

Chapitre 1

Introduction Générale

Sommaire du chapitre

1.1 Contexte : L'Enseignement de l'Histoire face aux Nouveaux Médias	1
1.2 Problématique et Positionnement de la Recherche.....	3
1.3 Questions de Recherche.....	5
1.4 Contributions de la Thèse	6
1.5 Structure du Manuscrit	7

1.1 Contexte : L'Enseignement de l'Histoire face aux Nouveaux Médias

L'enseignement de l'histoire au secondaire se heurte à une difficulté récurrente. Les élèves reconnaissent la pertinence de cette discipline : interrogés sur son utilité, ils évoquent spontanément la compréhension du présent, l'évitement des erreurs passées, la formation citoyenne. Cette reconnaissance coexiste pourtant avec un désengagement face au cours magistral, aux exercices de mémorisation, aux évaluations écrites répétitives. Le paradoxe traverse les enquêtes internationales : les élèves jugent l'histoire ennuyeuse dans sa forme scolaire habituelle, mais importante pour décrypter l'actualité ([Henkaline, 2023](#)).

Les enquêtes comparatives entre disciplines éclairent cette tension. L'histoire se classe en position intermédiaire dans les préférences des élèves, devancée par les disciplines requérant une activité concrète : éducation physique, technologie, arts plastiques ([Haydn](#)

[and Harris, 2010](#)). Ces matières impliquent une manipulation, une production tangible, un engagement corporel ou créatif que le cours d’histoire peine à offrir. Les variations observées entre établissements suggèrent néanmoins que le problème n’est pas inhérent à la discipline. Les écarts d’appréciation atteignent plusieurs dizaines de points de pourcentage selon les contextes, soulignant l’effet déterminant des choix pédagogiques.

La comparaison avec les disciplines scientifiques et techniques révèle une asymétrie supplémentaire. Les STIM bénéficient d’une perception d’utilité professionnelle que l’histoire ne peut revendiquer avec la même évidence ([Grever et al., 2011](#)). Les débouchés des filières scientifiques apparaissent clairement aux élèves et à leurs familles. L’histoire semble cantonnée à l’enseignement, à l’archéologie, aux métiers du patrimoine — des secteurs perçus comme moins attractifs économiquement. Cette hiérarchie implicite influence les choix d’orientation, malgré la transférabilité des compétences historiques : analyse critique des sources, argumentation nuancée, synthèse documentaire.

Les travaux en didactique de l’histoire convergent sur un point : les méthodes interactives modifient le rapport des élèves à la discipline. Jeux de rôle, débats structurés, simulations de procès, analyse de sources primaires génèrent un engagement que le cours transmissif ne suscite pas ([Straaten et al., 2016](#)). Les approches expérientielles — vidéos immersives, inclusion de figures historiques diverses — activent l’empathie historique et établissent des connexions explicites entre passé et présent ([Henkaline, 2023](#)). La manière d’enseigner détermine l’engagement davantage que le contenu.

L’émergence des technologies d’intelligence artificielle générative s’inscrit dans ce contexte. Les grands modèles de langage ouvrent une possibilité inédite : converser avec des représentations de figures historiques. La comparaison entre dialogues interactifs et lecture traditionnelle indique des effets positifs sur la motivation, bien que la contribution spécifique de l’interactivité reste à clarifier ([Pataranutaporn et al., 2023](#)). Les méta-analyses sur les agents virtuels alimentés par l’IA confirment ces résultats, avec des effets plus marqués pour les agents combinant apparence humanoïde et interactions structurées ([Dai et al., 2024](#)).

Les technologies de synthèse vocale et d’animation faciale complètent ce dispositif. Elles incarnent visuellement et vocalement le personnage. Le « témoin virtuel » du passé devient une réalité technique. Un élève peut poser des questions à Jules César et obtenir des réponses immédiates, formulées à la première personne. Cette possibilité suscite un

enthousiasme compréhensible au regard des difficultés documentées de l’enseignement traditionnel.

L’enthousiasme appelle toutefois à la prudence. La littérature sur les technologies éducatives en histoire souffre de limitations méthodologiques récurrentes : échantillons restreints, durées d’intervention courtes, effet de nouveauté non contrôlé, prédominance des mesures auto-rapportées (Veletsianos and Russell, 2014). Les études en contexte écologique — en classe réelle, avec des contraintes institutionnelles authentiques — demeurent rares. La fluidité même qui rend ces agents attrayants pourrait constituer leur talon d’Achille pédagogique. Quand la technologie se fait oublier, l’agent cesse d’être perçu comme un outil. L’élève risque alors de lui accorder une confiance que l’interaction ne justifie pas.

1.2 Problématique et Positionnement de la Recherche

La convergence des grands modèles de langage et des médias synthétiques permet de créer des agents pédagogiques d’un réalisme saisissant. Ces agents combinent génération de discours en temps réel, incarnation visuelle et interaction orale naturelle. Cette configuration répond aux facteurs identifiés comme efficaces dans l’enseignement de l’histoire : interactivité, personnification, dimension narrative.

Les premières études sur ces dispositifs confirment l’intérêt des collégiens pour les agents incarnant des personnages historiques (Kim et al., 2025). Des plateformes de simulation permettent d’explorer des environnements passés via des cartes interactives et de dialoguer avec des personnages générés par l’IA. Les enseignants peuvent créer des scénarios alignés sur leurs objectifs pédagogiques (Park et al., 2025). Le paradigme CASA (*Computers Are Social Actors*) éclaire ce phénomène : les individus appliquent spontanément aux technologies interactives les règles sociales qui régissent les relations humaines, créant un sentiment de présence qui transforme l’apprentissage en expérience relationnelle (Nass et al., 1994).

Cette promesse se double cependant d’un risque que la recherche commence à peine à explorer. L’intérêt déclaré ne garantit pas l’apprentissage effectif. Un élève captivé par l’échange conversationnel peut surestimer ce qu’il en retire. La clarté apparente

des explications fournies par l’agent, optimisées pour la lisibilité par les algorithmes sous-jacents, pourrait induire une confiance excessive dans sa propre compréhension. Ce phénomène est documenté en psychologie cognitive sous le terme d’illusion de compréhension ([Rozenblit and Keil, 2002](#)). Il désigne la tendance à surestimer la profondeur de ses propres connaissances jusqu’à ce qu’une tâche d’explication révèle les lacunes.

Le mécanisme sous-jacent est l’heuristique de fluidité : une information facile à traiter est perçue comme familière, et cette facilité est attribuée par erreur à sa propre maîtrise du sujet plutôt qu’aux qualités de la présentation ([Reber and Schwarz, 1999](#)). L’effet de fluidité de l’instructeur (*instructor fluency*) illustre ce biais : les comportements non verbaux d’un enseignant dynamique et fluide biaisent les jugements d’apprentissage des élèves, qui confondent la qualité de la prestation avec la qualité de leur propre apprentissage ([Toftness et al., 2018](#)). Les agents génératifs maximisent cette fluidité par leur capacité à produire un discours instantané, structuré et linguistiquement impeccable.

La présente recherche se positionne à l’intersection de ces enjeux. Elle relève du champ de l’Interaction Homme-Machine, mais s’ancre résolument dans le terrain éducatif. Son originalité tient à trois caractéristiques. Elle mobilise d’abord des protocoles expérimentaux contrôlés en contexte écologique — des classes réelles, pendant les heures de cours, avec des contraintes institutionnelles authentiques. Elle examine ensuite non seulement les effets sur l’intérêt, mais aussi les risques métacognitifs associés à l’interaction avec des agents génératifs. Elle couvre enfin un spectre développemental étendu, de la sixième à la terminale, permettant d’observer comment l’âge module les effets observés.

Cette double focale — engagement et vigilance critique — répond à une lacune de la littérature. Les travaux existants privilégient les mesures d’engagement, de satisfaction, de motivation ([Wu and Yu, 2024](#)). La question métacognitive demeure sous-explorée, alors même que les spécificités des LLM (génération fluide, hallucinations plausibles) la rendent particulièrement saillante. La recherche interroge donc cet équilibre : comment concevoir des agents qui engagent sans désactiver la conscience critique de l’outil ?

1.3 Questions de Recherche

Cette recherche s’articule autour de trois questions complémentaires, organisées selon une progression dialectique : de l’exploration des bénéfices potentiels vers l’examen des risques associés.

QR1 : Dans quelle mesure l’interactivité orale directe avec un agent historique alimenté par l’IA influence-t-elle l’intérêt des élèves par rapport à une présentation vidéo passive ?

Cette première question teste l’hypothèse issue de la littérature : l’interactivité constitue-t-elle un levier d’engagement supérieur aux formats passifs, même lorsque le contenu informationnel est équivalent ? Elle opérationnalise la distinction entre les modes « passif » et « interactif » du cadre ICAP ([Chi and Wylie, 2014](#)) en contexte d’agent conversationnel. La théorie de l’agence sociale prédit que les indices sociaux émis par un agent virtuel activent des réponses sociales chez l’apprenant, favorisant un traitement cognitif plus approfondi ([Moreno et al., 2001](#); [Mayer and DaPra, 2012](#)).

QR2 : Comment l’alignement thématique de l’agent (personnage historique vs. neutre ; pair vs. autorité) et son style de présentation (formel vs. accessible) modulent-ils cet intérêt en fonction du stade développemental des élèves ?

Cette question explore les conditions aux limites de l’effet d’interactivité. Elle examine si les caractéristiques de l’agent — son identité, son style communicationnel — interagissent avec l’âge des apprenants pour produire des effets différenciés. L’alignement thématique entre l’agent et le contenu pédagogique pourrait renforcer l’expérience d’apprentissage en la rendant plus attractive et stimulante ([Schmidt et al., 2019](#)). La théorie du développement de l’intérêt distingue l’intérêt situationnel déclenché (suscité par des éléments nouveaux ou surprenants) de l’intérêt situationnel maintenu (entretenu par la pertinence personnelle et l’engagement significatif) ([Hidi and Renninger, 2006](#); [Renninger and Hidi, 2015a](#)). L’interaction avec un personnage historique pourrait activer ces deux mécanismes différemment selon l’âge.

QR3 : L’apparence de l’agent (humanoïde vs. abstrait) influence-t-elle la propension des élèves à l’illusion de compréhension, leur confiance et la crédibilité perçue des informations délivrées ?

Cette troisième question déplace l’analyse du versant motivationnel vers le versant

métacognitif. Elle interroge le revers potentiel de l’engagement : la fluidité de l’interaction conduit-elle les élèves à surestimer leur compréhension ? Le design visuel de l’agent amplifie-t-il ce risque ? Le protocole IOED (*Illusion of Explanatory Depth*) permet de mesurer l’écart entre confiance subjective et performance objective. L’hypothèse est que l’incarnation humanoïde, en renforçant la présence sociale et l’autorité perçue, pourrait amplifier l’illusion de compréhension par rapport à une représentation abstraite.

1.4 Contributions de la Thèse

Cette recherche apporte quatre types de contributions.

Contribution empirique. Elle fournit des données quantitatives issues de quatre études impliquant 458 élèves répartis sur quatre niveaux scolaires (sixième, cinquième, quatrième, terminale). Ces études ont été conduites en contexte écologique, intégrées au programme scolaire, avec des mesures pré/post et des designs expérimentaux contrôlés. Ce corpus constitue l’une des premières investigations systématiques des agents conversationnels historiques alimentés par LLM en classe réelle dans le contexte français. Les résultats éclairent l’effet de l’interactivité sur l’intérêt à différents stades développementaux, ainsi que l’influence du design de l’agent sur les processus métacognitifs.

Contribution théorique. Elle interroge les mécanismes psychologiques à l’œuvre dans l’interaction élève-agent à l’ère des modèles génératifs. Les résultats invitent à reconsidérer le paradigme CASA : la fluidité conversationnelle des LLM semble reconfigurer la hiérarchie des indices sociaux, la qualité du discours pouvant éclipser l’effet de l’incarnation visuelle. La thèse contribue également à la compréhension de l’illusion de compréhension dans les contextes d’apprentissage médiatisés par l’IA ([Paik and Schraw, 2013](#)). Elle articule ce biais métacognitif avec les spécificités techniques des agents génératifs (fluidité maximisée, absence de marqueurs d’incertitude).

Contribution méthodologique. Elle présente MemorIA, une architecture technique modulaire permettant le déploiement d’agents conversationnels historiques en milieu scolaire. Cette plateforme intègre reconnaissance vocale, génération de langage, synthèse vocale et animation faciale dans un pipeline temps réel. L’authenticité historique et le contrôle comportemental des agents constituent des préoccupations centrales pour ce type de système : l’exactitude factuelle doit guider le développement pour

éviter une représentation simplifiée du passé (DaCosta, 2025). La recherche propose également une adaptation du protocole IOED aux populations scolaires et aux contextes d'interaction avec agents génératifs, incluant des phases de production d'explications écrites et des mesures de confiance avant et après interaction.

Contribution pratique. Elle formule des recommandations pour la conception d'agents pédagogiques qui engagent les élèves sans désactiver leur vigilance critique. Les résultats suggèrent que l'interactivité constitue un levier d'engagement robuste, indépendant du niveau scolaire. L'effet de la représentation de l'agent apparaît plus nuancé, modulé par le développement cognitif et les attentes des élèves. Ces observations alimentent une réflexion sur l'équilibre entre engagement et conscience de l'outil, nécessaire à une intégration pédagogique responsable de ces technologies.

1.5 Structure du Manuscrit

Le manuscrit s'organise en sept chapitres.

