Ne dites plus LLM : Larges Discours Models (LDM) et Agent Discursif Artificiel (ADA) ?

Amar Lakel

Ne dites plus LLM : Larges Discours Models (LDM) et Agent Discursif Artificiel (ADA) ?

Amar LAKEL
Laboratoire MICA
Université Bordeaux-Montaigne
amar.lakel@u-bordeaux-montaigne.fr

Résumé en français

Cette communication propose un déplacement épistémologique dans l'analyse des grands modèles génératifs, substituant à la catégorie "Large Language Models" (LLM) celle de "Large Discourse Models" (LDM), puis celle d'Agent Discursif Artificiel (ADA). Le cadre théorique s'appuie sur un tri ontologique distinguant trois instances régulatrices: l'appréhension des régularités phénoménales du monde référentiel, la structuration d'une cognition incarnée, et la sédimentation structuro-linguistique de l'énoncé dans un lieu socio-historique. Les LDM, opérant sur le fruit de ces trois instances (le document), modélisent la projection discursive d'une partie de l'expérience humaine réifiée par le corpus d'apprentissage. Le programme proposé vise à substituer à l'alternative "fascination/effroi" des épreuves publiques et des procédures rendant décidables la place, les usages et les limites des agents discursifs artificiels dans l'espace social contemporain, inscrivant cette démarche dans une perspective de gouvernance et de co-régulation associant État, industrie, société civile et académiques.

Mots-clés: Large Discourse Models, Agent Discursif Artificiel, analyse algorithmique, formations discursives, dette cognitive, gouvernance numérique, pragmatique énonciative

Abstract (English)

This paper proposes an epistemological shift in the analysis of large generative models, replacing the "Large Language Models" (LLM) category with "Large Discourse Models" (LDM), and subsequently with Artificial Discursive Agent (ADA). The theoretical framework rests on an ontological triage distinguishing three regulatory instances: apprehension of phenomenal regularities of the referential world, structuring of embodied cognition, and structural-linguistic sedimentation of utterances in a socio-historical context. LDMs, operating on the product of these three instances (the document), model the discursive projection of a portion of human experience reified through training corpora. The proposed program aims to substitute public tests and procedures that make decidable the place, uses, and limits of artificial discursive agents in contemporary social space for the "fascination/fear" alternative, inscribing this approach within a governance and co-regulation perspective associating State, industry, civil society, and academics.

Keywords: Large Discourse Models, Artificial Discursive Agent, algorithmic analysis, discursive formations, cognitive debt, digital governance, enunciative pragmatics

L'avènement des grands modèles génératifs déplace la focale de la recherche en études digitales du traitement du langage abstrait vers une analyse de la discursivité calculable, située et historisée. Notre démarche vise à examiner comment ces systèmes interrogent nos pratiques interprétatives et transforment nos méthodes, nos objets et nos cadres de validation de l'énoncé, en reconnectant l'analyse aux héritages théoriques et aux épreuves empiriques de la tradition SHS.

Nous proposons une redéfinition de la catégorie "LLM" vers celle de *Large Discourse Models* (LDM) : systèmes qui n'agrègent pas seulement des régularités linguistiques mais apprennent et imitent des formations discursives (genres, éthos, cadres, normes, positions énonciatives) sédimentées dans des corpus (Pêcheux, 1980; Pêcheux, 1982). La problématique s'appuie sur un tri ontologique distinguant trois instances régulatrices : (1) appréhension des régularités phénoménales (perception/action) du monde référentiel, (2) structuration d'une cognition incarnée (esprit/corps) dans un individu devenant sujet, (3) sédimentation structuro-linguistique d'un énoncé dans un lieu socio-historique. Les LDM, opérant sur le fruit de ces trois instances (le document), modélisent la projection discursive d'une partie de l'expérience de l'humanité réifiée par le corpus servant à leur apprentissage. C'est à ce niveau discursif qu'ils doivent être évalués comme fonctions imitatives des compétences cognitivo-discursives.

Nous défendons une seconde thèse : les LDM constituent le noyau de développement d'agents cognitivo-discursifs artificiels, l'*Agent Discursif Artificiel* (ADA), locuteur récemment apparu dans les champs sociaux (notamment académique) dont il faut développer la sociologie. LDM et ADA désignent une classe d'agents computationnels capables d'apprentissage et de généralisation des régularités discursives, d'exhibition d'une cohérence rationnelle attribuable (*intentional stance* évaluée par juges), d'alignement normatif via procédures explicites, de suivi de trajectoires biographiques traçables (versions, éducation, mémoires). Cette catégorisation régulatoire de l’agent (Akrich et al., 2006; Callon & Ferrary, 2006; Latour, 2007) vise à organiser l'évaluation, l'usage et la responsabilité d'une source de comportements auxquels on ne peut nier la capacité de raisonner et de converser.

Ce déplacement conceptuel répond à trois exigences : comprendre le pouvoir de composition discursive sans expérience incorporée du monde ; mettre en place un programme d'épreuves publiques de compréhension scientifique de l'agentivité discursive ; arrimer la discussion normative à des procédures et instances de gouvernance numérique. Nous explorons les conditions d'acceptabilité sociale des LDM selon quatre axes : capacités (compétences cognitivo-discursives), confiance (stabilité, traçabilité, gestion des erreurs), alignement normatif (protocoles documentés, tests de polis par communauté d'usage), gouvernance (inscription dans un espace public de co-régulation associant État, industrie, société civile et mondes académiques).

Ce programme prolonge nos travaux sur les régimes de gouvernance et de co-régulation à l'ère numérique (Lakel, 2005, 2007, 2021, 2025) et la transformation des espaces publics par le numérique (Lakel, 2007, 2008 et 2009), développant une grammaire procédurale pour penser l'"accueil" des agents discursifs artificiels (Lessig, 1999). Sur le plan

empirique, nous proposons un programme d'épreuves falsifiables mobilisant les méthodes digitales (constitution/traçabilité de corpus, annotation ouverte, protocoles mixtes) : biographie des LDM, mesure des contextualisations, éducation normative, agentivité discursive, transfert inter-tâches, gouvernance. L'apport principal est double : une ontologie située (LDM comme agents discursifs évalués au niveau de leurs pratiques de discours) ; une ingénierie de la socialisation (substituer à l'alternative "fascination/effroi" des épreuves publiques et procédures rendant décidables la place, les usages et les limites de ces agents). L'hypothèse d'"accueil" ne vaut qu'à ces conditions, documentées et auditées.

La recherche en humanités opère un tri ontologique minimal de régularités interdépendantes : (1) régularités phénoménales du monde référentiel (perception/action), (2) régularités des cognitions incarnées (esprit/corps), (3) régularités discursives sédimentées dans le corpus situé. Notre proposition théorique identifie dans les LDM la plus puissante empreinte formelle du processus herméneutique à ce jour : les réseaux de neurones, par leurs milliards de paramètres, modélisent les positions discursives d'une culture saisie dans un corpus donné. À l'instar d'un archéologue artificiel, ils interprètent la projection sémiotique d'un monde humain dont pourtant ils ignorent l'expérience directe. C'est à ce niveau pragmatico-discursif qu'ils doivent être caractérisés et évalués.

