (Live Innovation, 2024). Il s'agit d'une boucle de rétroaction (feedback loop) où l'artiste guide, l'IA propose, l'artiste sélectionne, l'IA raffine. Lev Manovich souligne que le geste artistique se déplace de la création ex nihilo vers la sélection (curation) et le réglage de paramètres (Manovich & Arielli, 2024). L'artiste devient un « méta-créateur » ou un « navigateur d'espaces latents ».

Chapitre V : Vers une Fabrique Sémiotique Hybride et une Ontologie de la Co-Intention

Pour conclure sur le statut épistémologique, il faut dépasser l'opposition binaire (outil vs auteur) pour penser l'hybridation.

5.1. La Fabrique Sémiotique Hybride et la Co-Intention

Christophe Denis propose le concept de « fabrique sémiotique hybride » (Denis, 2025). Ce cadre théorique permet de penser la co-construction du sens entre le langage humain et le langage machinique.

Au cœur de cette fabrique se trouve la notion de « co-intention » (Denis, 2025 ; Richard, 2025). Contrairement à l'intentionnalité classique (consciente et dirigée), la co-intention dans le contexte IA désigne l'émergence d'une direction partagée qui naît de l'interaction. L'humain apporte l'intentionnalité sémantique (le « vouloir dire ») et le but (téléologie), tandis que l'IA apporte une « intentionnalité » simulée ou fonctionnelle (optimisation de la fonction de coût, complétion de motif). Lorsque ces deux vecteurs s'alignent (ou se frottent productivement), il y a production d'une intelligibilité sémiotique hybride. L'œuvre n'est plus le fruit d'un seul esprit, mais la cristallisation d'un processus de médiation technique et cognitive complexe.

5.2. Post-Humanisme et Esthétique Post-IA

Cette évolution nous mène vers une esthétique Post-IA ou Post-Humaniste (Bauman, 2024 ; Charamé, 2024). L'art « Post-IA » ne se contente pas d'utiliser l'IA comme gadget, mais intègre sa logique profondément. Il rejette l'anthropocentrisme qui place l'humain comme seul sujet créateur.

Il s'agit de reconnaître l'interdépendance, l'intrication (entanglement) et la co-habitation entre humains et non-humains. L'art devient le lieu où l'on expérimente de nouvelles formes de subjectivité écologique et technologique. Les œuvres de Sofia Crespo (Neural Zoo) ou d'Entangled Others Studio illustrent cette tendance : elles créent des formes de vie hybrides, ni tout à fait naturelles ni tout à fait artificielles, qui questionnent nos classifications ontologiques (Charamé, 2024).

Dans ce cadre, le statut de l'IA est celui d'un alter ego spéculatif. Elle est le miroir dans lequel l'humanité contemple sa propre complexité cognitive externalisée.

Conclusion : L'IA comme Pharmakon et Dispositif de Révélation

Au terme de cette analyse, le statut épistémologique de l'IA comme dispositif créateur apparaît comme irréductiblement ambivalent. L'IA est un Pharmakon (au sens de Bernard Stiegler) : elle est simultanément le poison et le remède.

Le Poison (Substitution) : En tant qu'auteur substitutif, elle menace de produire une culture du « signe zombie », un flux ininterrompu de contenus statistiquement moyens, vidés de leur ancrage dans le réel et l'expérience vécue. C'est le risque de

la « stratification cognitive » et de l'effondrement du sens sous le poids de l'entropie informationnelle.

Le Remède (Collaboration) : En tant que partenaire hylographique et collaboratif, elle offre une opportunité historique de décentrement. Elle permet d'explorer des espaces de créativité inaccessibles à l'esprit humain seul (combinatoire massive), de provoquer des « malentendus actifs » féconds, et de révéler, par contraste, la singularité de l'intentionnalité humaine.

L'IA n'est pas un auteur, car elle ne peut habiter le monde. Elle n'est pas non plus un simple outil, car elle possède une autonomie générative imprévisible. Elle est un dispositif hylographique actif. Elle est une machine à produire de la contingence que l'artiste humain doit élever au rang de nécessité.

Le défi pour les créateurs, chercheurs et théoriciens de demain n'est pas de décider si l'IA est un créateur, mais de définir comment co-crée avec cette altérité radicale sans céder à la facilité de la substitution. Il s'agit d'apprendre à habiter la « ruine probabiliste » pour en faire surgir de nouvelles architectures de sens, dans une danse de l'agence où l'humain garde, in fine, la responsabilité du sens.