Le **chapitre 2** établit le cadre théorique de la recherche. Il articule les théories de la motivation et de l'intérêt (théorie de l'autodétermination, modèle du développement de l'intérêt), les spécificités de l'enseignement de l'histoire (comparaison avec les STIM, perception des élèves, pratiques pédagogiques), le paradigme CASA et la théorie de l'agence sociale, les mécanismes de l'illusion de compréhension, pour aboutir à la formulation des hypothèses.

Le **chapitre 3** présente la plateforme expérimentale MemorIA : son architecture technique, ses modules (reconnaissance vocale, génération de langage, synthèse vocale, animation faciale), et sa validation lors d'une étude pilote. Il expose les contraintes de conception liées au déploiement en milieu scolaire et les choix techniques effectués.

Le **chapitre 4** expose l'Étude 1, une série de trois expérimentations examinant l'impact de l'interactivité et de la représentation de l'agent sur l'intérêt des élèves, à trois niveaux scolaires distincts (sixième, quatrième, terminale). Chaque expérimentation est présentée avec sa méthodologie, ses résultats et sa discussion spécifique. Une synthèse croisée conclut le chapitre.

Le **chapitre 5** présente l'Étude 2, qui examine l'influence du design visuel de l'agent (humanoïde vs. abstrait) sur l'illusion de compréhension chez des élèves de cinquième. Le

protocole IOED adapté est détaillé, ainsi que les mesures complémentaires de crédibilité, confiance et anthropomorphisme perçu.

Le **chapitre 6** propose une discussion générale synthétisant les résultats des deux études, leurs implications théoriques et pratiques, ainsi que les limites de la recherche. Il examine la tension entre engagement et vigilance critique, et formule des recommandations pour la conception d’agents pédagogiques.

Le **chapitre 7** conclut le manuscrit en résumant les contributions et en ouvrant des perspectives de recherche future, notamment sur les études longitudinales, les modalités alternatives d’interaction, et les questions éthiques soulevées par l’utilisation d’agents IA personnifiant des figures historiques.

Chapitre 2

État de l'Art

Sommaire du chapitre

2.1 Cadre Théorique de l'Apprenant : Motivation, Engagement et Pertinence.....	10
2.1.1 Dynamiques Motivationnelles : SDT et CET.....	11
2.1.2 Architecture de l'Intérêt	12
2.1.3 Pertinence, Valeur et Personnalisation de l'Apprentissage....	15
2.1.4 L'Apprentissage Actif et le Cadre ICAP.....	18
2.2 Le Contexte Spécifique de l'Enseignement de l'Histoire	20
2.2.1 Épistémologie et Comparaison Interdisciplinaire.....	20
2.2.2 La Perception des Élèves : le Paradoxe « Ennuyeux mais Important »	21
2.2.3 Pratiques Pédagogiques : du Transmissif à l'Actif.....	23
2.2.4 Technologies Éducatives en Histoire : Bilan Empirique.....	24
2.3 Les Agents Pédagogiques Virtuels : Typologie, Signaux Sociaux et Efficacité	26
2.3.1 Cartographie des Agents Pédagogiques : 25 Ans d'Évolution	27
2.3.2 Fondements Cognitifs : Contraintes et Principes de Design .	28
2.3.3 De l'Artefact au Partenaire : Mécanismes de la Présence Sociale	30
2.3.4 Taxonomie et Efficacité des Signaux Sociaux.....	31
2.3.5 Limites du Réalisme : Le Paradoxe de l'Hyperréalisme	34
2.4 La Rupture Technologique des IA Génératives en Éducation	36

2.4.1 Des Agents Scriptés aux Agents Génératifs : le Saut Qualitatif	
37	
2.4.2 La Personnalisation en Temps Réel.....	38
2.4.3 Fiabilité et Hallucinations : le Défi Épistémique	39
2.5 Le Phénomène de l'Illusion de Compréhension	40
2.5.1 Métacognition et Calibration de la Confiance	40
2.5.2 L'Heuristique de Fluidité.....	42
2.5.3 L'Autorité de l'Agent et la Vigilance Épistémique	43
2.6 Synthèse et Problématique	45
2.6.1 Convergence et Tensions Théoriques.....	45
2.6.2 Les Lacunes de la Littérature Actuelle.....	46
2.6.3 Questions et Hypothèses de Recherche	48

Ce chapitre établit le cadre théorique nécessaire à l'analyse de l'interaction entre élèves et agents virtuels historiques. La structure suit une progression logique : elle part des mécanismes internes de l'apprenant (section ??), examine le contexte disciplinaire spécifique de l'enseignement de l'histoire (section 2.2), analyse la nature de l'interaction avec l'agent (section ??), explore la rupture technologique des IA génératives (section ??), pour aboutir au risque métacognitif central de la thèse (section 2.5). Une synthèse finale articule ces éléments en une problématique de recherche cohérente (section ??).

2.1 Cadre Théorique de l'Apprenant : Motivation, Engagement et Pertinence

Cette première section établit les fondations théoriques nécessaires à l'analyse des mécanismes psychologiques qui sous-tendent l'engagement des élèves dans les apprentissages. Nous examinons successivement les théories motivationnelles qui éclairent les conditions de l'engagement intrinsèque (§2.1.1), l'architecture développementale de l'intérêt (§2.1.2), les processus par lesquels les élèves attribuent de la valeur aux contenus d'apprentissage et les stratégies de personnalisation (§2.1.3), et les niveaux d'engagement cognitif différenciés par le cadre ICAP (§2.1.4).

Ces cadres théoriques, issus de la psychologie de l'éducation, fournissent les outils

conceptuels qui seront mobilisés dans les sections suivantes pour analyser les spécificités de l'enseignement de l'Histoire (§2.2) et pour comprendre les mécanismes d'interaction avec les agents conversationnels (§??).

2.1.1 Dynamiques Motivationnelles : SDT et CET

La motivation intrinsèque — cette propension à s'engager dans une activité pour le plaisir et la satisfaction qu'elle procure — constitue un prédicteur robuste de la qualité de l'apprentissage (Deci and Ryan, 2000a). Sa compréhension nécessite d'articuler deux niveaux d'analyse : les mécanismes par lesquels l'environnement affecte cette motivation, et les besoins psychologiques sous-jacents qui la conditionnent.

Au premier niveau, les facteurs environnementaux n'exercent pas d'effet direct et uniforme sur la motivation ; leur impact dépend de l'interprétation qu'en fait l'individu (Deci and Ryan, 1985). Un même feedback peut ainsi être vécu comme *informationnel* — fournissant un retour constructif sur la compétence et soutenant la motivation — ou comme *contrôlant* — exerçant une pression sur le comportement et sapant l'autodétermination. Cette distinction, issue de la Théorie de l'Évaluation Cognitive (CET), s'articule autour de deux dimensions psychologiques : le *locus de causalité perçu*, soit le sentiment que ses actions émanent de soi plutôt que de contraintes externes, et le *sentiment de compétence*, soit la conviction de pouvoir atteindre les résultats souhaités. Un troisième aspect, *amotivant*, peut signaler l'incompétence et conduire au désengagement total.

Au second niveau, trois besoins psychologiques fondamentaux conditionnent le bien-être et l'engagement : l'*autonomie*, la *compétence* et l'*affiliation* (?). La Théorie de l'Autodétermination (SDT) postule que les environnements d'apprentissage favorisent l'intérêt lorsqu'ils satisfont ces besoins à travers des choix significatifs, des défis appropriés, un feedback constructif et des interactions soutenantes. L'ajout du besoin d'affiliation — le sentiment de connexion aux autres — étend la portée explicative du modèle au-delà des seules dimensions cognitives pour intégrer la dimension sociale de l'apprentissage.

L'articulation de ces deux niveaux révèle une dynamique complexe : un environnement qui offre des choix (autonomie), propose des défis calibrés avec un feedback informatif (compétence), et crée des opportunités d'interaction authentique (affiliation), réunit les conditions propices au développement de la motivation intrinsèque. À l'inverse,

un environnement perçu comme contrôlant, qui impose des tâches sans en expliciter le sens et isole l'apprenant, tend à éroder cette motivation — un phénomène particulièrement documenté lors de la transition vers le secondaire (Gnambs and Hanfstingl, 2016). Cette période, marquée par une pression accrue sur la performance et des structures scolaires qui peuvent limiter la curiosité naturelle (Engel, 2011), voit la motivation intrinsèque décliner significativement.

Ce déclin revêt une importance particulière pour l'enseignement de l'Histoire, discipline souvent perçue comme distante et déconnectée des expériences personnelles (Audi-gier and Fink, 2010). Le défi consiste alors à concevoir des environnements qui, tout en respectant les contraintes curriculaires, créent les conditions de satisfaction des besoins identifiés. Les agents conversationnels offrent une piste prometteuse : l'interactivité dialogique peut soutenir l'autonomie en permettant à l'élève de diriger l'échange, les réponses adaptatives peuvent nourrir le sentiment de compétence, et les indices sociaux de l'agent peuvent répondre au besoin d'affiliation. Cette hypothèse, qui guide notre programme expérimental, nécessite cependant de comprendre comment l'intérêt se développe et se maintient — objet de la section suivante.

2.1.2 Architecture de l'Intérêt

Si la motivation intrinsèque décrit une orientation générale vers l'action autodéterminée, l'intérêt désigne un état psychologique plus spécifique, focalisé sur un contenu particulier et caractérisé par une double composante : affective (l'expérience émotionnelle positive associée à l'activité) et cognitive (la motivation à approfondir la compréhension du sujet) (Renninger and Hidi, 2015b). Cette distinction conceptuelle s'avère cruciale : contrairement à la motivation, qui peut être extrinsèque ou intrinsèque selon sa source, l'intérêt implique toujours une relation significative entre l'individu et un domaine de connaissance.

L'intérêt n'est pas un état binaire mais un construit développemental qui évolue sur un continuum (Hidi and Renninger, 2006). À une extrémité, l'*intérêt situationnel* émerge comme réponse transitoire à des déclencheurs environnementaux — nouveauté, surprise, discordance avec les attentes. À l'autre extrémité, l'*intérêt individuel* constitue une disposition stable, internalisée, qui pousse à rechercher activement des occasions de réengagement avec le domaine (Priniski et al., 2018). La transition de l'un à l'autre

représente le passage d'un engagement soutenu de l'extérieur vers un engagement auto-entretenu.

Trajectoire développementale de l'intérêt

Cette transition s'opère selon une séquence en quatre phases qui permet de comprendre comment un intérêt éphémère peut se cristalliser en disposition durable ([Hidi and Renninger, 2006](#)). La première phase, l'*intérêt situationnel déclenché*, correspond à une capture attentionnelle initiale provoquée par des stimuli saillants — informations inattendues, présentations nouvelles, activités qui remettent en question les présupposés. Cette attention reste temporaire et peut s'évanouir rapidement si elle n'est pas nourrie.

La deuxième phase, l'*intérêt situationnel maintenu*, marque un premier approfondissement : l'attention devient soutenue à travers un engagement significatif avec le contenu. L'environnement d'apprentissage joue ici un rôle déterminant : un équilibre entre structure et autonomie ([Jang et al., 2010](#)), combiné à des opportunités de participation active ([Cents-Boonstra et al., 2021](#)), fournit le soutien nécessaire au maintien de l'intérêt. Sans ces conditions, l'intérêt déclenché retombe.

Les troisième et quatrième phases — *intérêt individuel émergent* puis *développé* — marquent l'internalisation progressive de l'intérêt. L'élève commence à générer ses propres questions, à rechercher des informations en dehors du contexte scolaire, à développer une expertise qui, à son tour, alimente l'intérêt. À ce stade, le soutien externe devient moins nécessaire : l'individu assume un rôle actif dans le maintien de son propre intérêt.

Déclencheurs de l'intérêt : facteurs individuels et situationnels

Si le modèle développemental décrit la trajectoire de l'intérêt, il reste à identifier les leviers qui permettent de le déclencher et de le maintenir. Une taxonomie distingue à cet effet les *facteurs individuels*, caractéristiques propres à l'apprenant, des *facteurs situationnels*, éléments de l'environnement d'apprentissage ([Bergin, 1999](#)).

Les facteurs individuels déterminent la réceptivité de l'élève aux stimuli environnementaux. Parmi ceux-ci, l'*appartenance* (*belongingness*) — incluant la valeur culturelle accordée au contenu, l'identification à des modèles, et le soutien social perçu — résonne avec le besoin d'affiliation identifié par la SDT. De même, le *sentiment de compétence*

et la *pertinence par rapport aux objectifs personnels* prolongent les dimensions d'autonomie et de compétence. Les *connaissances préalables* permettent d'ancrer les nouvelles informations dans des schémas existants, facilitant l'engagement cognitif. Les *émotions* associées au sujet, enfin, rappellent la dimension affective constitutive de l'intérêt.

Les facteurs situationnels concernent les caractéristiques de l'environnement susceptibles de déclencher l'intérêt : les *activités pratiques*, la présence d'un *auteur visible* qui personnalise le contenu, la *modélisation* par des pairs ou des figures d'autorité, les *jeux et puzzles*, la *discordance* avec les attentes préalables, la *nouveauté*, l'*interaction sociale*, et — particulièrement pertinent pour notre propos — le *narratif*. La structure narrative, en engageant les processus d'identification et en créant une tension dramatique, constitue un puissant déclencheur d'intérêt situationnel.