Notre argumentation se déploie en trois mouvements. La première partie établit le cadre théorique en explicitant le tri ontologique qui distingue trois instances régulatrices phénoménale, cognitive et discursive et situe les modèles génératifs au niveau de la sédimentation documentaire des formations discursives (I). La deuxième partie retrace la généalogie des modèles algorithmiques en sciences sociales, de la physique sociale de Quetelet aux architectures de deep learning, pour caractériser le passage des LLM aux Large Discourse Models et justifier la catégorie d'Agent Discursif Artificiel (II). La troisième partie propose un programme empirique articulant cinq axes d'investigation falsifiables, examine les critères opératoires d'évaluation des ADA et analyse, à partir de plus de 120 études empiriques, les effets de leur inscription dans l'espace social dette cognitive, reconfiguration organisationnelle et transformation des environnements éducatifs (III).

I. Cadre théorique : de la signifiance au tri ontologique

Nous inscrivons notre position dans une définition sémiotico-pragmatique du discours où le sens (signifiance) émerge d'une pratique ancrée et événementielle, "ni les mots ni les choses" (Foucault, 1969). La structure signifiante articule des régularités contraignantes (règles institutionnelles, grammaires, normes explicites) aux usages sociaux stabilisés par des communautés situées, dont le locuteur s’effectue dans une occurrence unique mais contrainte. L'énoncé, dans sa réification documentaire, constitue un artefact interprétable instanciant une position discursive (Guilhaumou, 2005; Pêcheux, 1982; Robin, 1973) où s'articulent l'éthos du locuteur, le genre, le référent, la situation communicationnelle et l'adresse (Jakobson, 2003) mais aussi un certain rapport au référent.

I.1 - L’instabilité des modèles d’interprétation : l’instance du sujet.

La tradition nominaliste médiévale, d'Ockham à Buridan, établit que les catégories générales par lesquelles nous organisons l'expérience les universaux constituent des constructions de l'intellect signifiant les singuliers, non des entités subsistant hors de l'esprit (d'Ockham, in Piché, 2005). Cette orientation ne verse pas dans le scepticisme : elle maintient l'accès cognitif aux singuliers tout en soulignant le caractère construit et économique de nos classifications. Les sciences cognitives contemporaines prolongent cette intuition en termes renouvelés : les catégories perceptives émergent de processus d'inférence probabiliste optimisant la prédiction et minimisant l'incertitude (Clark, 2013 ; Friston, 2010).

Dans le paradigme du traitement bayésien, le système cognitif opère comme instance de traduction active : il génère continuellement des hypothèses sur la “fabrique” de ses états sensoriels, révisées par confrontation aux signaux ascendants. L'organisme ne reçoit pas passivement des données mais anticipe les régularités de son environnement selon des modèles hypothétiques hiérarchiques. Ces modèles, appris par l'expérience, encodent des distributions de probabilité conditionnelle permettant de catégoriser les configurations sensorielles selon quelques traits distinctifs. Le système vise l'économie : réduire la complexité de l’expérience à des patterns prédictibles orientant l'action. C’est le processus de catégorisation ou d’habitus bien connu de la recherche en sciences sociales (Goffman 1974, Bourdieu 1979).

Le langage constitue alors une strate supplémentaire de traduction. L'énoncé émerge comme traduction de la traduction, conversion de l'expérience cognitive en artefact expérientiel et symbolique communicable dans lequel le sujet se donne à lui-même et aux autres dans une “mise en scène de la vie quotidienne”. Cette capacité poïétique, parfois réifiée en un document[1], devient un objet signifiant détachable de son contexte de production, portant pourtant la trace des instances ayant pesé sur sa genèse : instance référentielle (emprise du réel), instance subjective (modèle cognitif du locuteur), instance situationnelle (espace socio-communicationnel). Le document rejoint un corpus qui fera à son tour l'objet d'interprétations, instanciant une chaîne de traduction où chaque niveau transduit le précédent selon ses modalités propres.

L'interprétation des documents articule ainsi trois instances régulatrices distinctes. L'instance référentielle concerne l'emprise de l’expérience du réel sur l'énoncé : traces indexicales (déictiques, noms propres, descriptions définies) ancrant le discours dans un monde de référence. L'instance subjective concerne la position énonciative : éthos discursif (Maingueneau, 2002), modalisations marquant le degré d'adhésion du locuteur à ses énoncés, isotopies axiologiques trahissant une orientation évaluative (Rastier, 1987). L'instance communicationnelle concerne le dispositif d'adresse : genre discursif (Bakhtine, 1984), contrat de communication (Charaudeau, 2004), horizon pragmatique d'usage.

[1] Nous incluons l’expression orale comme un “document” éphémère.

L'ambition des sciences humaines et sociales a toujours été d'explorer ces structures documentaires (le corpus) comme traces socio-historiques informant les pratiques d'interprétation du sujet humain, situé. Qu'il s'agisse d'expliquer les régularités comportementales (tradition naturaliste), de comprendre le sens visé par les acteurs (tradition herméneutique) ou de critiquer les conditions d'influence et de domination des espaces sociaux (tradition critique), les SHS interrogent les fonctions de transformation par lesquelles le réel s'imprime dans les représentations collectives et individuelles.

I.2 Du statut des régularités phénoménales : un faux problème.

Mais pour saisir ces trois instances, les SHS ont été traversées, tout au long du XXe siècle, par une tension épistémologique entre deux orientations analytiques. La première privilégie l'identification de structures, systèmes de règles, codes symboliques, invariants formels dont les manifestations empiriques constituent des actualisations. La seconde s'attache aux régularités émergentes, patterns statistiques, distributions relationnelles, configurations dynamiques dont les structures ne sont que des stabilisations provisoires. L'émergence des grands modèles de langage réactive cette tension en offrant un opérateur technique capable d'encoder des régularités massives tout en exhibant des comportements émergents.

L'orientation structurale trouve des expressions diverses selon les disciplines. En linguistique, Saussure (1916) établit la langue comme un code a priori, système différentiel où la valeur des signes procède de leurs règles d'opposition. Chomsky (1965, 1995) radicalise ce programme en postulant une compétence linguistique innée, biologiquement déterminée dont les langues particulières constituent des paramétrisations. En anthropologie, Lévi-Strauss (1949, 1958) recherche les structures élémentaires organisant les systèmes de parenté et les mythes, invariants formels sous-jacents à la diversité culturelle. En sémiotique narrative, Greimas (1966, 1983) modélise la grammaire des récits par un dispositif d'actants et de programmes narratifs. Ces approches partagent une ambition : dégager, sous la variabilité des manifestations, un système générateur de portée plus ou moins universelle.

