



RÉSEAU DES ÉTABLISSEMENTS D'ENSEIGNEMENT SUPÉRIEUR DU DÉVELOPPEMENT DURABLE
CRÉATEUR D'AVENIR

Guide n°10

Esprit critique des étudiants à l'ère de l'IA

Entretenir et développer l'esprit critique des étudiants à l'ère de l'IA

Guide à l'attention des établissements du RESDD



RÉSEAU DES ÉTABLISSEMENTS D'ENSEIGNEMENT
SUPÉRIEUR DU DÉVELOPPEMENT DURABLE
CRÉATEUR D'AVENIR



INSTITUT
POLYTECHNIQUE
DE PARIS

ENTPE
L'école de l'aménagement durable des territoires



ENM
École Nationale
de la Météorologie



École
du Service Public
de la Mer



Université
Gustave Eiffel



Juin 2026

Ce guide est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International. Pour consulter une copie de cette licence, rendez-vous sur <https://creativecommons.org/licenses/by-nc/4.0/>

Synthèse

L'IA générative produit des réponses fluides, structurées et apparemment fiables, ce qui peut fragiliser l'exercice de l'esprit critique : elle crée une illusion de compétence et d'autorité qui affaiblit les réflexes de vérification.

Les travaux récents suggèrent que l'usage régulier de l'IA peut favoriser un glissement du délestage cognitif, initialement limité à l'exécution tout en préservant la maîtrise du jugement, vers une délégation plus large du raisonnement lui-même. Ce glissement peut affaiblir progressivement les capacités de jugement critique. L'essor de l'IA agentique, qui prend en charge des processus et des raisonnements complets, et non plus seulement des tâches ponctuelles, pourrait accentuer ce risque.

L'enjeu est majeur pour les établissements du RESDD. Pour les étudiants, un esprit critique affaibli risque d'altérer leur parcours professionnel. Pour les établissements, c'est la réussite de leur mission même d'enseignement, et la valeur du diplôme, qui sont en jeu. L'esprit critique est aussi la condition de l'apprentissage permanent, qui devient la norme à l'ère de l'IA.

Il s'agit aussi d'une opportunité : à mesure que l'IA démocratise la capacité à produire, la capacité à juger pourrait devenir de plus en plus centrale dans l'exercice des métiers d'ingénieurs. La combinaison entre esprit critique, expertise technique et maîtrise de l'IA devient un avantage distinctif

Trois principes d'action sont proposés aux établissements : ancrer l'esprit critique dans les formations disciplinaires, miser sur les activités pratiques régulières, et calibrer la vigilance en fonction des enjeux.

Ce guide propose d'agir sur trois dimensions de l'esprit critique appliqué aux productions de l'IA :

1. Critiquer les réponses : détecter les erreurs factuelles, les hallucinations, les biais, les approximations, les manques.
2. Critiquer les sources : tracer la provenance des informations, distinguer les faits des opinions, croiser avec des sources primaires.
3. Critiquer le système : comprendre les mécanismes de production de l'IA, ses cheminements de pensée (CoT), ses biais systémiques, ses limites structurelles, et les enjeux sociétaux de son déploiement.

Six recommandations d'actions opérationnelles sont proposées :

Priorité	#	Recommandation	Impact en jeu
Indispensable	1	Former les étudiants à détecter les erreurs et hallucinations de l'IA	Fiabilité du travail académique
Indispensable	2	Intégrer systématiquement la vérification des sources et du raisonnement dans les travaux utilisant l'IA	Rigueur intellectuelle
Recommandé	3	Concevoir des activités d'« audit critique » de l'IA dans chaque discipline	Culture de la vérification
Recommandé	4	Former les étudiants aux biais systémiques de l'IA et à leurs conséquences disciplinaires	Conscience des limites structurelles
Souhaitable	5	Développer l'esprit critique à l'ère de l'IA agentique : auditer un processus, pas seulement un résultat	Préparation aux usages avancés
Souhaitable	6	Évaluer et piloter le développement de l'esprit critique à l'ère de l'IA	Amélioration continue

Principaux livrables opérationnels du guide : exercices de détection d'erreurs (A), checklist de vérification des sources (B), fiches d'activités d'audit critique par discipline (C), fiche pédagogique sur les biais systémiques (D), exercice d'audit d'un processus agentique (E), indicateurs de pilotage (F).