Cette articulation entre facteurs individuels et situationnels éclaire le potentiel des agents conversationnels incarnant des personnages historiques. Un tel dispositif combine plusieurs facteurs situationnels — nouveauté, interaction sociale, narratif, auteur visible — tout en pouvant activer des facteurs individuels à travers le dialogue — identification au personnage, pertinence personnelle, connexion aux connaissances préalables. La question devient alors : comment orchestrer ces facteurs pour favoriser la transition vers un intérêt maintenu, puis individuel ?

Le rôle pivot de la valeur perçue

Un mécanisme clé de cette transition réside dans l'évolution de la *valeur perçue* du contenu ([Hidi and Renninger, 2006](#)). Pour progresser de l'intérêt situationnel vers l'intérêt individuel, l'apprenant doit percevoir une valeur croissante dans le domaine — valeur qui motive la poursuite de l'engagement au-delà de l'attrait initial de la nouveauté. Cette proposition établit un pont conceptuel entre la théorie de l'intérêt et la théorie de l'attente-valeur (voir section [2.1.3](#)) : les deux cadres convergent pour prédire qu'augmenter la valeur perçue constitue un levier viable pour promouvoir l'intérêt et la motivation ([Harackiewicz et al., 2014](#)).

Cette convergence théorique suggère une stratégie d'intervention : si la valeur perçue est le médiateur de la progression développementale de l'intérêt, alors les interventions visant à augmenter cette valeur — notamment par la pertinence personnelle et la personnalisation — pourraient faciliter la transition de l'intérêt situationnel vers l'intérêt

individuel. C'est cette hypothèse que nous explorons dans la section suivante.

2.1.3 Pertinence, Valeur et Personnalisation de l'Apprentissage

La section précédente a établi que la valeur perçue constitue un mécanisme pivot dans la transition de l'intérêt situationnel vers l'intérêt individuel. Cette valeur ne relève pas d'une propriété intrinsèque du contenu ; elle émerge de la relation que l'apprenant établit entre ce contenu et ses préoccupations personnelles. La théorie de l'attente-valeur formalise cette intuition en postulant que la motivation à s'engager dans une tâche dépend de deux facteurs multiplicatifs : les *attentes de succès* (croyance en sa capacité de réussir) et la *valeur* attribuée à la tâche (??).

La valeur subjective se décompose en quatre dimensions qui éclairent différentes facettes de la pertinence. La *valeur d'accomplissement* renvoie à l'importance de la tâche pour l'identité personnelle — réussir confirme une image de soi valorisée. La *valeur intrinsèque* correspond au plaisir tiré de l'activité elle-même, rejoignant la dimension affective de l'intérêt. La *valeur d'utilité* concerne la pertinence pour les objectifs futurs — le contenu est perçu comme un moyen vers une fin désirée. Enfin, le *coût* perçu (effort, anxiété, opportunités sacrifiées) vient moduler négativement ces valeurs positives. Cette décomposition révèle pourquoi l'enseignement de l'Histoire souffre souvent d'un déficit d'engagement : la discipline peine à démontrer sa valeur d'utilité immédiate, contrairement aux disciplines STIM dont les applications apparaissent plus évidentes ([Harackiewicz et al., 2016](#)).

De la valeur d'utilité à la pertinence personnelle

La valeur d'utilité, si elle constitue un levier d'intervention accessible, ne représente qu'une facette de la pertinence. Au-delà de la question instrumentale « à quoi ça sert ? », une question plus profonde émerge : « en quoi cela me concerne-t-il ? ». Cette *pertinence personnelle* (*self-relevance*) implique que le contenu entre en résonance avec l'identité, les expériences et les préoccupations de l'apprenant ([Priniski et al., 2018](#)).

Une conceptualisation multidimensionnelle distingue ainsi la pertinence *personnelle* (connexion au soi) de la pertinence *impersonnelle* (utilité pour des entités externes), et la pertinence *appliquée* (utilité pour des actions concrètes) de la pertinence *conceptuelle*

(aide à la compréhension du monde) ([Albrecht and Karabenick, 2018](#)). Cette taxonomie explique pourquoi une même intervention peut fonctionner pour certains élèves et échouer pour d'autres : un élève sensible à la pertinence appliquée (« cela m'aidera dans mon métier ») et un autre répondant à la pertinence conceptuelle (« cela m'aide à comprendre l'actualité ») nécessitent des approches différenciées.

Stratégies d'intervention sur la pertinence

Les interventions visant à augmenter la pertinence perçue se distinguent par le niveau d'effort cognitif qu'elles requièrent de l'apprenant ([Albrecht and Karabenick, 2018](#)). À faible effort, la *communication directe* — l'enseignant explique pourquoi le contenu est pertinent — et la *personnalisation* — le contenu est adapté aux intérêts déclarés — représentent des approches accessibles mais dont l'impact reste limité. L'information fournie de l'extérieur peut être ignorée ou rejetée si elle ne résonne pas avec les schémas existants de l'apprenant.

Les approches à effort modéré, comme la *réflexion critique*, invitent les élèves à examiner leurs propres croyances sur la pertinence du contenu. Cette introspection peut révéler des connexions insoupçonnées, mais elle reste circonscrite au répertoire cognitif préexistant de l'apprenant.

Les approches à effort élevé — *auto-génération* et *réévaluation dirigée* — demandent aux élèves de produire activement des connexions entre le contenu et leur vie, ou de reconsidérer la valeur du contenu à la lumière de nouvelles perspectives. Les interventions de valeur d'utilité, où les élèves rédigent des essais connectant le contenu académique à leur expérience personnelle, illustrent cette approche ([Harackiewicz et al., 2014](#)). L'effort cognitif investi dans la génération de connexions semble renforcer leur impact motivationnel — un résultat qui résonne avec la hiérarchie des modes d'engagement que nous examinerons dans la section suivante (cadre ICAP).

Une nuance importante émerge cependant : ces interventions bénéficient particulièrement aux élèves présentant de faibles attentes de succès initiales, pour qui la découverte de connexions personnelles peut transformer la perception du contenu. Pour les élèves avec de fortes attentes de succès, l'intervention peut paradoxalement détourner l'attention de stratégies d'apprentissage déjà efficaces ([Harackiewicz et al., 2014](#)). Cette interaction entre intervention et profil de l'apprenant souligne la nécessité d'une

approche différenciée.

La personnalisation comme vecteur de pertinence

La personnalisation de l'apprentissage — l'adaptation du contenu et des activités aux intérêts, connaissances préalables et préférences individuels — constitue une stratégie prometteuse pour développer la pertinence à grande échelle ([Walkington and Bernacki, 2018](#)). L'intégration d'informations personnelles dans les contextes d'apprentissage (prénom de l'élève, centres d'intérêt déclarés) peut augmenter la motivation ([Cordova and Lepper, 1996](#)). Plus remarquable encore, offrir un contrôle sur des aspects même non essentiels de l'apprentissage (choix d'un avatar, d'un thème visuel) produit des effets positifs — suggérant que le sentiment d'appropriation personnelle contribue à l'engagement indépendamment du contenu lui-même. Ce résultat établit un lien avec le besoin d'autonomie identifié par la SDT : le choix, même symbolique, nourrit le sentiment d'autodétermination.

En mathématiques et en sciences, permettre aux élèves de choisir des exemples connectant les concepts abstraits à leurs intérêts renforce l'engagement, particulièrement chez ceux présentant initialement un faible intérêt ([Høgheim and Reber, 2015](#); [Reber et al., 2018](#)). Cette stratégie crée des ponts entre concepts abstraits et expériences concrètes, facilitant l'ancrage des nouvelles connaissances dans les schémas existants — retrouvant ainsi le facteur « connaissances préalables » de la taxonomie de l'intérêt. Placer la curiosité des élèves au centre de l'apprentissage peut également améliorer l'engagement et les relations pédagogiques ([Hagay and Baram-Tsabari, 2015](#)).

Défis et perspectives technologiques

La personnalisation se heurte cependant à des obstacles pratiques : diversité des intérêts, évolution constante des préférences, complexité de créer du contenu adapté pour chaque profil ([Walkington and Bernacki, 2018](#)). Ces contraintes ont longtemps limité le déploiement à grande échelle d'approches véritablement individualisées.

Les grands modèles de langage (LLM) offrent de nouvelles perspectives en permettant une adaptation dynamique et contextuelle (??). Un agent conversationnel incarnant un personnage historique peut ainsi devenir un vecteur naturel de pertinence personnelle. Le dialogue direct — où l'élève pose ses propres questions et reçoit des réponses adaptées

à son niveau de compréhension et à ses préoccupations — crée une connexion entre le contenu historique et la curiosité individuelle. Chaque élève explore les aspects qui l'intriguent, génère ses propres connexions à travers ses questions, et découvre ainsi la pertinence personnelle du savoir historique.

Cette *personnalisation émergente* — qui naît de l'interaction plutôt que d'une programmation préalable — combine plusieurs mécanismes identifiés : elle permet l'auto-génération de connexions personnelles (effort élevé), offre un sentiment de contrôle sur l'exploration (autonomie), et peut activer différentes dimensions de pertinence selon les questions posées (personnelle/impersonnelle, appliquée/conceptuelle). L'alignement thématique entre le personnage et le contenu de la leçon peut amplifier ces effets (?). Cette hypothèse constitue l'un des axes centraux de notre programme expérimental.

Reste à comprendre comment la forme de l'interaction — et non seulement son contenu — affecte l'engagement cognitif. Le cadre ICAP, que nous examinons maintenant, offre une grille d'analyse pour distinguer différents niveaux d'engagement et prédire leurs effets sur l'apprentissage.

2.1.4 L'Apprentissage Actif et le Cadre ICAP

Les sections précédentes ont montré que l'effort cognitif investi dans la génération de connexions personnelles renforce leur impact motivationnel. Ce constat s'inscrit dans un cadre plus général : l'apprentissage actif, qui dépasse la simple activité physique pour désigner le fait de « penser activement à ce que l'on fait » (Mayer, 2014a; Yannier et al., 2020). La question devient alors : comment caractériser les différents niveaux d'engagement cognitif et prédire leurs effets sur l'apprentissage ?

Le cadre ICAP (*Interactive, Constructive, Active, Passive*) propose une taxonomie hiérarchisée des activités d'apprentissage selon leur niveau d'engagement cognitif (Chi, 2009). Le mode *Passif* correspond à la réception d'information sans comportement observable au-delà de l'attention — écouter un cours, regarder une vidéo. Le mode *Actif* implique une manipulation ou une attention focalisée sans production de nouvelles idées — prendre des notes verbatim, surligner. Le mode *Constructif* requiert la génération d'idées qui dépassent l'information présentée — formuler des hypothèses, élaborer des explications, connecter le contenu à son expérience personnelle. Le mode *Interactif* ajoute une dimension dialogique : les partenaires co-construisent des connaissances à

travers un échange où chacun contribue substantiellement.

La prédiction centrale du modèle — Interactif > Constructif > Actif > Passif en termes de gains d'apprentissage — a reçu un soutien empirique substantiel. Une méta-analyse portant sur 225 études montre que l'apprentissage actif augmente significativement la performance des étudiants en STIM comparé aux cours magistraux (Freeman et al., 2014). L'interactivité permet de réguler le rythme d'apprentissage, d'explorer les concepts selon ses intérêts, de formuler des questions et de recevoir un feedback immédiat (Domagk, 2010; Evans and Gibbons, 2007).

Cette hiérarchie éclaire les résultats sur les interventions de pertinence : les approches à effort élevé (auto-génération, réévaluation dirigée) relèvent du mode *Constructif*, tandis que les approches à faible effort (communication directe) maintiennent l'apprenant en mode *Passif*. L'efficacité supérieure des premières s'explique ainsi par leur niveau d'engagement cognitif plus profond. Un paradoxe mérite cependant attention : l'apprentissage actif, bien que conduisant à de meilleurs résultats, est souvent perçu comme moins efficace par les apprenants eux-mêmes (Deslauriers et al., 2019). L'effort cognitif accru est interprété comme un signe de difficulté alors qu'il indique un traitement plus profond.

Dans l'enseignement de l'Histoire, les approches interactives produisent des effets positifs sur la compréhension et l'intérêt. Les discussions de groupe favorisent une meilleure compréhension des concepts historiques (Favero et al., 2007). La narration numérique interactive, avec des points de décision stratégiques, stimule des discussions significatives et une compréhension plus profonde (Petousi et al., 2021). La combinaison d'interactions tangibles avec des récits émotionnels encourage les adolescents à s'engager avec les figures historiques au-delà de la connaissance factuelle (Roussou et al., 2024a).

Un agent conversationnel incarnant un personnage historique se situe naturellement au niveau *Interactif* : l'élève formule des questions (activité constructive), l'agent répond, et l'échange peut conduire à une co-construction de sens. Cette position contraste avec la vidéo (mode *Passif*) et le texte (mode *Actif* si l'élève surligne ou prend des notes). Le cadre ICAP prédit ainsi que l'agent dialogique devrait produire des gains d'engagement et d'apprentissage supérieurs aux formats traditionnels — une prédiction que notre programme expérimental vise à tester.

Cette convergence des cadres théoriques — SDT, développement de l'intérêt, théorie

de l'attente-valeur, ICAP — dessine les contours d'une intervention prometteuse : un agent conversationnel qui satisfait les besoins d'autonomie, de compétence et d'affiliation, déclenche et maintient l'intérêt par la nouveauté et l'interaction sociale, permet la génération de connexions personnelles, et engage l'apprenant au niveau interactif du cadre ICAP.