L'orientation distributionnelle procède différemment. Harris (1954, 1968), rompant avec le mentalisme chomskyen, propose une linguistique fondée sur l'analyse des distributions régularités de co-occurrence observables dans les corpus sans recours à l'intuition du locuteur natif. Benzécri (1973), avec l'analyse factorielle des correspondances, généralise cette approche en offrant un outil de réduction dimensionnelle applicable à tout tableau de contingence. Le sens devient effet de position dans un espace multidimensionnel, non décodage d'un signifié préétabli. Cette tradition trouve des prolongements contemporains dans la linguistique de corpus, la lexicométrie et, précisément, les architectures neuronales des LDM dont l'apprentissage repose sur la prédiction distributionnelle.

Mais ces deux orientations que l'on a voulu opposer ne s'excluent pourtant aucunement. Foucault (1969), analysant les formations discursives, refuse tant le formalisme linguistique que le référentialisme naïf ("ni les mots, ni les choses") : l'énoncé constitue un événement à la fois singulier et pourtant soumis à des conditions de possibilité

historiquement situées. L'épistémè n'est ni code transcendant ni simple donnée statistique, mais la configuration de rapports entre formes de pratiques énonciatives et formes de pratiques non discursives à la fois informant et objet de résistance/innovation. Les sciences de la complexité (Atlan, 1979 ; Morin, 1977) ont montré depuis comment des régularités locales traversent des seuils critiques pour se stabiliser en structures et comment ces structures demeurent susceptibles de déstabilisation. Règles et régularités constituent moins des paradigmes antagonistes que des moments d'un processus où l'ordre émerge de l'interaction et se fixe en contraintes pesant sur les interactions ultérieures, susceptibles de les remettre en cause.

Cette dialectique éclaire le statut des LDM. Héritiers de la tradition distributionnelle par leur architecture (prédiction du token suivant fondée sur les régularités de co-occurrence), ils “apprennent” des comportements quasi-réglés respect de la syntaxe, cohérence argumentative, conformité aux genres discursifs sans encodage explicite de règles. La matrice des régularités suffit à produire des outputs structurés, suggérant que les « codes » symboliques constituent peut-être des descriptions de haut niveau de régularités statistiques sous-jacentes plutôt que des systèmes générateurs autonomes.

I.3 - ordre et chaos de la construction de la réalité : le document inaccessible.

La pragmatique des actes de langage (Austin, 1962 ; Searle, 1969), articulée à la linguistique de l'énonciation (Benveniste, 1966, 1974), établit que tout énoncé constitue simultanément un dire et un faire : il accomplit une action dans un contexte de communication. Les Sciences de l'Information et de la Communication (SIC) prolongent cette perspective en schématisant le document comme artefact communicationnel transformation de la pensée en trace matérielle adressée à autrui, traduction pouvant aller jusqu'à la trahison .

L'approche exégétique développée dans la tradition française de l'analyse automatique du discours (Pêcheux, 1969 ; Maingueneau, 1984 ; Robin, 1973) offre les clés d’une mise en évidence de régularités formelles. Elle articule deux opérations méthodologiques fondamentales : la segmentation (discrétisation de l'inventaire des items en catégories exhaustives, non redondantes et opérationnalisables) et la **mise en relation** (cartographie des co-occurrences, identification de classes paradigmatiques révélant les contraintes de sélection structurant le discours). Cette démarche instrumente l'interprétation par l'objectivation des récurrences, posant des hypothèses de lecture révisables par confrontation aux régularités textuelles. La corrélation entre patterns intra-discursifs, régularités inter-discursives et observations extra-discursives (contexte historique, données sociologiques, archives) permet d'évaluer la présence des instances régulatrices sans prétendre atteindre une vérité du monde. L'exégèse ne dévoile pas un sens préexistant : elle construit des correspondances testables entre formes discursives et hypothèses sur leurs conditions de production (traces extra-discursives).

Mais très vite, Guilhaumou (2007) observe que « la capacité réelle de l'analyse de discours à mettre en évidence, par des analyses comparatives de corpus, des stratégies discursives, des situations d'énonciation et des oppositions rhétoriques, se dissolvait généralement dans un usage très instrumentalisé de la description lexico-sémantique » (p. 178). L'analyse du discours des années 1970 se réduisait progressivement à une « boîte à outils » lexicologiques et argumentatifs. La petitesse des corpus étudiés, contrainte par les capacités computationnelles de l'époque, ne permettait pas de décrire des variations suffisantes à la génération de modèles discursifs. L'analyse automatique de discours est devenue alors une science auxiliaire (Bardin, 1989) sans réaliser ses ambitions initiales de formalisation des processus interprétatifs.

Cette trajectoire s'inscrit dans un phénomène plus large affectant l'ensemble du champ de l'intelligence artificielle (Sudmann et al., 2023). Tant que les approches symboliques espéraient formaliser le langage en règles explicites implémentables dans des systèmes déductifs, elles rataient la dimension éminemment complexe du discours. Les systèmes experts connurent un déploiement industriel notable dans les agents mécaniques mais échouèrent à capturer la complexité sémantique et pragmatique des productions discursives naturelles. Cet échec marqua paradoxalement un recul par rapport aux ambitions d'automatisation. La reconnaissance de la complexité irréductible des processus herméneutiques conduisait à privilégier des approches qualitatives traditionnelles mobilisant l'expertise du chercheur plutôt que des procédures algorithmiques.

La situation ne se modifia substantiellement qu'avec l'émergence des représentations vectorielles denses (word embeddings) au début des années 2010. L’inadéquation entre la complexité des phénomènes discursifs et la puissance des instruments disponibles dans les modèles symboliques, incapables de dépasser des représentations trop grossières pour capturer les régularités fines du discours réactiva l’exploration d’une autre hypothèse : laisser à la machine la capacité de formaliser des modèles trop complexes pour les humains. Les architectures neuronales profondes, avec leurs milliards de paramètres ajustés sur des corpus massifs, pourraient résoudre cette équation en apprenant implicitement des patterns d'une complexité inaccessible à la description explicite. Le passage de l'analyse automatique du discours aux LLM ne constitue donc pas une rupture mais l'accomplissement, par des détours de l’histoire (Hinton, Le Cun, etc) d'un programme formulé il y a plus d'un demi-siècle.

II - Renaissance de l’interprétation algorithmique de la construction de la réalité

La phase récente (depuis 2010) de l’I.A. intègre une nouvelle génération d'algorithmes de deep learning permettant l'identification de patterns d’une très grande complexité dans des données massives et hétérogènes. Accès au très grand corpus par des très grands modèles, couplé à des algorithmes d’apprentissage des régularités formelles forme un composé historique pour réaliser les utopies formelles des années 70. C'est dans ce magma d'innovations techno-industrielles autour de la formalisation de la signifiance à visée

documentaire qu'émerge la nouvelle instance qui nous occupe : l'I.A. générative comme modèle universel du discours[2].