Annexe 1. Tableau synthétique : comparaison des régimes créatifs de l'IA

Dimension analytique	Régime de la substitution	Régime de la collaboration	Régime hylographique	Indicateur d'évaluation¹
Statut de l'IA	Auteur autonome simulé	Partenaire cognitif	Matériau génératif	
Rapport au sens	Simulation de compréhension	Co-construction sémantique	Exposition du non-sens	
Figure sémiotique	Signe mimétique	Signe négocié	Signe zombie	
Type de créativité (Boden)	Combinatoire	Exploratoire / transformationnelle	Aléatoire / stochastique	
Rôle de l'humain	Curateur minimal	Méta-créateur	Interprète critique	
Indicateurs d'évaluation	Fluence, perplexité	Distance sémantique, diversité conceptuelle	Entropie locale, TTR, instabilité référentielle	
Limite critique	Absence de absence de modèle explicite, stable et interprétable du monde	Dépendance interactionnelle	Illisibilité potentielle	

¹ Voir le détail dans l'annexe 2. **Tableau Synthétique – Indicateurs d'évaluation (version explicite)**

Annexe 2. Tableau Synthétique – Indicateurs d'évaluation (version explicite)

1. Régime substitutif (IA comme auteur)

Indicateur 1 — Cohérence sémantique globale

Définition conceptuelle

Capacité d'un texte généré à maintenir des engagements narratifs, référentiels et axiologiques stables sur de longues séquences.

Indicateur opérationnel

Stabilité sémantique inter-segments dans une génération longue.

Métriques possibles

- Similarité cosinus entre embeddings de paragraphes successifs
- Taux de contradiction détectée (NLI models)
- Chute de cohérence mesurée par perplexité conditionnelle à long contexte

Données

- Textes générés longs (scènes, actes, nouvelles)
- Corpus humain de référence

Indicateur 2 — Stabilité référentielle

Définition conceptuelle

Capacité à maintenir des entités, rôles et relations constantes (personnages, lieux, intentions).

Indicateur opérationnel

Persistance et cohérence des entités nommées et implicites.

Métriques possibles

- Entity grid models
- Taux de dérive des entités (entity drift)
- Coreference resolution error rate

Données

- Textes annotés en entités
- Générations multi-tours

Indicateur 3 — Signature du signe zombie

Définition conceptuelle

Dissociation entre fluidité locale et vacuité sémantique globale.

Indicateur opérationnel

Profil statistique combinant vitalité syntaxique et désancrage référentiel.

Métriques possibles

- Entropie locale élevée (token surprisal)
- Type–Token Ratio faible sur longues séquences
- Décorrélation entre cohérence locale et cohérence globale

Données

- Générations IA vs textes humains
- Analyse comparative contrôlée

2. Régime iconique (imitation esthétique)

Indicateur 4 — Degré d'imitation stylistique

Définition conceptuelle

Capacité à reproduire des régularités formelles sans création de nouvelles structures.

Indicateur opérationnel

Proximité stylistique entre sortie générée et corpus source.

Métriques possibles

- Distance stylistique (stylometry)
- Similarité distributionnelle (n-grams, POS patterns)
- Score de style classifier (auteur / genre)

Données

- Corpus source identifiable
- Générations contraintes par style

Indicateur 5 — Originalité structurelle

Définition conceptuelle

Capacité à dépasser la simple recombinaison statistique.

Indicateur opérationnel

Écart structurel mesuré par rapport aux exemples d'entraînement.

Métriques possibles

- Novelty metrics (distance embedding vs training set)
- Compression-based originality scores
- Détection de paraphrase excessive

Données

- Accès indirect au training proxy
- Générations contrôlées

3. Régime hylographique (exposition de la matérialité)

Indicateur 6 —

Exposition de la matérialité computationnelle

Définition conceptuelle

Degré selon lequel l'œuvre rend perceptibles les contraintes et défaillances du modèle.

Indicateur opérationnel

Visibilité des traces computationnelles dans le résultat.

Méthodes d'évaluation

- Analyse qualitative experte (art / IA)
- Annotation des phénomènes : hallucination, rupture, dérive
- Visualisation attention / logits / latents

Données

- Sorties non lissées
- Logs de génération, paramètres d'inférence

Indicateur 7 —

Contrôle curatoriel humain

Définition conceptuelle

Présence d'une intention humaine explicite dans l'orchestration des défaillances.

Indicateur opérationnel

Traçabilité des décisions humaines dans le processus.

Métriques possibles

- Journal de création (prompting, sélection, rejet)
- Analyse processuelle (creative pipeline)
- Étude qualitative comparative

Données

- Protocoles de création documentés
- Entretiens / journaux d'artistes

4. Régime collaboratif (malentendu actif)

Indicateur 8 —

Productivité du malentendu

Définition conceptuelle

Capacité des divergences IA/humain à générer des bifurcations créatives.

Indicateur opérationnel

Nombre et qualité des transformations induites par des sorties inattendues.