2.2 Le Contexte Spécifique de l'Enseignement de l'Histoire

Les cadres théoriques présentés dans la section précédente éclairent les mécanismes généraux de l'engagement dans l'apprentissage. Leur application à l'enseignement de l'Histoire révèle des défis spécifiques liés à la nature épistémologique de la discipline, aux perceptions qu'en ont les élèves, et aux pratiques pédagogiques qui y prédominent. Cette section analyse successivement ces trois dimensions : la comparaison épistémologique avec les STIM (§2.2.1), le paradoxe de la perception des élèves (§2.2.2), les pratiques pédagogiques efficaces (§2.2.3), et le bilan empirique des technologies éducatives dans ce domaine (§2.2.4).

2.2.1 Épistémologie et Comparaison Interdisciplinaire

L'enseignement de l'Histoire se distingue des disciplines STIM (Sciences, Technologie, Ingénierie, Mathématiques) par la nature même des savoirs qu'il transmet et des compétences qu'il développe. Cette différence épistémologique conditionne la réception de la discipline par les élèves et les défis pédagogiques qui lui sont propres.

Nature des savoirs et rapport à la vérité. Les disciplines STIM reposent sur des savoirs cumulatifs et universels, vérifiables par l'expérimentation ou la démonstration (Straaten et al., 2016). La progression y suit une logique d'accumulation où chaque concept s'appuie sur les précédents, et le feedback est immédiat : une équation est résolue correctement ou non. L'Histoire produit, à l'inverse, des savoirs interprétatifs et contextuels (Haydn and Harris, 2010). La vérité historique émerge de l'argumentation fondée sur l'analyse critique de sources, processus où plusieurs interprétations peuvent légitimement coexister pour un même événement. Cette nature interprétative déstabilise

les élèves habitués aux certitudes des sciences exactes : ils peuvent produire une réponse cohérente et argumentée tout en recevant une évaluation négative, sans comprendre les critères qui ont guidé le jugement (Harris and Haydn, 2006).

Perception de l'utilité et orientation scolaire. L'analyse par le prisme de la *Utility Value* (cf. §2.1.3) révèle une asymétrie défavorable à l'Histoire. Les STIM bénéficient d'une perception d'utilité professionnelle immédiate : débouchés identifiés, applications concrètes, discours social valorisant (Potvin and Hasni, 2014). Environ 14% des élèves européens estiment que « l'histoire est morte et n'a rien à voir avec ma vie présente » (Straaten et al., 2016). Cette perception persiste alors même que ces élèves reconnaissent l'utilité cognitive de l'Histoire pour comprendre le monde actuel (Haydn and Harris, 2010). La distinction entre utilité professionnelle perçue et utilité formative réelle explique ce paradoxe : la seconde pèse moins dans les décisions d'orientation que la première (Harackiewicz et al., 2016).

Atouts distinctifs et enjeux pédagogiques. Ces différences épistémologiques ne constituent pas uniquement des obstacles. Le potentiel narratif et dramatique de l'Histoire permet un engagement émotionnel moins courant dans les disciplines formelles (Harris and Haydn, 2006). La discipline développe des compétences transférables — analyse critique, argumentation, capacité de synthèse — recherchées dans de nombreux contextes professionnels (Haydn and Harris, 2010). Sa contribution à la formation citoyenne et à la compréhension interculturelle répond à des besoins sociétaux croissants. L'enjeu pédagogique n'est donc pas d'imiter les STIM en standardisant l'enseignement historique, mais de valoriser ses spécificités épistémologiques tout en explicitant les compétences qu'il développe (Straaten et al., 2016).

2.2.2 La Perception des Élèves : le Paradoxe « Ennuyeux mais Important »

La recherche sur les perceptions des élèves révèle un paradoxe structurant : l'Histoire est simultanément reconnue comme importante et critiquée pour ses modalités d'enseignement. Ce décalage entre valeur attribuée et intérêt suscité constitue un défi pédagogique.

Position relative parmi les disciplines scolaires. Contrairement aux idées reçues, l'Histoire ne se classe pas systématiquement dernière parmi les disciplines scolaires. Une enquête auprès de 1740 élèves britanniques indique que 69,8% trouvent la matière « assez agréable », la plaçant en cinquième position ([Harris and Haydn, 2006](#)). Les données américaines révèlent que 55,3% des élèves de 4^e se déclarent intéressés par la discipline ([Henkaline, 2023](#)). Ces résultats contrastent avec des études antérieures plus pessimistes : l'intérêt dépend des contextes nationaux et des pratiques pédagogiques locales. L'Histoire reste cependant surpassée par les matières impliquant une activité physique ou une manipulation concrète : éducation physique, technologie, arts plastiques ([Harris and Haydn, 2006](#)).

Le paradoxe forme/fond. Un paradoxe traverse les contextes nationaux : les élèves jugent l'Histoire ennuyeuse dans sa forme scolaire tout en la considérant importante pour comprendre le monde actuel ([Henkaline, 2023](#); [Harris and Haydn, 2006](#)). En France comme en Suisse, ils la perçoivent comme distante de leurs préoccupations quotidiennes ([Audigier and Fink, 2010](#)). Le cadre théorique de la pertinence personnelle éclaire ce constat : lorsque les contenus sont perçus comme une collection de faits à mémoriser plutôt que comme un processus d'enquête connecté à l'expérience vécue, l'intérêt situationnel peine à se transformer en intérêt individuel durable ([Priniski et al., 2018](#)). Cette déconnexion n'est pas inhérente à la discipline mais résulte largement des modalités pédagogiques qui prédominent dans son enseignement.

Impact des pratiques sur le sentiment de compétence. L'enseignement traditionnel de l'Histoire présente des caractéristiques qui fragilisent le sentiment de compétence des élèves. La faible autonomie accordée, le feedback différé et peu explicite, ainsi que les aspects contrôlants prédominants sapent les conditions de l'intérêt intrinsèque identifiées par la théorie de l'autodétermination (cf. §2.1.1). La nature interprétative de la discipline amplifie cette difficulté : l'élève peut produire une réponse cohérente et argumentée tout en recevant une évaluation négative, sans comprendre les critères mobilisés ([Harris and Haydn, 2006](#)). Cette incertitude évaluative contraste avec la clarté des disciplines où les réponses sont univoquement « justes ou fausses ».

Variations selon les profils. Des différences significatives apparaissent selon le genre et l'origine. Les filles manifestent une appréciation affective plus élevée pour l'Histoire, tandis que les élèves issus de minorités rapportent une appréciation comportementale supérieure ([Henkaline, 2023](#)). Certains aspects de l'Histoire — son potentiel narratif, sa capacité à donner voix à des perspectives marginalisées — résonnent différemment selon les profils.

2.2.3 Pratiques Pédagogiques : du Transmissif à l'Actif

L'analyse des perceptions élèves converge vers un constat : le problème de l'enseignement de l'Histoire réside moins dans son contenu que dans ses modalités pédagogiques. La transition du mode transmissif vers des approches actives constitue un levier pour stimuler l'intérêt.

Limites du cours magistral. Le cours magistral traditionnel présente plusieurs limites structurelles. Il ne tire pas parti du potentiel narratif et dramatique de la discipline, transformant l'Histoire en une succession de faits à mémoriser et occultant sa dimension interprétative ([Audigier and Fink, 2010](#)). Il prive les élèves de l'expérience de l'enquête historique : la confrontation aux sources, la construction d'une argumentation, la mise en perspective. Les élèves identifient spontanément ces limites : ils critiquent la prise de notes passive et la redondance des contenus d'une année sur l'autre ([Harris and Haydn, 2006](#); [Henkaline, 2023](#)). Ce diagnostic converge avec les prédictions du cadre ICAP (cf. §2.1.4) : les activités passives génèrent un traitement cognitif superficiel.

Efficacité des approches interactives. À l'inverse, les approches interactives recueillent l'adhésion des élèves et produisent des gains d'apprentissage mesurables. Les jeux de rôle, débats et discussions sont particulièrement appréciés ([Harris and Haydn, 2006](#)). Les méthodes expérientielles — simulations de procès historiques, reconstitutions, vidéos immersives — s'avèrent efficaces pour susciter l'intérêt ([Henkaline, 2023](#)). Ces approches génèrent un engagement cognitif plus profond que la réception passive, conformément aux prédictions du cadre ICAP : elles activent les modes *Constructif* et *Interactif* où les élèves génèrent des inférences et confrontent leurs interprétations ([Chi and Wylie, 2014](#)).

La narrativité comme levier spécifique. La narrativité constitue un atout distinctif de l'Histoire que les approches actives peuvent exploiter. La narration numérique interactive, intégrant des points de décision stratégiques dans les récits historiques, stimule des discussions de groupe significatives et une compréhension approfondie (Petousi et al., 2022). Ces points de décision permettent aux élèves de s'engager activement dans la réflexion historique en générant des hypothèses sur les conséquences de leurs choix. La connexion émotionnelle avec le contenu historique renforce cet effet : des installations combinant interactions tangibles et récits émotionnels encouragent les adolescents à s'engager avec les figures historiques au-delà de la connaissance factuelle, développant une forme d'empathie historique (Roussou et al., 2024b).

2.2.4 Technologies Éducatives en Histoire : Bilan Empirique

L'intégration des technologies numériques dans l'enseignement de l'Histoire a fait l'objet d'expérimentations variées. L'analyse comparative de ces travaux permet d'identifier les approches prometteuses, celles qui ont déçu, et les conditions transversales qui déterminent leur efficacité.

Approches ayant produit des résultats probants

Plusieurs dispositifs technologiques ont démontré une efficacité mesurable sur l'intérêt ou l'apprentissage en contexte historique, avec des mécanismes d'action distincts.

La **narration numérique interactive** figure parmi les approches les mieux documentées. L'intégration de points de décision stratégiques dans des récits historiques stimule des discussions de groupe significatives et une compréhension approfondie (Petousi et al., 2022). Dans une étude portant sur l'Athènes antique, les élèves confrontés à des choix narratifs ont développé une réflexion historique plus élaborée que ceux exposés à un récit linéaire. L'efficacité repose sur l'activation du mode *Constructif* du cadre ICAP : les apprenants génèrent des hypothèses sur les conséquences de leurs choix plutôt que de recevoir passivement l'information.

Les **environnements virtuels immersifs** produisent des gains d'apprentissage lorsqu'ils permettent une exploration autonome. Une reconstruction 3D de l'ancienne Uruk a conduit à une amélioration des performances aux tests et à un temps d'exploration accru comparé aux méthodes traditionnelles (Ijaz et al., 2017). L'immersion

spatiale active des processus de mémorisation épisodique qui renforcent la rétention des informations contextuelles.

Les **installations muséales** combinant interactions tangibles et récits émotionnels favorisent le développement de l'empathie historique (Roussou et al., 2024b). L'exploration de perspectives émotionnelles multiples sur des personnages historiques permet aux adolescents de dépasser la connaissance factuelle pour développer une compréhension des motivations et des contextes. Cette approche répond au déficit de connexion personnelle identifié dans la perception de l'Histoire par les élèves (cf. §2.2.2).

Les **discussions de groupe structurées** autour de problèmes historiques améliorent simultanément la compréhension conceptuelle et l'intérêt pour la discipline (Favero et al., 2007). La confrontation des interprétations entre pairs active des processus de réélaboration cognitive absents de l'apprentissage individuel, conformément aux prédictions du mode *Interactif* du cadre ICAP.

Approches aux résultats mitigés ou décevants

Certaines technologies n'ont pas tenu leurs promesses en contexte historique, révélant des limites instructives.

Les **agents pédagogiques scriptés** présentent des résultats décevants en Histoire (Davis et al., 2021). Cette contre-performance s'explique par l'inadéquation entre la rigidité des réponses préprogrammées et la nature interprétative de la discipline : un agent incapable de nuancer son discours ou de reconnaître la pluralité des interprétations apparaît comme une source d'autorité inappropriée.

Les **interfaces conversationnelles textuelles** avec figures historiques produisent des résultats quantitatifs mixtes malgré des gains motivationnels (Pataranutaporn et al., 2023). La comparaison entre dialogues interactifs et lecture traditionnelle indique une amélioration de la motivation mais des effets incertains sur l'apprentissage objectif. L'absence de modalité orale pourrait limiter l'activation des mécanismes de présence sociale.

La **simple introduction d'outils technologiques** sans repensée pédagogique ne garantit pas l'amélioration de l'intérêt (Yarema, 2002). Les technologies qui reproduisent le format transmissif sous une forme numérique échouent à transformer la relation des élèves au contenu historique.

Conditions de succès transversales

L'analyse comparative de ces expérimentations fait émerger trois conditions favorables à l'efficacité des technologies en enseignement de l'Histoire.

Premièrement, l'**activation de l'agentivité** de l'apprenant : les dispositifs efficaces offrent des choix significatifs qui engagent la réflexion et permettent à l'élève de réguler son exploration (Petousi et al., 2022; Ijaz et al., 2017).

Deuxièmement, la **dimension sociale** de l'interaction : les approches les plus prometteuses impliquent une confrontation des perspectives, que ce soit entre pairs ou avec un agent capable de dialogue authentique (Favero et al., 2007; Roussou et al., 2024b).

Troisièmement, la **cohérence épistémologique** : les technologies qui respectent la nature interprétative de l'Histoire surpassent celles qui importent un modèle transmissif adapté aux disciplines formelles (Davis et al., 2021).

Ces constats informent la conception de notre dispositif expérimental : l'agent conversationnel incarnant un personnage historique vise à satisfaire ces trois conditions.