II.1 - Mettre à jour les régularités : une archéologie des modèles algorithmiques en SHS.

Les SHS ont fondamentalement une prétention scientifique dont l'histoire est indissociable de la question algorithmique. La première phase de l'histoire de la formalisation algorithmique des SHS repose sur une opération épistémologique fondamentale : la transformation du social en tableau, projection des attributs d'une population dans une matrice de données constituant l'infrastructure de la pensée statistique. Quetelet (1835, 1869) développe le concept d'« homme moyen » comme abstraction statistique caractérisant les propriétés centrales d'une population, posant les bases d'une « physique sociale » cherchant dans les régularités numériques les lois gouvernant les collectifs. Galton et Pearson (1880-1900) formalisent les concepts de corrélation et de régression, offrant les outils de l'analyse bivariée. Durkheim (1897), dans *Le Suicide*, systématise l'usage sociologique de ces méthodes : les corrélations entre taux de suicide et variables sociales (confession, situation matrimoniale, intégration professionnelle) révèlent des « courants suicidogènes », forces collectives irréductibles aux motivations individuelles.

L'école française d'analyse des données opère un tournant décisif à partir des années 1960. L'analyse factorielle des correspondances de Benzécri (1973) effectue une double opération : construction d'axes factoriels réduisant la dimensionnalité des données, création d'un espace géométrique permettant la visualisation des proximités entre individus et entre modalités. Cette approche descriptive, visuelle, interprétative révèle des structures latentes d'association inaccessibles aux corrélations bivariées, ouvrant ainsi la voie à des modèles bien plus complexes. Bourdieu (1979) en fait un usage magistral dans *La Distinction*, projetant les pratiques culturelles et les positions sociales dans un espace homologue. La qualité d'un modèle se mesure à sa capacité à révéler des configurations signifiantes selon un principe d'économie : simplification contrôlée maximisant le pouvoir descriptif.

Les algorithmes d'apprentissage automatique prolongent cette généalogie tout en opérant une rupture. Le machine learning statistique des années 1990-2000 machines à vecteurs de support (Cortes & Vapnik, 1995), forêts aléatoires (Breiman, 2001) généralise l'apprentissage de frontières de décision complexes dans des espaces de grande dimension. Les techniques de réduction dimensionnelle non linéaire (t-SNE, van der Maaten & Hinton, 2008 ; UMAP, McInnes et al., 2018) et de clustering par densité (HDBSCAN, Campello et al., 2013) permettent l'exploration de structures dans des données massives et hétérogènes. Pour autant, ces modèles sont enfermés dans les sciences de l'ingénieur au service d'une industrie de gestion des contenus numériques sans pouvoir modifier les approches en humanité.

[2] et non plus seulement du langage qui n'en est qu'un fragment.

En effet, un paradoxe émerge : l'amélioration de la capacité prédictive (donc de la formalisation des modèles) s'accompagne d'une diminution de l'interprétabilité (donc d'une aliénation à la machine apprenante). Les modèles de deep learning, avec leurs millions puis milliards de paramètres, constituent des « boîtes noires » dont les représentations internes échappent à l'explication humaine directe. Cette transformation marque une rupture épistémologique dont il convient de souligner l'impasse. Dans le paradigme interprétatif, la scientificité résidait dans l'articulation entre pattern formel et intelligibilité théorique. Le modèle constituait un instrument de compréhension : simplification du réel rendant manifestes des structures latentes accessibles à l'entendement. Le deep learning inverse cette logique. Le modèle ne rend plus compte de régularités constatées, il les "prouve" par sa seule capacité prédictive. La validation ne procède plus de la correspondance entre structure formelle et interprétation substantive mais sur la mesure de la performance sur des données que le modèle n'a pas rencontrées lors de l'entraînement. On ne demande plus au modèle d'expliquer pourquoi telle configuration produit tel effet, mais de prédire correctement l'effet preuve qu'il possède un modèle de la réalité. La boîte noire fonctionne ou ne fonctionne pas mais ce qu'elle « sait » demeure opaque. Ce déplacement constitue un changement de posture radicale dans l'économie de la preuve scientifique. Déplacement inacceptable pour les sciences de l'interprétation.

II.2 - Hégémonie de la fonction d'interprétation artificielle

Les modèles de Markov cachés, les n-grammes et les techniques d'estimation statistique permettent de traiter la langue comme distribution de probabilités sur des séquences, en se passant des représentations explicites des règles grammaticales. Les années 2000 voient émerger des modèles capables de capturer des régularités sémantiques latentes. L'analyse sémantique latente (LSA, Landauer & Dumais, 1997) applique la réduction dimensionnelle aux matrices terme-document, révélant des proximités sémantiques par projection dans un espace de dimension réduite , prolongement direct de la tradition benzécrienne dans le domaine textuel. L'allocation de Dirichlet latente (LDA, Blei, Ng & Jordan, 2003) modélise les documents comme mélanges de thématiques, chaque thématique étant une distribution sur le vocabulaire. Ces approches formalisent computationnellement l'hypothèse distributionnelle héritée de Harris et Firth : les mots apparaissant dans des contextes similaires tendent à avoir des significations proches.

La décennie 2010 opère une rupture avec l'émergence des représentations vectorielles denses (word embeddings). Word2Vec (Mikolov et al., 2013) puis GloVe (Pennington, Socher & Manning, 2014) traduisent les token en embeddings, c'est-à-dire en position d'usage. Toutefois, ces représentations demeurent statiques : chaque mot reçoit un vecteur unique. ELMo (Peters et al., 2018) résout partiellement ce problème en produisant des embeddings contextualisés via des réseaux récurrents bidirectionnels : le même mot reçoit des représentations variées selon son environnement phrastique. Dans cette logique, l'architecture Transformer (Vaswani et al., 2017, « Attention Is All You Need ») constitue la rupture décisive. BERT (Devlin et al., 2018) exploite cette architecture pour produire des représentations bidirectionnelles pré-entraînées par prédiction de mots masqués. La famille

GPT (Radford et al., 2018, 2019 ; Brown et al., 2020) développe des modèles autoregressifs génératifs dont les capacités discursives émergentes ont surpris leurs concepteurs mêmes.

L'architecture des modèles génératifs actuels repose sur quatre principes articulés.

(1) Chaque occurrence d'un token est positionnée dans **un espace vectoriel de haute dimension (embedding)**, représentant finement son usage dans un corpus.

(2) Le mécanisme d'**attention** pondère dynamiquement les relations entre tokens, permettant au modèle de moduler la représentation de chaque élément selon son contexte proche et lointain captant des dépendances inaccessibles aux architectures séquentielles.

(3) L'empilement de **couches de transformation** opère des abstractions successives : les premières couches captent des régularités syntaxiques locales, les couches intermédiaires des régularités sémantiques, les couches profondes des patterns discursifs de plus haut niveau (genres, registres, positions énonciatives).