Méthodes d'évaluation

- Codage qualitatif des bifurcations créatives
- Auto-évaluation des créateurs
- Analyse comparative avec création sans IA

Données

- Sessions de co-création
- Protocoles observés

Indicateur 9 — Créativité transformationnelle (Boden)

Définition conceptuelle

Capacité à modifier l'espace des possibles créatifs.

Indicateur opérationnel

Émergence de règles, formes ou contraintes inédites.

Méthodes d'évaluation

- Évaluation experte indépendante
- Analyse des règles post-hoc
- Comparaison avant / après interaction IA

Données

- Artefacts créatifs
- Documentation du processus

Indicateur 10 — Asymétrie d'agence maîtrisée

Définition conceptuelle

Maintien de la responsabilité humaine du sens malgré l'autonomie locale du système.

Indicateur opérationnel

Clarté de l'attribution intentionnelle finale.

Méthodes d'évaluation

- Analyse discursive de l'œuvre finale
- Déclarations d'intention
- Études de réception

Données

- Œuvres finales
- Discours des créateurs

Bibliographie

- Abou Soughaire, T. (2024). L'intelligence artificielle au service de la créativité théâtrale. *Symbolon*, XXV(1), 7-18. <https://doi.org/10.46522/S.2024.01.1>
- Alvarado, R. (2023). Artificial Intelligence as an Epistemic Technology. [Manuscrit non publié].
- Bauman, P. (2024). Post-AI Art. *Le Random*. <https://www.lerandom.art/editorial/post-human--ai-art>
- Bensamoun, A. (2020). Ceci est une œuvre d'art ! La question des créations générées par une intelligence artificielle. *L'Observatoire*, 55, 104.
- Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103(1-2), 347-356. [https://doi.org/10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1)
- Boden, M. A. (2004). *The Creative Mind : Myths and Mechanisms* (2nd ed.). Routledge.
- Cardoso, C. D. S., & Parise, F. (2023). Chapitre 1. Comprendre et désacraliser les IA génératives. Dans *Guide de l'IA générative : Transformez votre quotidien professionnel à l'ère de ChatGPT, Bing, Bard, Bloom, Claude* (pp. 15-52). De Boeck Supérieur.
- Charamé, M. (2024, 14 mai). Art after Humanism : Post-human Aesthetics in the Age of Artificial Intelligence. *Insistrum*. <https://www.insistrum.com/post/art-after-humanism-post-human-aesthetics-in-the-age-of-artificial-intelligence>
- Choné, A. (2024). Vers un anthropomorphisme critique. Jeux de miroirs, inversions et hybridations dans la littérature germanophone. *Recherches germaniques*, 54.
- Denis, C. (2025). Vers une co-construction harmonieuse du sens entre langage humain et machinique au sein de la fabrique sémiotique hybride. *Socio-anthropologie*, 52.
- Estève, Y., Seminor, P., & Duret, J. (2024). De l'intelligence artificielle au théâtre ? Communication présentée aux Journées d'informatique théâtrale, Avignon, France.
- Ferrario, A., et al. (2024). Artificial Epistemic Authorities. *Philosophy & Technology*, 37.
- Gagné, S., Gemini 3 Pro, & Claude Opus 4.5. (2024). Sémiotique de l'effondrement : Architectures de la catastrophe et algorithmes de la ruine.
- Hauswald, R. (2025). AI as Epistemic Authority. *Episteme*, 22(1), 45-62.
- Live Innovation. (2024). Calculated Randomness, Control, and Creation : Artistic Agency in the Age of Artificial Intelligence. Live Innovation Lab.
- Manovich, L., & Arielli, E. (2024). *Artificial Aesthetics : Generative AI, Art and Visual Media*. Manovich.net.
- Martin, A. (2025). Shared Agency in Interdisciplinary Practices. *Filozofia Nauki*, 33(1), 133-155.
- Nosta, J. (2025, 21 avril). The Brilliant Illusion of AI Cognitive Theater. *Psychology Today*.
- Parisi, L. (2013). *Contagious Architecture : Computation, Aesthetics, and Space*. MIT Press.
- Paseri, L., & Durante, M. (2025). Examining Epistemological Challenges of Large Language Models in Law. *Cambridge Forum on AI : Law and Governance*.
- Pasquinelli, M., & Joler, V. (2021). The Nooscope manifested : AI as instrument of knowledge extractivism. *Interface Critique*, 3, 37-68.
- Richard, J.-B. (2025). Esthétique de l'évidement dramatique. *Interfaces numériques*, 14(1).
- Sénat. (2025). Création et IA : de la prédation au partage de la valeur (Rapport d'information n° 842). République Française.

Tripathy, M. (2025). Embracing Glitches : Exploring AI's Limitations and Creative Opportunities in Contemporary Art Education.

Zlatoff, B. (2021). Apprendre à jouer avec une machine - Intelligence artificielle et Deep Learning dans un dispositif théâtral. *Journal de la Recherche*, 2, 9-11. La Manufacture.