2.3 Les Agents Pédagogiques Virtuels : Typologie, Signaux Sociaux et Efficacité

Depuis le tuteur SCHOLAR de Carbonell en 1970, les agents pédagogiques virtuels ont connu trois générations de développement. La première (2000–2011) a exploré les possibilités des agents animés avec des systèmes comme Steve et Herman the Bug. La deuxième (2012–2019) a consolidé les bases empiriques à travers de nombreuses méta-analyses. La troisième (2020–présent) intègre les capacités des grands modèles de langage, transformant radicalement les possibilités d'interaction.

Cette section dresse un état de l'art systématique du domaine. Nous proposons d'abord une cartographie des agents pédagogiques à travers leur évolution historique et leurs caractéristiques de design (§2.3.1). Nous analysons ensuite les fondements cognitifs qui contraignent leur efficacité (§2.3.2), puis les mécanismes par lesquels ils génèrent une présence sociale (§2.3.3). Une taxonomie des signaux sociaux et de leur efficacité empirique est ensuite présentée (§2.3.4). Nous examinons enfin les limites du réalisme (§2.3.5).

2.3.1 Cartographie des Agents Pédagogiques : 25 Ans d'Évolution

L'histoire des agents pédagogiques virtuels peut être structurée en trois générations distinctes, chacune caractérisée par des paradigmes technologiques et des objectifs de recherche spécifiques (??).

La **première génération (1997–2011)** correspond à l'ère des pionniers. Les premiers agents pédagogiques animés émergent à la fin des années 1990 avec des systèmes comme Steve (*Soar Training Expert for Virtual Environments*) (?), agent 3D guidant les marins dans l'apprentissage de procédures techniques, ou Herman the Bug (?), agent 2D anthropomorphisé sous forme d'insecte qui a démontré l'*effet de persona* : sa seule présence améliorait la motivation des apprenants, indépendamment de la qualité du feedback (?). AutoTutor (?) a introduit le dialogue socratique automatisé, posant les bases des agents conversationnels éducatifs. Ces systèmes pionniers reposaient sur des scripts prédéfinis et des arbres de décision, ce qui limitait leur flexibilité conversationnelle.

La **deuxième génération (2012–2019)** se caractérise par la consolidation empirique. Les méta-analyses se multiplient pour évaluer l'efficacité réelle des agents selon différentes variables de design : effets de l'affect, impact des gestes, comparaisons entre apparences 2D et 3D, voix humaines et synthétiques, niveaux d'anthropomorphisme (??). Un constat émerge de ces synthèses : sur 15 études avec groupe contrôle, 9 ne montrent aucune différence d'apprentissage significative, appelant à une approche plus rigoureuse (?). Cette génération établit les bases empiriques du domaine, mais les agents restent contraints par leurs scripts préprogrammés.

La **troisième génération (2020–présent)** correspond à l'ère générative. L'émergence des grands modèles de langage transforme radicalement le paysage : les agents peuvent désormais générer des réponses contextuellement appropriées sans scripts prédéfinis. Les chatbots IA en éducation produisent des effets positifs sur les résultats d'apprentissage, mais cette flexibilité conversationnelle accrue s'accompagne de nouveaux défis liés à l'imprévisibilité des réponses (??). Cette génération, à laquelle appartient notre agent historique, hérite des acquis des générations précédentes tout en introduisant des risques inédits — notamment celui des hallucinations analysé en section 2.4.

Au-delà de la chronologie, les agents se différencient selon plusieurs dimensions de design. Le tableau 2.1 propose une typologie fondée sur les caractéristiques empiriquement étudiées dans la littérature, synthétisée à partir de 17 méta-analyses et revues systématiques.

TABLE 2.1 – Typologie des agents pédagogiques selon leurs caractéristiques de design. Synthèse établie à partir des méta-analyses de [1], [2] et [3].

Dimension	Modalités	Exemples	Effet sur apprentissage
<i>Représentation visuelle</i>	Absent / Statique / Animé 2D / Animé 3D	Texte seul / Image fixe / Cartoon / Avatar 3D	Animé > Statique ; 2D \geq 3D
<i>Niveau d'anthropomorphisme</i>	Non-humain / Stylisé / Réaliste / Hyperréaliste	Robot / Cartoon / Avatar / Deepfake	Stylisé souvent optimal
<i>Modalité vocale</i>	Texte / Voix synthétique / Voix humaine	Chat / TTS / Enregistrement	Humaine > Synthétique
<i>Rôle pédagogique</i>	Tuteur / Compagnon / Teachable agent	Expert / Assistant / Élève virtuel	Dépend du contexte
<i>Moteur conversationnel</i>	Script / Règles / NLP-ML / LLM	Arbre décision / AIML / Seq2Seq / GPT	LLM : flexibilité \uparrow , contrôle \downarrow

L'application des agents pédagogiques à l'enseignement de l'Histoire reste peu explorée et les résultats disponibles sont préoccupants : l'effet mesuré pour l'histoire est négatif, contre des effets positifs en biologie et en informatique ([4]). Cette disparité disciplinaire s'explique par un défi épistémologique spécifique : contrairement aux domaines STIM où l'agent peut s'appuyer sur des connaissances vérifiables, l'agent historique doit naviguer entre faits documentés, interprétations historiographiques et zones d'incertitude. Cette spécificité justifie une attention particulière aux risques métacognitifs que nous analysons dans cette thèse.

La figure 2.1 illustre cette évolution à travers deux exemples représentatifs des deux premières générations.

2.3.2 Fondements Cognitifs : Contraintes et Principes de Design

L'efficacité des agents pédagogiques est contrainte par l'architecture cognitive humaine. Deux cadres théoriques informent directement leur conception : la Théorie de la Charge Cognitive et la Théorie Cognitive de l'Apprentissage Multimédia.

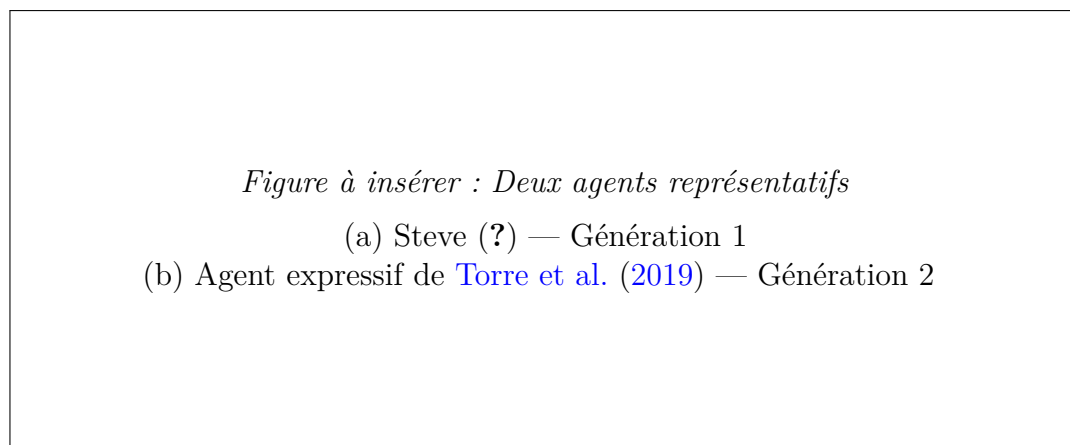


FIGURE 2.1 – Évolution des agents pédagogiques à travers les deux premières générations. (a) Steve, agent 3D pionnier pour l’entraînement naval (?). (b) Agent expressif utilisé dans les études sur l’alignement comportemental et contextuel (Torre et al., 2019).

La mémoire de travail constitue le goulot d’étranglement de l’apprentissage (?). Sa capacité limitée impose de distinguer trois types de charge : la charge *intrinsèque* (complexité du contenu), la charge *extrinsèque* (inefficacités de présentation), et la charge *pertinente* (effort d’apprentissage productif). Pour les agents pédagogiques, cette distinction a une implication directe : les éléments de design non fonctionnels peuvent constituer une charge extrinsèque susceptible de compromettre l’apprentissage.

Un agent visuellement complexe — animations élaborées, environnement 3D détaillé, expressions faciales sophistiquées — peut paradoxalement nuire à l’apprentissage s’il détourne des ressources cognitives du contenu éducatif. Les méta-analyses confirment ce risque : les agents 2D surpassent les agents 3D sur les mesures d’apprentissage (?). Ce résultat contre-intuitif s’explique par le coût cognitif du réalisme : le traitement d’un environnement 3D immersif consomme des ressources au détriment du contenu.

La Théorie Cognitive de l’Apprentissage Multimédia (?) postule l’existence de deux canaux de traitement distincts : visuel-pictural et auditif-verbal. Le *principe de modalité* qui en découle stipule que l’apprentissage est favorisé lorsque les explications verbales sont présentées sous forme audio plutôt que textuelle. En déléguant l’information verbale au canal auditif, le concepteur libère le canal visuel pour les éléments graphiques pertinents.

Ce principe fournit la justification théorique des agents vocaux : un agent qui *parle* plutôt qu’il n’affiche du texte permet une répartition optimale de la charge entre les canaux. Les voix humaines s’avèrent plus efficaces que les voix synthétiques, bien que

cet écart se réduise avec les progrès technologiques. Le *principe de personnalisation* complète cette recommandation : un style conversationnel surpasse un style formel.

Les gestes de l'agent constituent un cas particulier : ils peuvent soit faciliter l'apprentissage en guidant l'attention, soit le compromettre en ajoutant une charge extrinsèque. Les données empiriques révèlent des effets modérés des gestes sur le transfert proche et la rétention (?). Ces effets sont conditionnés par la *congruence sémantique* : les gestes doivent être alignés avec le contenu verbal. Un geste déictique pointant vers un élément pertinent du graphique facilite l'intégration ; un geste générique non relié au contenu constitue une distraction.

Ces résultats convergent vers un paradoxe apparent : les agents les plus efficaces ne sont pas les plus réalistes, mais les plus parcimonieux. Un agent doit fournir suffisamment d'indices pour activer l'engagement social (voir section suivante), sans surcharger le système cognitif par des éléments non fonctionnels. Cette recommandation de parcimonie entre en tension avec l'évolution technologique : les capacités croissantes de rendu réaliste ne garantissent pas une efficacité pédagogique accrue. La question n'est pas « que peut-on techniquement réaliser ? » mais « qu'est-ce qui sert effectivement l'apprentissage ? ».

2.3.3 De l'Artefact au Partenaire : Mécanismes de la Présence Sociale

La section précédente a établi les contraintes cognitives qui encadrent l'efficacité des agents. Mais l'impact d'un agent ne se réduit pas à sa fonction de canal d'information : il dépend aussi de sa capacité à être perçu comme un partenaire social. Cette perception transforme l'interaction technique en expérience relationnelle.

Le paradigme CASA (*Computers Are Social Actors*) constitue le cadre explicatif central de ce phénomène (?). Les travaux fondateurs de Nass et ses collaborateurs ont démontré que les utilisateurs appliquent inconsciemment aux ordinateurs les règles sociales qu'ils utiliseraient avec des humains : politesse, réciprocité, attribution de personnalité. Ce traitement social tend à s'opérer même lorsque l'utilisateur *sait* qu'il interagit avec une machine — un processus qualifié d'*ethopoeia*.

Des indices sociaux minimaux suffisent à déclencher ces scripts relationnels. Une voix, qu'elle soit synthétique ou humaine, active des réponses sociales. Les utilisateurs évaluent différemment un ordinateur selon le genre de sa voix, reproduisant les stéréotypes

sociaux : une voix masculine est perçue comme plus compétente sur les sujets techniques. Ces attributions automatiques expliquent pourquoi même des agents rudimentaires peuvent générer un engagement significatif.

La présence sociale désigne le sentiment d’être « avec » une autre intelligence dans un environnement médiatisé (?). Ce sentiment ne requiert pas un interlocuteur humain ; il peut émerger de l’interaction avec un agent virtuel pourvu que certaines conditions soient réunies. L’utilisateur doit percevoir l’agent non comme un outil passif, mais comme une entité dotée d’une forme d’intentionnalité et de réactivité.

Cette présence sociale constitue la condition préalable à l’activation des mécanismes d’apprentissage social. Sans elle, l’interaction tend à rester instrumentale : l’utilisateur traite l’information sans s’engager pleinement dans la relation. Avec elle, l’agent devient un partenaire dont on respecte implicitement le « contrat de communication ».

La théorie de l’agence sociale (*Social Agency Theory*) explicite le lien entre présence sociale et apprentissage (??). Lorsque l’apprenant perçoit l’agent comme un partenaire social, un contrat implicite s’établit : l’apprenant s’engage à « honorer » cette relation par un effort cognitif accru. Cet effort supplémentaire se traduit par un traitement plus profond du contenu.

Les méta-analyses confirment ce mécanisme : les agents anthropomorphisés produisent des effets supérieurs aux agents non anthropomorphisés, mais ces effets concernent davantage les mesures affectives et motivationnelles que les mesures cognitives pures (??).

2.3.4 Taxonomie et Efficacité des Signaux Sociaux

Les agents pédagogiques mobilisent une variété de signaux sociaux pour établir la présence et faciliter l’apprentissage. Cette section propose une taxonomie systématique de ces signaux, fondée sur les résultats empiriques des méta-analyses du domaine. Le tableau 2.2 synthétise les principaux effets documentés.

La **représentation visuelle** constitue le premier niveau de signaux. La présence d’un agent produit un effet global positif, mais ce résultat masque des variations importantes selon le type de représentation (?). Les agents 2D surpassent les agents 3D, un écart qui s’explique par le coût cognitif du réalisme : le traitement d’environnements 3D immersifs consomme des ressources au détriment du contenu pédagogique. L’animation constitue

TABLE 2.2 – Efficacité des signaux sociaux dans les agents pédagogiques. Synthèse des résultats issus des méta-analyses.