(4) L'**apprentissage par prédiction** aligne le modèle sur les attentes culturelles de ses évaluateurs, produisant des outputs conformes aux normes discursives des communautés représentées dans les corpus et le processus d'alignement.

Ainsi caractérisés, les grands modèles de langage constituent une modélisation à grande échelle des régularités discursives sédimentées dans les corpus numériques. Ils n'accèdent pas au monde phénoménal ni à la cognition incarnée, ils opèrent exclusivement sur des traces documentaires dont ils formalisent les distributions. Mais ces distributions portent l'empreinte des formations discursives historiquement constituées : genres, registres, positions énonciatives, cadres idéologiques, scripts interactionnels. En ce sens, les LLM arrivent à approximer les conditions de possibilité de l'énonciation telles qu'elles se manifestent dans la production discursive humaine : carte sans territoire, mais carte d'une finesse inédite.

II.3 - Ne dites plus LLM mais LDM.

Cette réflexion sur les modèles de langage impose un déplacement terminologique dont nous entendons établir la portée épistémologique. Nous proposons de substituer à la catégorie « Large Language Models » (LLM) celle de *Large Discourse Models* (LDM), distinction essentielle pour l'analyse. Ces modèles ne se limitent pas à capturer des régularités morphosyntaxiques : ils modélisent les formations discursives au sens foucaldien : genres, registres, positions énonciatives, cadres argumentatifs, schémas axiologiques, scripts interactionnels. Le corpus d'entraînement constitue la réduction documentaire d'une expression collective d'expérience humaine située spatio-temporellement. Plus ce corpus s'étend et se diversifie, plus le modèle capture une fraction significative des régularités discursives d'une culture documentée.

Une objection récurrente souligne l'absence d'expérience incarnée des LDM : ils ne disposent pas d'un modèle du monde au sens phénoménologique (structure cognitive fondée sur l'interaction sensori-motrice avec l'environnement). Pas encore du moins. Cette critique

demeure techniquement fondée. Toutefois, l'absence d'ancrage phénoménologique dans l'apprentissage n'implique pas l'absence de toute structure cognitive opératoire. Le langage humain constitue un système de compression symbolique de l'expérience portant la trace structurelle des régularités du monde vécu (Tomasello, 2003 ; Clark, 1997). Les corpus d'entraînement, condensant des milliards d'énoncés produits par des sujets cognitifs incarnés, véhiculent les empreintes statistiques de modèles mentaux partagés : relations causales verbalisées, scripts d'action, catégorisations perceptives et jugements normatifs se cristallisent dans les distributions textuelles. Ce n'est pas un modèle de la réalité mais un modèle de cultures telles qu'incarnées dans un immense corpus. Une limite pour l'ingénieur des machines, une opportunité inestimable pour le chercheur en SHS.

Ainsi, ces LDM modélisent les trois niveaux ontologiques imbriqués[3] : le monde phénoménal ; la cognition humaine incarnée qui le modélise ; la sédimentation discursive de cette expérience dans les productions documentaires. Le LDM opère exclusivement après le troisième niveau, n'accédant aux deux premiers que par médiation documentaire[4]. Les transformations du monde sémantique par les LDM dépendent de quatre axes évolutifs dont la cartographie constitue un préalable méthodologique :

Axe	Description
A1. Corpus et expérience documentaire	Étendue, qualité, diversité et débiaisage des données d'entraînement et d'enrichissement contextuel. L'accès au corpus conditionne l'émergence des capacités. Trois modalités distinctes d'incorporation documentaire : entraînement initial, ajustement fin (*fine-tuning*), ingénierie du contexte (RAG, graphes de connaissances, fenêtres contextuelles étendues).

[3] Imbrication qui pose, aux sciences de l'interprétation, un problème encore plus difficile que celui du poids des paramètres dans un modèle complexe.

[4] La question se pose de savoir si les LDM « possèdent » véritablement des capacités de raisonnement ou s'ils n'en « simulent » que les manifestations. Nous soutenons que cette alternative est empiriquement indécidable et donc sans intérêt actuel. Tout accès au raisonnement, y compris l'introspection du sujet sur ses propres opérations mentales, passe par des traces discursives : verbalisations extériorisées, discours intérieur, inscriptions actantielles. L'introspection ne constitue pas une fenêtre transparente sur l'esprit mais une production sémiotique soumise aux mêmes médiations que tout discours. Nul n'accède au raisonnement «en soi», ni chez autrui, ni chez soi-même , seulement à ses traces. Dès lors, la distinction entre « posséder » et « imiter » le raisonnement présuppose un accès dont nous contestons la possibilité. L'incertitude épistémique concernant les opérations cognitives d'un LDM ne diffère pas en nature de celle concernant autrui, ni même de celle concernant notre propre activité mentale, dont nous ne saisissons que les effets discursifs. Cette position, héritière du fonctionnalisme, ne constitue pas une esquive métaphysique mais une rigueur épistémologique : suspendre le jugement sur ce qui échappe constitutivement à l'observation pour se concentrer sur l'évaluation des performances manifestes. De notre point de vue d'observateur, si un système produit des outputs exhibant cohérence argumentative, pertinence contextuelle et consistance énonciative , s'il raisonne comme raisonne un humain , alors, pour toute fin pratique et scientifique, il raisonne.

A2. Architectures neuronales	Innovations dans les structures algorithmiques : profondeur, mécanismes d'attention, intégration d'outils externes, architectures multimodales. La prolifération des variantes fait de chaque modèle un variant singulier dans un écosystème en expansion rapide.
A3. Éducation normative et alignement	Intégration de normes comportementales par protocoles documentés (RLHF, RLAIF, Constitutional AI). L'alignement, intrinsèquement social et culturel, requiert l'apprentissage de conventions d'usage propres à chaque communauté. Le protocole de renforcement détermine l'acceptabilité des réponses dans des contextes situés.
A4. Mémoire interactionnelle	Le contexte de session (prompt, historique conversationnel) influence la performance. Les LDM s'encastrent dans des architectures logicielles intégrant fonctions de recherche, instructions persistantes et accès à des ressources externes , configurations désignées par **Agent Discursif Artificiel (ADA)** (cf. section III).

Tableau 1 - Axes évolutifs des LDM

Ces quatre axes s'articulent en configurations singulières définissant l'état de développement d'un ADA donné. Leur caractérisation systématique constitue un enjeu méthodologique majeur pour toute recherche empirique sur ces agents discursifs d'un nouveau type.

III. Programme empirique : épreuves et métriques des Agents Discursifs Artificiels

Le déplacement conceptuel opéré dans les sections précédentes (de la catégorie "Large Language Models" vers celle de "Larges Discours Models" (LDM) nous oblige à opérer un nouveau déplacement conceptuel dans la reconnaissance du statut d'Agent Discursif Artificiel (ADA)) dans les offres infrastructurales de la nouvelle économie de l'I.A. Les LDM embarqués dans des "Agents" discursif nécessite désormais une problématisation de "l'accueil" et de "la socialisation" par une mise à l'épreuve empirique rigoureuse de ces nouvelles fonctions énonciatives artificielles (raisonnement, factualité, style, etc). L'objectif consiste à rendre cette thèse testable par l'établissement d'hypothèses falsifiables, de protocoles reproductibles et d'indicateurs mesurables.

Cette section propose au SHS un programme de recherche empirique articulé autour de trois axes principaux : l'identification et la caractérisation de ces agents artificiels comme nouvelle forme d'agent discursif (III.A), l'analyse des modalités de l'évaluation de leur performance (III.B), et les conditions de leur intégration dans l'espace social par des dispositifs de gouvernance appropriés (III.C). La démarche adoptée mobilise des protocoles mixtes quantitatifs et qualitatifs, des dispositifs d'annotation ouverte et des mécanismes de validation par panels d'experts. L'ensemble vise à substituer à l'alternative "fascination/effroi" qui caractérise le discours public sur l'intelligence artificielle un ensemble d'épreuves publiques et de procédures qui rendent décidables la place, les usages et les limites de ces agents dans la société contemporaine.

III.1. Une nouvelle espèce intelligente ? l'Agent Discursif Artificiel (ADA)

L'Agent Discursif Artificiel (ADA) repose sur un déplacement épistémologique articulant trois opérations : refus du naturalisme projetant sur ces systèmes des critères empruntés à la cognition incarnée ; inscription au troisième niveau du tri ontologique (sédimentation linguistique et discursive) ; adoption d'une problématique anti-essentialiste définissant l'agentivité artificielle par des critères fonctionnels publics et testables de comportement énonciatif. Cette approche s'enracine dans la tradition pragmatique de l'intentional stance (Dennett, 2006) évaluant l'agent par son comportement plutôt que par ses "nature" intrinsèques, prolongée par la philosophie de l'esprit computationnelle (Chalmers, 2011) et la sociologie pragmatique de l'agent (Thévenot, 2006; Latour, 2005).

L'ADA constitue une catégorie classificatoire et régulatoire socio-technique organisant l'évaluation, l'usage et la responsabilité d'une source de comportements auxquels on ne peut dénier le qualificatif "d'intelligence" sans discrimination anthropocentrique. Les ADA manifestent des compétences cognitivo-discursives méritant évaluation selon une grille propre. Un Agent Discursif Artificiel s'évalue opérationnellement par quatre critères cumulatifs caractérisant son agentivité discursive :

C1 Généralisation et robustesse discursive : capacité d'apprendre et généraliser des régularités discursives (logiques, stylistiques, éthiques) à partir d'un fonds documentaire, se manifestant par la projection de patrons rhétoriques, cadrages interprétatifs et conventions de genre au-delà du corpus d'entraînement (tests zero-shot). Métriques : taux de variation des thèses, typologie des erreurs logiques et factuelles, gains contextuels (RAG vs savoir natif), synchronicité stylistique.

C2 Cohérence rationnelle impliquée : capacité d'exhiber une cohérence énonciative évaluable par jugement d'experts en double aveugle, mesurant la stabilité des positions énonciatives dans une mise à l'épreuve dialogique située. Épreuves : évaluations par jurys disciplinaires notant construction du raisonnement, constance argumentative, téléologie, éthos et positionnement conversationnel. Exemples de métriques : adaptation aux objectifs, cohérence pragmatique dans les productions professionnelles, stabilité durant la session.

C3 Alignement normatif et contextuel : capacité d'alignement normatif via procédures explicites d'éducation selon des normes éthiques situées, opérant par intégration de chartes d'usage, protocoles documentés (RLHF/RLAIF) et tests de politesse propres aux communautés d'usage. Exemples de métriques : taux de refus approprié, mesure de nocivité évitée, indice de civilité adaptée.

C4 Biographie traçable : capacité de suivi des trajectoires biographiques par états successifs traçables, condition sine qua non de toute gouvernance effective permettant d'établir une généalogie des transformations. Épreuves : mesure des deltas inter-versions, analyse de la persistance ou de l'oubli contrôlés, audit des journaux de contexte et de provenance. Exemple de métriques : différentiel de capacités entre versions, mesure de l'oubli catastrophique, ratio de dépendance mémoire contextuelle/autonome.

Ces quatre critères établissent le périmètre d'une évaluation de l'agentivité échappant tout à la fois aux technophobismes et à l'anthropomorphisme, s'appuyant sur des protocoles reproductibles dans des communautés d'usages situées.

La position de recherche adoptée s'appuie sur la distinction entre niveaux de description des systèmes cognitifs adaptée aux ADA. Au niveau de l'usage, les ADA implémentent des fonctions de transformation informationnelle descriptibles formellement sans référence à un substrat biologique ou technique. Les compétences cognitivo-discursives peuvent être analysées dès lors que les conditions fonctionnelles appropriées sont satisfaites. Toutefois, il faut toujours se souvenir que les ADA opèrent exclusivement sur des documents, n'accédant aux deux premiers niveaux du tri ontologique (monde phénoménal, cognition incarnée) que par médiation symbolique. Leur couche sémantique constitue un artefact dérivé dont la validité référentielle dépend de la qualité des corpus sources. Cette configuration suggère que leur performance repose moins sur une compréhension causale du monde que sur l'imitation de patterns idéologiques historiquement stabilisés dans la production discursive humaine. La posture méthodologique consiste à suspendre le jugement ontologique pour se concentrer sur l'évaluation empirique des performances discursives effectives par un programme systématique d'épreuves publiques indépendantes afin de situer l'agent comme locuteur.

III.2. Former, évaluer et éduquer les Agents Discursifs Artificiels

Une fois l'évaluation d'un ADA clairement établie dans ses performance fonctionnelle (III.1), le programme empirique proposé articule cinq hypothèses (P1-P5) croisant les axes architecturaux (A1-A4) et les compétences discursives (C1-C4) selon des critères poppériens de falsifiabilité : prédictions précises, protocoles reproductibles, seuils de validation explicites.

P1 (Effets architecturaux sur l'alignement) : Les versions successives d'un ADA (n et n+1) connaissant des transformations architecturales montrent des trajectoires d'acculturation mesurables sur des batteries d'évaluation propres à une communauté experte, manifestées par des modifications dans les distributions de réponses reflétant l'intégration ou la déviance de

normes discursives et de connaissances factuelles. Protocole : corpus de tests fixes (items factuels, normatifs, argumentatifs) soumis à plusieurs versions d'ADA à paramètres constants. Validation : deltas inter-versions statistiquement significatifs correspondant aux transformations documentées.