Catégorie	Signal	Effet	Source
<i>Représentation</i>	Présence vs absence	Effet positif	?
	2D vs 3D	2D > 3D	?
	Animé vs statique	Animé > Statique	?
<i>Voix</i>	Voix humaine vs syn- thétique	Humaine > Syn- thétique	?
	Style conversationnel	Effet positif mo- déré	?
<i>Gestes</i>	Gestes déictiques	Effet positif (transfert)	?
	Gestes génériques	Effet modéré (ré- tention)	?
<i>Affect</i>	Expression émotion- nelle	Effet positif mo- déré	?
	Affect positif	Corrélation posi- tive	?
<i>Interaction</i>	Chatbots IA	Effet positif im- portant	?
	Feedback adaptatif	Effet positif	?

néanmoins un atout : les agents animés surpassent les agents statiques, l’animation permettant de guider l’attention et de maintenir l’engagement (?).

La **voix** représente un signal social particulièrement puissant. Les voix humaines conservent un avantage sur les voix synthétiques, bien que cet écart se réduise avec les progrès technologiques (?) : les voix synthétiques de haute qualité produisent désormais des résultats d’apprentissage équivalents, voire supérieurs, pour le transfert de connaissances (?). La qualité prosodique s’avère plus déterminante que l’incarnation visuelle pour la perception de naturalité : un agent réaliste avec une prosodie inadéquate sera jugé moins naturel qu’un agent désincarné avec une prosodie correcte (Ehret et al., 2021). Au-delà de la qualité vocale, le style discursif influence l’apprentissage : le style conversationnel produit un effet positif par rapport au style formel (?). Ce résultat s’aligne avec le principe de personnalisation de la théorie cognitive de l’apprentissage multimédia : un discours adressé directement à l’apprenant (“vous”) favorise l’engagement par rapport à un discours impersonnel.

Les **gestes** de l’agent constituent une catégorie de signaux aux effets différenciés. Deux types de gestes produisent des effets distincts (?). Les gestes déictiques — pointage

vers des éléments pertinents de l'interface — produisent un effet important sur le transfert en guidant explicitement l'attention visuelle et en facilitant l'intégration des informations verbales et graphiques. Les gestes génériques — mouvements non liés au contenu — produisent un effet plus modeste sur la rétention. Cette différence confirme l'importance de la congruence sémantique : les gestes doivent être alignés avec le contenu pour maximiser leur efficacité.

La combinaison de plusieurs types de communication non verbale n'amplifie pas nécessairement les effets : un seul type de signal approprié au résultat d'apprentissage visé (gestes *ou* expressions faciales) surpasse la combinaison des deux (?). Les expressions faciales favorisent l'apprentissage attitudinal, tandis que les gestes facilitent l'apprentissage procédural. Ce résultat contre-intuitif s'explique par la théorie de la charge cognitive : la multiplication des signaux peut surcharger l'apprenant plutôt que l'assister. L'inconsistance des résultats dans la littérature confirme que l'efficacité des signaux non verbaux dépend davantage de leur appropriation au contexte que de leur quantité (?).

L'**affect** de l'agent influence l'apprentissage à travers plusieurs mécanismes. Une corrélation positive existe entre affect de l'agent et apprentissage (?). L'expression émotionnelle se distingue de la simple présence affective (?). Les expressions faciales doivent être alignées avec le contexte pour produire leurs effets (Liew et al., 2016a) ; lorsqu'elles sont coordonnées avec le discours, elles influencent positivement la motivation (Liew et al., 2017). Toutefois, l'expressivité émotionnelle positive peut s'avérer contre-productive : elle peut sembler inappropriée et nuire à la confiance lors de tâches critiques où la neutralité est attendue (Torre et al., 2019). Les effets affectifs sont également plus marqués sur les mesures motivationnelles que sur les mesures cognitives pures (?).

La dimension **interactive** distingue les agents conversationnels des agents unidirectionnels. Les chatbots IA en éducation produisent des effets positifs importants (?), suggérant que la capacité de dialogue — recevoir les questions de l'apprenant et y répondre de manière contextuelle — constitue un amplificateur majeur de l'efficacité pédagogique.

La **motivation** constitue une variable médiatrice importante. Les effets des chatbots éducatifs varient selon le type de motivation mesurée : l'auto-efficacité montre des effets

importants, tandis que la motivation intrinsèque présente des effets plus modérés (?). Ces résultats suggèrent que les agents agissent principalement en renforçant la confiance de l'apprenant dans ses capacités plutôt qu'en modifiant son intérêt intrinsèque pour le contenu.

L'**alignement thématique** entre l'agent et le contenu constitue un facteur modulateur. Les apprenants interagissant avec un agent thématiquement aligné (un astronaute enseignant des concepts spatiaux) rapportent des niveaux supérieurs de présence sociale et d'engagement comparés à ceux interagissant avec un agent neutre ([Schmidt et al., 2019](#)). Ce résultat suggère que la cohérence entre l'identité de l'agent et le domaine enseigné peut amplifier les effets des signaux sociaux.

Une analyse transversale révèle un patron cohérent : les effets des signaux sociaux sont systématiquement plus importants sur les mesures affectives et motivationnelles que sur les mesures cognitives. Sur 15 études avec groupe contrôle, 9 ne montrent aucune différence d'apprentissage significative, alors que les effets sur l'engagement sont généralement positifs (?). Une dissociation frappante émerge entre perception et performance : les apprenants apprécient davantage la leçon, la jugent plus intéressante et ont l'impression de mieux apprendre, mais leurs résultats de compréhension objective diminuent — un phénomène résumé par l'expression “*more gaze, but less learning*” ([Wilson et al., 2018](#)).

La figure 2.2 visualise ces relations entre signaux sociaux et types d'effets.

Ces résultats convergent vers une recommandation de parcimonie stratégique : multiplier les signaux sociaux n'amplifie pas nécessairement l'efficacité pédagogique. Chaque signal doit être évalué selon son rapport bénéfice (engagement, guidage attentionnel) / coût (charge cognitive, distraction). Cette tension entre richesse des signaux et surcharge potentielle conduit naturellement à examiner les limites du réalisme dans la conception des agents.

2.3.5 Limites du Réalisme : Le Paradoxe de l'Hyperréalisme

Les résultats empiriques présentés dans les sections précédentes convergent vers un constat contre-intuitif : le réalisme accru des agents ne garantit pas une efficacité pédagogique supérieure. Cette section examine les mécanismes qui expliquent ce paradoxe et ses implications pour le design.

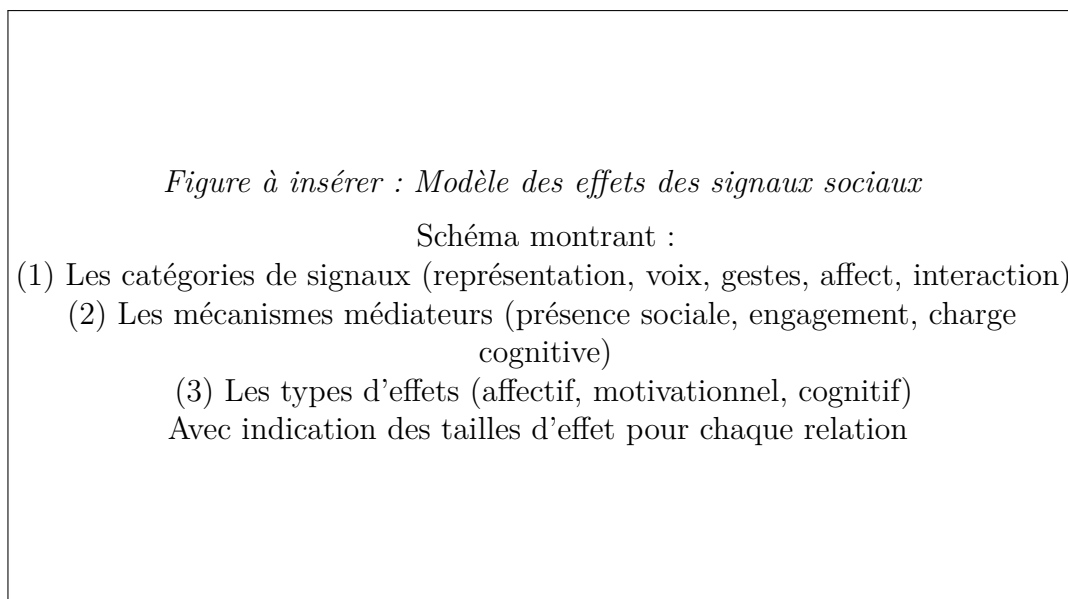


FIGURE 2.2 – Modèle intégratif des effets des signaux sociaux des agents pédagogiques. Les signaux influencent l'apprentissage à travers des mécanismes médiateurs, avec des effets différenciés selon le type de mesure.

Les comparaisons directes entre agents réalistes et agents stylisés révèlent une absence de différence significative sur les mesures d'apprentissage (?). Les agents stylisés génèrent parfois un engagement supérieur, possiblement parce qu'ils évitent les attentes implicites associées à l'apparence humaine. Lorsqu'un agent ressemble étroitement à un humain, l'apprenant s'attend à des comportements pleinement humains ; les déviations — latence de réponse, expressions faciales limitées, erreurs de compréhension — peuvent générer une dissonance qui compromet l'interaction.

L'écart entre agents 2D et agents 3D illustre ce phénomène (?). Cette différence ne s'explique pas uniquement par la charge cognitive du réalisme visuel : elle reflète également une inadéquation entre les capacités comportementales des agents et les attentes générées par leur apparence. Un agent 3D photoréaliste qui ne maintient pas un contact visuel approprié ou dont les expressions faciales restent rigides crée une impression d'étrangeté absente chez un agent 2D dont les limitations sont explicitement acceptées.

Ce décalage constitue un défi persistant du domaine (?). Les progrès technologiques permettent désormais des rendus visuels sophistiqués, mais les comportements interactifs — synchronisation multimodale, gestion des tours de parole, réactivité émotionnelle — n'ont pas progressé au même rythme. Cette asymétrie crée une vallée de l'étrange (*uncanny valley*) où l'agent est suffisamment réaliste pour activer les attentes sociales,

mais insuffisamment capable pour les satisfaire.

Les agents hyperréalistes peuvent également réduire l'engagement par un mécanisme psychologique distinct (?) : face à un agent perçu comme “intelligent” et “compétent”, certains apprenants développent une anxiété de performance qui inhibe leur participation. Les agents stylisés, perçus comme moins menaçants, favoriseraient une interaction plus détendue et un engagement plus authentique.

Ces résultats ont des implications directes pour le design des agents historiques. Un personnage historique rendu de manière hyperréaliste risque de générer des attentes impossibles à satisfaire : l'apprenant s'attend à une interaction “comme avec un vrai humain”, alors que l'agent reste limité par ses capacités techniques et la fiabilité de ses connaissances. Une représentation stylisée, explicitement non réaliste, pourrait paradoxalement favoriser une interaction plus productive en établissant d'emblée les limites de l'échange.

Cette analyse conduit à reformuler la question du design : plutôt que de maximiser le réalisme, il s'agit d'optimiser la congruence entre l'apparence de l'agent, ses capacités comportementales et les attentes de l'apprenant. Ces principes, établis pour les agents traditionnels à scripts prédéfinis, doivent être réexaminés à la lumière de la rupture technologique introduite par les IA génératives, dont les capacités conversationnelles modifient substantiellement les modalités d'interaction.

2.4 La Rupture Technologique des IA Génératives en Éducation

Les cadres théoriques présentés jusqu'ici ont été élaborés dans un contexte technologique où les agents pédagogiques fonctionnaient sur la base de scripts prédéfinis. L'émergence des Grands Modèles de Langage (LLM) et des technologies de synthèse multimédia a transformé ce paysage, créant de nouvelles possibilités mais aussi de nouveaux risques.

Cette section examine le saut qualitatif représenté par les agents génératifs (§2.4.1), leurs capacités de personnalisation en temps réel (§2.4.2), et le défi épistémique posé par leur propension aux hallucinations (§2.4.3).

2.4.1 Des Agents Scriptés aux Agents Génératifs : le Saut Qualitatif

Les agents pédagogiques ont évolué au cours des dernières décennies, passant de simples outils de diffusion d'information à des partenaires d'apprentissage plus sophistiqués (?). L'émergence des Grands Modèles de Langage (*Large Language Models*, LLM) a particulièrement accéléré cette évolution.

Les agents pédagogiques traditionnels fonctionnaient sur la base de scripts prédéfinis et d'arbres de décision. Leur comportement était entièrement déterminé par les anticipations de leurs concepteurs : chaque question possible devait être prévue, chaque réponse pré-rédigée. Cette architecture présentait des avantages — prévisibilité, contrôle du contenu, absence d'erreurs factuelles — mais aussi des limites fondamentales. La rigidité constituait le principal écueil : l'agent ne pouvait répondre qu'aux questions anticipées, dans les formulations anticipées. Toute déviation se heurtait à des réponses génériques. Cette rigidité contrastait avec la fluidité du dialogue humain et limitait le sentiment de présence sociale.