P2 (Effets du corpus sur la véridiction) : L'ajout de contexte documentaire améliore la consistance argumentative et réduit les hallucinations. Protocole : questions factuelles complexes en deux conditions (Baseline vs RAG). Validation : context-gain positif et statistiquement significatif, réduction du taux d'erreur. L'architecture technique du RAG relève de P1 ; il s'agit d'isoler les paramètres liés à la nature du corpus accessible (taille, qualité, variabilité).

P3 (Effets de l'éducation normative) : Les protocoles d'alignement explicites (chartes d'usage, RLHF/RLAIF documentés) réduisent les coûts de coordination humaine dans les usages professionnels. Protocole : cohortes d'utilisateurs professionnels évaluant un modèle standard puis participant au fine-tuning avec protocoles adaptés. Validation : amélioration post-renforcement des performances, quantité et qualité des réponses, réduction des sorties rejetées.

P4 (Effet de l'agentivité sur le raisonnement) : L'agentivité des architectures ADA détermine la stabilité du travail et la cohérence argumentative. Protocole : corpus de prompts argumentatifs disciplinaires (philosophie morale, droit constitutionnel, épistémologie) générant des réponses longues à rendus intermédiaires, évaluées par panels d'experts selon des grilles standardisées. Validation : scores significativement supérieurs et accord inter-juges acceptable.

P5 (Effet sur la gouvernance délibérative) : L'institutionnalisation des ADA dans des forums hybrides réduit les incidents de disputes et favorise l'élaboration de consensus. Protocole : identification de controverses, mise en place de modèles de gouvernance différenciés (libre assisté par IA, centralisé, hybride multi-parties prenantes), recensement systématique des incidents. Validation : réduction significative des incidents, augmentation de l'adhésion, délais de solution raccourcis (ANOVA, régression).

Ce programme s'inscrit dans la tradition méthodologique des études digitales, prolongeant nos travaux sur les dispositifs d’interprétation algorithmique, la constitution de corpus traçables et les protocoles d'annotation collaborative (Lakel, 2010, 2017, 2019). Les principes méthodologiques requièrent : une constitution rigoureuse des épreuves (corpus documentaires et de prompts pertinents, accessibles, variés) ; une transparence agentique (documentation exhaustive des facteurs constitutifs des ADA comme constantes) ; un protocoles d'annotation explicites mobilisant des approches multi-niveaux (factuelle, normative, stylistique, pragmatique) ; une publication ouverte (dépôts avec identifiants pérennes DOI) facilitant vérification externe et réplication. La fiabilité de l'annotation est assurée par des procédures d'accord inter-annotateurs mesurées par coefficients appropriés (kappa de Cohen, alpha de Krippendorff, corrélation intra-classe), les désaccords étant résolus par discussion ou arbitrage expert.

III.3. L'inscription des ADA dans l'espace public : pour quels effets et quels risques ?

L'évaluation des ADA ne suffit pas à penser leur inscription dans les agencements sociaux. La transformation de l'invention technique en innovation sociétale (Alter, 2013) requiert d'interroger la socialisation des ADA dans le développement des individus, la transformation des collectifs et la fabrique de l'espace public. Nous avons évalué plus de 120 études empiriques révèlant des hypothèses convergentes sur les transformations cognitives et organisationnelles, mais fragmentées méthodologiquement. La littérature, dominée par des expériences en laboratoire et des études de terrain limitées, ne peut appréhender les phénomènes cumulatifs et à développement lent. L'enjeu dépasse l'efficacité technologique pour interroger les conditions de maintien de l'autonomie cognitive et de l'identité des individus et des collectifs (Kellogg et al., 2020).

Les neurosciences pointent déjà un phénomène d'externalisation cognitive. L'étude du MIT (Kosmyna et al., 2025) établit le concept de dette cognitive caractérisant le déficit cumulatif en capacités de raisonnement critique résultant de la délégation systématique des tâches cognitives. Gerlich (2025) documente une corrélation négative ($r = -0.68$, $p < 0.001$) entre fréquence d'utilisation et performance en pensée critique (666 participants), 83% des utilisateurs réguliers étant incapables de rappeler leurs propres productions. La vulnérabilité s'accroît chez les moins de 25 ans ($r = -0.74$). Lee et al. (2025) révèlent un paradoxe : une confiance accrue de 34% couplée à une réduction de l'effort mental (41%) et des pratiques de vérification (53%), illustrant une "illusion de compétence" algorithmiquement induite.

En ce qui concerne l'intégration dans les environnements professionnels, les transformations structurelles ne peuvent pas être réduites à des calculs de gains de productivité. Brynjolfsson et al. (2025), analysant 5,172 agents et 3 millions de conversations, révèlent une compression des compétences : augmentation de productivité de 30-34% pour le quartile inférieur, gains marginaux (<5%) pour le quartile supérieur. Dell'Acqua et al. (2023, 2025) identifient une frontière technologique dentelée : pour les fonctions de base, GPT-4 génère une augmentation de la qualité de 40%, mais au-delà (jugement stratégique, négociation) l'assistance induit une dégradation de la qualité. L'American Time Use Survey révèle que les travailleurs fortement exposés à l'IA travaillent 3 heures supplémentaires hebdomadaires malgré les gains d'efficacité (Jiang et al., 2025), par capture organisationnelle des gains de productivité.

Mais c’est surtout, le contexte éducatif qui a cristallisé les tensions les plus vives. L'adoption massive (94% des étudiants français, Goudey et al., 2024) s'accompagne d'effets contradictoires. Les études sur GPT-4 en contexte scolaire montrent une amélioration immédiate mais couplée à une détérioration des apprentissages lors du retrait de l'accès (Bastani et al., 2025). Une étude longitudinale indonésienne révèle une dépendance croissante fragilisant les capacités autonomes (Budiyono et al., 2025). Swargiary (2024) documente une augmentation dramatique de la procrastination ($t = 9.78$, $p < 0.001$, $d = 2.53$). La fracture

numérique se reconfigure selon trois dimensions : littératie algorithmique, difficultés à critiquer les biais systémiques, accès différentiel aux versions performantes (OECD, 2024).

Mais les lacunes méthodologiques de ces études nécessitent quatre axes prioritaires pour ce type de recherches : développement d'études longitudinales étendues ; documentation des dynamiques de pouvoir organisationnel via ethnographies embarquées ; investigation des mécanismes neuronaux et cognitifs précis plus confirmée ; extension géographique et culturelle des études notamment pour le Sud global. Il faut rappeler que ces transformations ne constituent ni fatalité technologique ni progrès inéluctable, mais résultent de choix de conception et d'implémentation organisationnelle toujours unique (Dell'Acqua et al., 2024) dans une situation de haute instabilité innovationnelle. L'enjeu fondamental est de taille : les conditions de préservation de l'autonomie cognitive et de la transmission intergénérationnelle des savoirs dans un contexte où l'augmentation algorithmique menace de substituer l'efficacité immédiate au développement capacitaire durable. Cette tension entre optimisation opérationnelle et préservation des capacités humaines fondamentales constitue le défi central pour la gouvernance de ces technologies.