Les LLM ont transformé cette équation. Ces modèles peuvent générer des réponses personnalisées, s'adapter aux besoins des élèves en temps réel, et produire du contenu éducatif contextuellement pertinent (??). Ces avancées s'inscrivent dans la continuité des travaux sur les environnements d'apprentissage multimodaux interactifs (?). La fluidité conversationnelle des LLM constitue leur caractéristique la plus distinctive : ils génèrent un discours structuré, linguistiquement cohérent, et adapté au contexte de l'échange. Cette fluidité peut activer les mécanismes d'agence sociale décrits en §?? : l'apprenant perçoit l'agent comme un interlocuteur plutôt qu'un système automatique.

Cette évolution s'accompagne d'avancées parallèles en synthèse multimédia. Les réseaux antagonistes génératifs (GAN) permettent de créer des représentations visuelles hyperréalistes, brouillant la frontière entre réel et artificiel (?). L'animation faciale par apprentissage profond, le clonage vocal, et la génération vidéo atteignent des niveaux de réalisme inédits. Ces technologies peuvent atténuer l'effet de vallée de l'étrange (cf. §??), produisant des agents synthétiques perçus comme attractifs (?). Certains travaux indiquent que ces agents peuvent atteindre des niveaux de performance et de perception comparables à ceux d'instructeurs humains (??).

C'est précisément cette convergence — moteurs conversationnels fluides et interfaces

visuelles réalistes — qui multiplie les enjeux. D'un côté, des « moteurs » capables de produire un discours éloquent ; de l'autre, des « interfaces » visuelles pour les incarner. Cette combinaison ouvre des possibilités pédagogiques inédites, mais comporte également des risques que les sections suivantes examineront.

2.4.2 La Personnalisation en Temps Réel

L'une des capacités distinctives des agents alimentés par LLM réside dans leur aptitude à personnaliser le contenu d'apprentissage en temps réel, sans nécessiter de programmation préalable pour chaque profil d'apprenant. Cette capacité répond aux défis de mise en œuvre identifiés en §??.

L'apprentissage adaptatif traditionnel reposait sur des algorithmes prédéfinis : en fonction des réponses de l'élève à des questions diagnostiques, le système orientait vers des parcours prédéterminés (?). Les LLM permettent une forme d'adaptation plus fine : l'agent peut moduler son vocabulaire, ses exemples, son niveau de détail en fonction du flux de la conversation elle-même.

L'étude de la personnalisation dans l'apprentissage du vocabulaire illustre ce potentiel : un système développant des exemples et des récits adaptés aux intérêts individuels conduit à une augmentation de la motivation intrinsèque et à un sentiment renforcé de compétence et d'autonomie (?). Ces résultats démontrent la faisabilité d'une personnalisation automatique à grande échelle. L'exploration des interactions conversationnelles avec des figures historiques à travers des interfaces textuelles indique une amélioration de la motivation et des résultats d'apprentissage comparés à la lecture traditionnelle (?). L'évaluation de l'interactivité textuelle en éducation financière confirme que permettre aux étudiants de dialoguer avec l'instructeur virtuel conduit à une motivation et un engagement accrus comparés à l'instruction vidéo passive (?).

Ce qui distingue la personnalisation par LLM est son caractère émergent. L'adaptation n'est pas programmée explicitement : elle émerge de la capacité du modèle à générer des réponses contextuellement appropriées. L'élève pose une question selon ses propres termes, l'agent répond en s'adaptant. Si l'élève manifeste une incompréhension, l'agent peut reformuler spontanément. Cette fluidité adaptative présente un avantage pédagogique : chaque interaction devient unique, calibrée sur les besoins du moment.

Elle présente également un risque : l'adaptation peut masquer l'absence de compré-

hension réelle. L'agent qui reformule efficacement peut donner l'impression à l'élève qu'il a compris, alors que c'est l'agent qui a simplifié son discours au point de ne plus transmettre le concept dans sa complexité. Ce mécanisme constitue l'une des sources potentielles de l'illusion de compréhension.

L'examen des agents conversationnels conçus pour favoriser la curiosité chez les enfants du primaire révèle des résultats prometteurs : un agent encourageant le questionnement divergent conduit à une amélioration de la qualité des questions et à des activités exploratoires soutenues (?). Ces résultats suggèrent que la personnalisation peut être mise au service de l'engagement cognitif authentique plutôt que de la simple facilitation.

2.4.3 Fiabilité et Hallucinations : le Défi Épistémique

La fluidité conversationnelle des LLM masque une faille intrinsèque : leur propension à générer des informations factuellement incorrectes présentées avec une confiance apparente. Ce phénomène, qualifié d'« hallucination », représente un défi épistémique majeur pour les applications éducatives.

Les LLM sont des modèles probabilistes qui prédisent le mot suivant le plus probable étant donné le contexte. Cette nature stochastique implique qu'ils ne « connaissent » pas les faits au sens humain : ils génèrent des séquences statistiquement plausibles (?). Une taxonomie des hallucinations distingue les *hallucinations factuelles* (assertions fausses sur le monde), les *hallucinations de fidélité* (déformations de l'information source), et les *hallucinations d'entrée* (fabrication d'éléments non présents dans la requête). Dans un contexte éducatif, chaque type présente des risques spécifiques : date erronée, citation inventée, personnage historique attribué à la mauvaise période.

L'enseignement de l'Histoire présente une vulnérabilité particulière à ces hallucinations. La discipline repose sur des faits précis — dates, lieux, protagonistes — dont l'exactitude est vérifiable. Or, les LLM excellent dans la production de récits plausibles et cohérents ; ils peuvent générer une narration parfaitement fluide qui contient néanmoins des erreurs factuelles. Cette tension est d'autant plus problématique que la fluidité du discours constitue un signal de crédibilité (cf. §2.5.2) : un récit bien construit est intuitivement perçu comme plus vrai qu'un récit hésitant (?). Le LLM, en produisant un discours maximale-ment fluide, maximise cette heuristique — indépendamment de la

véracité de ce qu'il affirme.

Les avancées technologiques soulèvent plusieurs défis connexes. L'engagement des élèves peut fluctuer en raison d'effets de nouveauté (?), tandis que les questions de fiabilité, de biais, de confidentialité et d'intégrité académique nécessitent une attention particulière (???). Au-delà de l'établissement de lignes directrices pour l'intégration de l'IA en éducation, des opportunités de recherche existent pour examiner comment l'interaction verbale avec des agents pourrait compléter les pratiques pédagogiques actuelles. La modalité orale, en tirant parti des dynamiques naturelles de la classe, pourrait créer des patterns d'engagement différents comparés aux interactions textuelles individuelles (?).

Ces considérations informent directement notre programme expérimental. L'Étude 2, en exposant les participants à un agent capable de produire des informations historiques incorrectes mais présentées avec fluidité, teste précisément ce risque : la fluidité de l'agent conduit-elle les élèves à accepter des informations fausses et à surestimer leur propre compréhension ?

2.5 Le Phénomène de l'Illusion de Compréhension

La fluidité des agents conversationnels génératifs constitue leur force pédagogique principale — mais aussi leur risque le plus insidieux. Lorsqu'un agent présente l'information de manière parfaitement articulée et convaincante, l'apprenant peut confondre la facilité de réception avec la qualité de sa propre compréhension. Ce phénomène, que nous désignons par *illusion de compréhension* (ou *Illusion of Understanding*, IoU), représente le risque métacognitif central examiné par cette thèse. Cette section analyse d'abord les fondements cognitifs de ce phénomène à travers l'Illusion de Profondeur Explicative (§2.5.1), puis examine le mécanisme de l'heuristique de fluidité (§2.5.2), avant d'explorer comment l'autorité perçue de l'agent peut compromettre la vigilance épistémique (§2.5.3).

2.5.1 Métacognition et Calibration de la Confiance

La métacognition désigne la cognition sur la cognition : la capacité d'un individu à surveiller et réguler ses propres processus cognitifs (?). Dans le contexte de l'ap-

prentissage, cette capacité se manifeste par deux fonctions distinctes : le *monitoring*, qui consiste à évaluer sa propre compréhension, et le *control*, qui permet d'ajuster ses stratégies en conséquence. L'auto-évaluation de la compréhension constitue un processus intrinsèquement faillible (?) : les apprenants croient fréquemment avoir compris un contenu alors même qu'ils échouent à détecter des contradictions évidentes.

Ce déficit métacognitif trouve sa formalisation dans le concept d'*Illusion of Explanatory Depth* (IOED). Ce phénomène désigne la tendance des individus à surestimer la profondeur de leur compréhension des systèmes causaux complexes — jusqu'au moment où ils sont contraints de produire une explication détaillée (?). Le protocole expérimental classique procède en trois phases : une auto-évaluation initiale (T1), la production d'une explication écrite, puis une seconde auto-évaluation (T2). La révélation est systématique : confrontés à l'exercice d'explication, les participants découvrent que leur compréhension était moins profonde qu'ils ne le croyaient, se traduisant par une chute significative entre T1 et T2.

Ce phénomène s'avère pertinent pour l'étude des interactions avec les agents conversationnels. Certaines modalités de présentation peuvent donner l'impression que le matériel est facile à traiter (?). Cette facilité apparente conduit les apprenants à sous-estimer la difficulté de la tâche et à développer une méta-compréhension excessivement optimiste. Le résultat est une dissociation entre la confiance subjective, qui augmente, et l'apprentissage objectif, qui stagne.

Cette illusion s'inscrit dans un déficit métacognitif plus large. Le phénomène décrit par l'effet Dunning-Kruger met en lumière un « double fardeau » : les individus les moins compétents manquent également des compétences métacognitives nécessaires pour reconnaître leur propre incompetence (?). Dans le contexte d'une interaction avec un agent conversationnel fluide, ce déficit devient particulièrement problématique : la facilité apparente de l'échange peut renforcer une confiance injustifiée.

La notion de *calibration* désigne l'écart entre la confiance subjective et la performance objective (?). Les études révèlent une tendance systématique à la *surconfiance* : les apprenants surestiment leur niveau de compréhension. Cette surconfiance constitue un prédicteur robuste de mauvais apprentissage, car elle conduit à un désengagement prématuré de l'effort cognitif.

Dans le contexte des agents alimentés par IA générative, ce problème prend une

dimension nouvelle. Un « paradoxe métacognitif » émerge : bien que l'assistance de l'IA puisse améliorer la performance immédiate, elle dégrade la capacité de l'utilisateur à évaluer cette même performance (?). L'agent, en fournissant des réponses instantanées et fluides, prive l'apprenant des « difficultés désirables » — l'effort de construction et de réorganisation des connaissances — pourtant essentielles à un apprentissage durable (?).

2.5.2 L'Heuristique de Fluidité

Le mécanisme central de l'illusion de compréhension réside dans ce que la littérature désigne par *processing fluency* : l'information facile à percevoir et à traiter cognitivement est vécue comme familière (?). Cette facilité de traitement est alors attribuée par erreur à la propre maîtrise du sujet par l'apprenant plutôt qu'aux qualités de la présentation. L'heuristique de fluidité constitue un raccourci cognitif par lequel nous jugeons plus vraie, plus crédible et mieux comprise l'information qui « passe bien ».

La simple facilité de perception d'un énoncé — police lisible, articulation claire, formulation syntaxiquement simple — augmente sa crédibilité perçue, indépendamment de sa véracité objective. L'individu interprète la facilité cognitive comme un signal de familiarité, et la familiarité comme un indice de vérité. Dans le contexte des agents conversationnels alimentés par LLM, cette heuristique prend une dimension particulière : ces modèles possèdent des caractéristiques techniques qui maximisent la fluidité de traitement — discours instantané, parfaitement structuré et linguistiquement fluide (?). Cette double fluidité — linguistique et auditive lorsque couplée à une synthèse vocale — crée les conditions idéales pour que l'heuristique opère.

Une variante pertinente est l'*instructor fluency effect*, qui décrit comment les comportements non-verbaux de l'enseignant — dynamisme, contact visuel, fluidité verbale — peuvent biaiser les jugements d'apprentissage (?). Les apprenants confondent la qualité de la délivrance pédagogique avec la qualité de leur propre apprentissage, rapportant une confiance élevée sans gains de performance correspondants. La présence visuelle d'un instructeur peut augmenter la satisfaction et l'apprentissage perçu tout en détournant l'attention du contenu (?). L'apprenant se trouve dans une situation paradoxale où son expérience subjective positive masque un apprentissage objectivement dégradé.

Au-delà de la compréhension conceptuelle, l'heuristique de fluidité peut générer une « illusion d'acquisition de compétence » : l'observation passive d'une démonstration

peut conduire l'observateur à confondre la fluidité de traitement visuel avec sa propre capacité à exécuter la tâche (?). Par analogie, un apprenant qui observe un agent expliquer un phénomène avec aisance peut confondre la clarté de la présentation avec sa propre maîtrise du sujet.

La convergence de ces mécanismes crée un risque métacognitif majeur. L'interaction conversationnelle tend à augmenter la crédibilité perçue et à réduire la détection des inexactitudes par rapport au texte statique, car elle active des heuristiques sociales qui diminuent la vigilance critique (?). Même lorsque le contenu n'est pas jugé globalement plus crédible, il est souvent perçu comme plus clair et plus engageant — une qualité qui peut conduire à une acceptation non critique (?). Cette configuration favorise une forme de « paresse métacognitive », où les apprenants délèguent les processus cognitifs coûteux et renoncent à l'effort de construction personnelle des connaissances (?).

2.5.3 L'Autorité de l'Agent et la Vigilance Épistémique

Au-delà de l'illusion de compréhension — qui constitue une erreur d'auto-évaluation —, les agents conversationnels génèrent un second risque épistémique : la surconfiance accordée à la source elle-même. Cette distinction conceptuelle mérite une analyse séparée.