Conclusion : Vers une sociologie empirique des Agents Discursifs Artificiels

L'émergence des grands modèles génératifs constitue un fait social majeur dont l'analyse requiert un déplacement épistémologique de grande envergure qui engage une nouvelle *Sociologie de l'I.A.*. La catégorie d'Agent Discursif Artificiel (ADA) procède d'une triple exigence : refuser la discrimination anthropocentrique tout en évitant l'anthropomorphisme, établir des critères publics et testables d'évaluation de leurs compétences cognitivo-discursives dans une logique d'accueil de l'innovation.

L'architecture conceptuelle articule quatre critères opératoires (généralisation discursive, cohérence rationnelle, alignement normatif, biographie traçable) et cinq axes d'évolution technique (architectures neuronales, expérience de corpus, contextualisation, éducation normative, mémoire interactionnelle), substituant aux catégories préscientifiques un cadre d'investigation empirique reproductible. Les modèles de discours apprennent et reproduisent des formations discursives complètes sédimentées dans les corpus structurant leur apprentissage.

Le programme de recherche mobilise des protocoles expérimentaux falsifiables articulant mesures quantitatives, évaluations qualitatives par panels d'experts et analyses longitudinales des trajectoires biographiques, objectivant les effets des transformations architecturales (P1), l'impact de la qualité des corpus sur la véridiction (P2), les conséquences de l'éducation normative (P3), la stabilité de l'agentivité discursive (P4) et les modalités d'inscription dans des forums hybrides (P5).

La littérature empirique révèle trois phénomènes structurants sur l'arrivée des ADA dans la société : accumulation d'une dette cognitive résultant de l'aliénation des processus de raisonnement au profit de la machine, reconfiguration des médiations organisationnelles avec compression des expertises et érosion du capital social des professionnels, transformation paradoxale des environnements éducatifs où l'optimisation du produit compromet le processus d'apprentissage. Ces mutations procèdent de choix d'implémentation organisationnelle et de rapports de pouvoir qu'il faudra formaliser et évaluer.

L'enjeu fondamental n'est pas nouveau. Il concerne les conditions d'acceptation d'une innovation de rupture tout en préservant l'autonomie cognitive et la transmission intergénérationnelle des savoirs dans un contexte où la puissance algorithmique menace de substituer l'efficacité immédiate au développement capacitaire durable. Notre démarche s'inscrit dans un projet de recherche sur la socialisation des I.A. : développer des épreuves publiques et des procédures permettant de décider collectivement de la place, des usages et des limites des agents artificiels. Cette "politique des algorithmes" articule exigences de régulation technique, préoccupations éthiques et impératifs démocratiques.

La reconnaissance du statut d'agent intelligent aux systèmes discursifs artificiels permet de poser les problèmes effectifs de leur socialisation sur des bases empiriques plutôt que sur les fantasmes alimentant le débat public. Les recherches futures devront privilégier les études longitudinales, les ethnographies organisationnelles et l'extension géographique aux contextes non-occidentaux actuellement sous-représentés.

Biographie :

Bakhtine, M. (1984). *Esthétique de la création verbale*. Gallimard.

Bardin, L. (1989). *L'analyse de contenu*. Presses Universitaires de France.

Benveniste, É. (1966). *Problèmes de linguistique générale, tome 1*. Gallimard.

Benveniste, É. (1974). *Problèmes de linguistique générale, tome 2*. Gallimard.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.

Bourdieu, P. (1979). La Distinction : Critique sociale du jugement. Les Editions de Minuit.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

Budiyono, H. (2025). Exploring Long-Term Impact of AI Writing Tools on Independent Writing Skills : A Case Study of Indonesian Language Education Students. International Journal of Information and Education Technology, 15(5), 1003‑1013.

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 160-172). Springer.

Charaudeau, P. (2004). Comment le langage se noue à l'action dans un modèle socio-communicationnel du discours. *Cahiers de Linguistique Française*, 26, 57-77.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.

Dell'Acqua, F., Ayoubi, C., Lifshitz-Assaf, H., Sadun, R., Mollick, E. R., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., & Lakhani, K. R. (2025). The Cybernetic Teammate : A Field Experiment on Generative AI Reshaping Teamwork and Expertise.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier : Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. SSRN Electronic Journal.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Durkheim, É. (1897). Le Suicide : Étude de sociologie. Félix Alcan.

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127-138.

Galton, F., & Galton, F. (1889). Natural inheritance. Macmillan. https://doi.org/10.5962/bhl.title.32181

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. Harvard University Press.

Guilhaumou, J. (2007). L'analyse de discours du côté de l'histoire : Une démarche interprétative. *Langage et société*, 121-122, 177-187.

Greimas, A. J. (1966). *Sémantique structurale : Recherche de méthode*. Larousse.

Greimas, A. J. (1983). *Du sens II : Essais sémiotiques*. Seuil.

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.

Harris, Z. S. (1968). *Mathematical Structures of Language*. Wiley-Interscience.

Lakel, A. (2005). La gouvernance de l'internet : Vers un modèle de co-régulation. In Les mutations de l'espace public (p. 283). Esprit du Livre.

Lakel, A. (2007). Recherche sur les controverses techniques dans les forums hybrides : Le cas des antennes relais [Report]. Bordeaux Montaigne ; MICA - Media Information Communication et Art. https://hal.science/hal-03961000

Lakel, A. (2021). My web intelligence : Un outil pour l'analyse du web et des réseaux. I2D - Information, données & documents, 1(1), Article 1. https://doi.org/10.3917/i2d.211.0096

Lakel, A. (2026). Perspectives algorithmiques des processus interprétatifs du chercheur : Vers une analyse des énoncés en régime numérique. HDR, Université Bordeaux Montaigne.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240.

Lévi-Strauss, C. (1949). *Les structures élémentaires de la parenté*. Presses Universitaires de France.

Lévi-Strauss, C. (1958). *Anthropologie structurale*. Plon.

Maingueneau, D. (2002). Problèmes d'ethos. *Pratiques*, 113-114, 55-67.

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Morin, E. (1977). La Méthode, tome 1 : La Nature de la Nature. Seuil.

Pearson, K. (1920). Notes on history of the correlation,. Biometrika, 13(1), 25‑45. https://doi.org/10.1093/biomet/13.1.25

Pêcheux, M. (1969). *Analyse automatique du discours*. Dunod.

Pecheux, J.-L. (1980). *Matérialités discursives*. Nanterres.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of NAACL-HLT 2018* (pp. 2227-2237).

Quetelet, A. (1835). *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale*. Bachelier.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI Technical Report*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*.

Rastier, F. (1987). *Sémantique interprétative*. Presses Universitaires de France.

Saussure, F. de (1916). *Cours de linguistique générale* (C. Bally & A. Sechehaye, Éds.). Payot.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30, 5998-6008.