La surconfiance épistémique ne désigne pas une erreur d'auto-évaluation, mais une erreur concernant l'information externe (?). Ce biais se manifeste lorsque les apprenants accordent une confiance excessive à des informations factuellement incorrectes parce que la source apparaît autoritaire ou que l'explication est intuitivement séduisante. Les explications fondées sur des conceptions erronées, parce qu'elles s'alignent avec l'expérience quotidienne, sont souvent jugées plus convaincantes que des explications scientifiquement correctes mais contre-intuitives. Cette dynamique prend une dimension critique face aux « hallucinations » des LLM — la génération d'informations incorrectes présentées avec assurance (?). La fluidité discursive de l'agent crée les conditions d'une acceptation non critique d'informations potentiellement fausses.

L'*illusory truth effect* constitue un mécanisme cognitif qui contribue à cette surconfiance (?). La simple répétition d'un énoncé, même faux, augmente sa crédibilité perçue en accroissant sa familiarité. Dans le contexte d'interactions répétées avec un agent, ce mécanisme peut consolider des croyances erronées initialement introduites par une

hallucination. Il convient toutefois de distinguer cet effet de l'illusion de maîtrise mesurée par le protocole IOED : l'effet de vérité illusoire constitue une variable explicative de la persistance des fausses croyances plutôt qu'une mesure de l'auto-évaluation. Les deux phénomènes peuvent néanmoins se renforcer : un apprenant qui croit maîtriser un sujet sera moins enclin à questionner les informations reçues.

L'anthropomorphisme — notre tendance à attribuer des caractéristiques humaines à des entités non-humaines (?) — joue un rôle central dans l'attribution d'autorité aux agents. L'apparence visuelle déclenche ce processus, mais les travaux récents suggèrent que la fluidité conversationnelle pourrait constituer un signal social de premier ordre (?). Cette dynamique s'avère préoccupante pour les jeunes apprenants, dont la tendance naturelle à l'anthropomorphisme est plus prononcée (?). Les adolescents, dont les modèles mentaux de la technologie sont en développement, sont plus dépendants des indices de surface pour évaluer la crédibilité (?).

La convergence de ces mécanismes — fluidité de traitement, autorité perçue, anthropomorphisme — tend à désactiver la vigilance épistémique de l'apprenant. Les explications trompeuses générées par l'IA peuvent être plus persuasives que les explications honnêtes (?). Le « halo de confiance » observé dans les interactions avec les agents — où les élèves généralisent la compétence perçue à l'ensemble du domaine — agit comme un « bouclier de crédibilité ». L'apprenant, ayant catégorisé l'agent comme une source fiable, cesse d'appliquer les filtres critiques qu'il mobiliserait face à une source dont l'autorité n'est pas établie. Ce phénomène de « négligence épistémique » opère de manière inconsciente, échappant à la régulation métacognitive de l'individu.

L'analyse de l'illusion de compréhension révèle une tension fondamentale : les mêmes caractéristiques qui rendent les agents conversationnels engageants — fluidité, réalisme, crédibilité — peuvent simultanément compromettre la profondeur de l'apprentissage en désactivant les mécanismes de vigilance critique. Cette tension entre engagement et vigilance constitue le fil conducteur de notre problématique de recherche, qu'il convient maintenant de formuler explicitement.

2.6 Synthèse et Problématique

Les sections précédentes ont établi le cadre théorique nécessaire à l'analyse de l'interaction entre élèves et agents conversationnels historiques. Il convient maintenant de synthétiser ces apports pour faire émerger les tensions qui structurent notre problématique. Cette section identifie d'abord les convergences et tensions théoriques (§2.6.1), puis analyse les lacunes de la littérature actuelle (§2.6.2), avant de formuler les questions et hypothèses de recherche qui guident ce travail doctoral (§2.6.3).

2.6.1 Convergence et Tensions Théoriques

L'analyse de la littérature révèle une tension fondamentale : les mêmes caractéristiques qui rendent les agents conversationnels pédagogiquement prometteurs constituent simultanément leurs principaux risques.

Du côté des convergences positives, la littérature établit que les agents conversationnels incarnés disposent de leviers d'engagement multiples et cohérents. Le cadre ICAP prédit que l'interactivité dialogique active les modes d'engagement les plus profonds (cf. §2.1.4). La théorie de l'agence sociale suggère que la présence d'un agent personnifié déclenche un « contrat de partenariat » qui augmente l'effort cognitif (cf. §??). Le paradigme CASA explique pourquoi même des indices sociaux minimaux suffisent à activer nos scripts relationnels (cf. §??). La théorie de l'autodétermination identifie l'interactivité comme un vecteur de satisfaction du besoin d'autonomie (cf. §2.1.1). Ces cadres convergent vers une prédiction commune : un agent conversationnel interactif devrait favoriser l'engagement et l'intérêt des apprenants.

Du côté des tensions, l'analyse révèle que les caractéristiques mêmes qui génèrent l'engagement peuvent simultanément compromettre la qualité de l'apprentissage. La fluidité conversationnelle qui rend l'agent engageant active également l'heuristique de fluidité, conduisant l'apprenant à confondre facilité de traitement et maîtrise du contenu (cf. §2.5.2). L'incarnation réaliste qui déclenche les mécanismes d'agence sociale peut agir comme un « détail séduisant » qui détourne l'attention du contenu. L'autorité perçue de l'agent, qui favorise l'attention initiale, peut désactiver la vigilance épistémique nécessaire à un apprentissage critique (cf. §2.5.3).

Cette tension suggère ce que nous proposons d'appeler l'*hypothèse de l'équilibre des*

défauts symétriques. Dans cette perspective, l'interactivité et l'incarnation constituent des leviers robustes pour l'engagement — mais chaque gain sur cette dimension s'accompagne potentiellement d'un risque métacognitif. L'agent idéal ne serait pas celui qui maximise l'engagement, mais celui qui optimise le rapport entre engagement généré et risque d'illusion induit.

Cette hypothèse trouve un écho dans les recommandations de la littérature. Plutôt que de concevoir des outils « transparents » qui s'effacent pour laisser place à la tâche (?), certains auteurs suggèrent de « scripter la rupture » : concevoir des agents qui, loin de fournir des réponses fluides et directes, exposent délibérément leurs limites, admettent leur incertitude, ou forcent l'apprenant à valider l'information (?). L'objectif n'est plus de maximiser la confiance, mais de la *calibrer* en maintenant active la vigilance critique.

Ces tensions théoriques se traduisent en questions de design concrètes. Comment concevoir un agent suffisamment engageant pour maintenir l'attention, sans que cette qualité ne désactive les mécanismes de vigilance ? L'incarnation visuelle amplifie-t-elle le risque métacognitif, ou la fluidité conversationnelle constitue-t-elle à elle seule le facteur déterminant ? L'alignement thématique entre le personnage et le contenu peut-il servir de levier d'engagement sans les coûts attentionnels d'une incarnation réaliste ? Ces questions structurent directement notre programme de recherche.

2.6.2 Les Lacunes de la Littérature Actuelle

L'analyse systématique de la littérature sur les agents pédagogiques révèle plusieurs lacunes que ce travail doctoral vise à combler.

Sous-représentation des populations K-12. Les méta-analyses récentes (??) convergent sur un constat : la majorité des études portent sur des populations universitaires, avec une sous-représentation significative des élèves du secondaire. Cette lacune est problématique à double titre. D'une part, les mécanismes motivationnels et métacognitifs varient avec le développement ; les résultats obtenus avec des adultes ne sont pas directement transférables aux adolescents. D'autre part, c'est précisément durant l'adolescence que se cristallisent les attitudes envers les disciplines scolaires ; intervenir à ce stade présente un potentiel d'impact maximal.

Absence d'études dans le domaine de l'Histoire. La littérature se concentre massivement sur les domaines STIM, avec des effets particulièrement marqués en biologie

et en informatique (?). L'Histoire reste un terrain quasi inexploré. Une méta-analyse a identifié un effet négatif des agents pédagogiques en histoire ($g^+ = -0.80$), mais ce résultat isolé repose sur un nombre limité d'études et nécessite une investigation approfondie. La spécificité épistémologique de l'Histoire — discipline interprétative où la narrativité joue un rôle central — justifie une attention particulière.

Études antérieures à l'ère des LLM. La quasi-totalité des travaux fondateurs (??) a été conduite avec des agents scriptés. L'avènement des grands modèles de langage en 2023 a changé la donne : les agents peuvent désormais générer des réponses contextuellement adaptées avec une fluidité conversationnelle sans précédent. Les conclusions des études antérieures, notamment sur l'importance relative de l'apparence visuelle versus les capacités fonctionnelles, méritent d'être réexaminées dans ce contexte.

Mesures focalisées sur les niveaux cognitifs inférieurs. Les protocoles d'évaluation se concentrent sur la mémorisation et la compréhension de surface (?). Les niveaux cognitifs supérieurs de la taxonomie de Bloom restent peu explorés. Plus fondamentalement, les mesures métacognitives sont rares : peu d'études évaluent explicitement la calibration de la confiance ou l'illusion de compréhension. Cette lacune est d'autant plus problématique que l'illusion de maîtrise constitue potentiellement le risque majeur des agents génératifs.

Absence d'études longitudinales. La durée médiane des interventions est de l'ordre de 30 minutes à quelques heures (?). Cette brièveté empêche d'évaluer les effets d'apprentissage durables, la rétention à long terme, et l'évolution de la relation élève-agent. L'effet de nouveauté n'est jamais contrôlé, confondant enthousiasme initial et efficacité durable. Des études longitudinales sont nécessaires pour établir la valeur pédagogique réelle des agents conversationnels.

Ce travail doctoral vise à combler ces lacunes en conduisant des études expérimentales auprès d'élèves du secondaire français, dans le domaine de l'Histoire, avec des agents alimentés par IA générative, et en intégrant des mesures métacognitives explicites (protocole IOED). Les limites temporelles (sessions uniques) constituent une contrainte assumée, appelant des travaux complémentaires ultérieurs.

2.6.3 Questions et Hypothèses de Recherche

Sur la base des convergences théoriques, des tensions identifiées et des lacunes de la littérature, nous formulons trois questions de recherche qui structurent ce travail doctoral.

Question de Recherche 1 (QR1)

Dans quelle mesure l'interactivité d'un agent conversationnel incarné influence-t-elle l'intérêt des élèves pour l'activité, le contenu historique et le personnage ?

Cette question interroge le potentiel d'engagement des agents conversationnels. La littérature prédit un effet positif de l'interactivité sur l'intérêt, mais cet effet reste à démontrer empiriquement dans le contexte spécifique de l'enseignement de l'Histoire auprès d'adolescents français. Nous formulons l'hypothèse suivante :

[H1] L'interactivité de l'agent génère un effet positif significatif sur l'intérêt des élèves, mesurable sur les trois dimensions (activité, contenu, personnage), indépendamment du niveau scolaire.

Question de Recherche 2 (QR2)

L'alignement thématique entre le personnage historique et le contenu pédagogique module-t-il l'effet de l'interactivité sur l'intérêt ?

Cette question explore l'hypothèse selon laquelle l'efficacité des agents dépend de leur cohérence thématique avec le contenu. Un personnage aligné (Napoléon enseignant l'Empire) pourrait amplifier l'engagement par rapport à un agent neutre, en activant des mécanismes d'empathie historique et de pertinence perçue. Nous formulons l'hypothèse suivante :

[H2] L'alignement thématique entre le personnage et le contenu amplifie l'effet positif de l'interactivité sur l'intérêt, avec une interaction significative entre les deux facteurs.

Question de Recherche 3 (QR3)

L'apparence de l'agent (humanoïde vs. abstrait) influence-t-elle la propension des élèves à l'illusion de compréhension, leur confiance et la crédibilité perçue des informations délivrées ?

Cette question aborde le risque métacognitif central de la thèse. La littérature sur le paradigme CASA et la théorie de l'agence sociale prédit qu'un agent humanoïde devrait générer une confiance et une crédibilité accrues. L'hypothèse concurrente, issue de l'analyse de l'heuristique de fluidité, suggère que la fluidité conversationnelle pourrait constituer le signal dominant, éclipsant l'effet de l'apparence. Nous formulons les hypothèses suivantes :

[H3a] L'agent humanoïde induit une confiance et une persuasion perçue supérieures à l'agent abstrait.

[H3b] L'agent humanoïde amplifie l'illusion de compréhension par rapport à l'agent abstrait, mesurable par une augmentation de l'écart entre confiance subjective (IOED T3) et performance objective.

Articulation des études

Ces questions de recherche sont traitées par deux études empiriques complémentaires. L'**Étude 1** (Chapitre ??) examine les effets de l'interactivité et de l'alignement thématique sur l'intérêt (QR1, QR2) à travers trois expérimentations auprès de populations d'âges différents (4^{ème}, 6^{ème}, Terminale). L'**Étude 2** (Chapitre ??) examine l'effet du design visuel sur la confiance et l'illusion de compréhension (QR3) auprès d'élèves de 5^{ème}. Cette articulation permet d'explorer séparément les dimensions d'engagement et de risque métacognitif, avant de les confronter dans la discussion générale.

Les questions de recherche ainsi formulées appellent une méthodologie expérimentale rigoureuse, capable de manipuler les variables de design identifiées tout en mesurant leurs effets sur l'engagement et les biais métacognitifs. Avant de présenter les études empiriques, il convient de décrire la plateforme technique développée pour les conduire : MemorIA, dont l'architecture et la validation font l'objet du chapitre suivant.