Table des matières

Synthèse.....	2
I. Esprit critique et IA : quels enjeux ?	4
1.1. Le problème : l'IA affaiblit l'esprit critique et la capacité de jugement.....	4
1.2. Les enjeux : ce que les étudiants et les établissements ont à perdre - et à gagner	5
1.3. Trois leviers pour développer l'esprit critique à l'ère de l'IA.....	6
II. Six recommandations d'actions opérationnelles à mener	7
Principes d'action et vue d'ensemble	7
#1. Former les étudiants à détecter les erreurs et hallucinations de l'IA.....	8
#2. Intégrer systématiquement la vérification des sources et du raisonnement dans les travaux utilisant l'IA	8
#3. Concevoir des activités d'« audit critique » de l'IA dans chaque discipline	9
#4. Former les étudiants aux biais systémiques de l'IA et à leurs conséquences disciplinaires.....	9
#5. Développer l'esprit critique à l'ère de l'IA agentique	10
#6. Évaluer et piloter le développement de l'esprit critique à l'ère de l'IA.....	11
III. Références utiles.....	12
BOITE A OUTILS.....	13
A - Exercices de détection d'erreurs et hallucinations de l'IA	13
B - Checklist de vérification des sources pour les travaux utilisant l'IA.....	13
C - Idées d'activités d'audit critique de l'IA par discipline	14
D - Fiche pédagogique : biais systémiques de l'IA	15
E - Exercice d'audit d'un processus agentique.....	16
F - Indicateurs de pilotage de l'esprit critique à l'ère de l'IA	16

I. Esprit critique et IA : quels enjeux ?

1.1. Le problème : l'IA affaiblit l'esprit critique et la capacité de jugement

Une fluidité trompeuse

La particularité de l'IA générative est de produire des réponses qui présentent une forte apparence de fiabilité : syntaxe correcte, structure logique, ton assuré, parfois même références bibliographiques. Cette fluidité est trompeuse, car elle peut masquer des faiblesses structurelles : hallucinations, biais, manque de profondeur ou absence de contextualisation.

Or la **fluidité linguistique de l'IA a des effets importants, car elle agit comme un « biais d'autorité »** : plus la réponse est bien écrite, plus l'étudiant tend à la considérer comme fiable. L'étude de Thesen & Park (2025) montre que les étudiants manquent souvent de l'expertise nécessaire pour détecter les hallucinations, en particulier dans les domaines où ils sont encore novices.

De la confiance excessive à la délégation du raisonnement

Le problème ne se limite pas à la difficulté de détecter les erreurs. Shaw et Nave (2026) décrivent un phénomène plus profond : la **délégation du raisonnement**. Dans leur expérimentation contrôlée, les participants consultent fréquemment l'IA pour résoudre des problèmes de réflexion et suivent très souvent ses réponses, y compris lorsqu'elles sont erronées. Plus préoccupant encore, le simple fait d'avoir consulté l'IA tend à renforcer la confiance accordée à la réponse retenue, même quand celle-ci est fausse¹.

L'usage de l'IA peut d'abord relever d'un simple délestage cognitif : l'étudiant délègue une tâche d'exécution - rechercher une information, reformuler un texte, effectuer un calcul - tout en conservant la capacité d'évaluer la qualité du résultat. Dans cette configuration, comme avec une calculatrice, l'outil assiste l'activité sans se substituer au jugement.

Mais les données expérimentales suggèrent que l'ampleur de ce que l'on délègue à l'IA peut augmenter avec le temps : on commence par déléguer l'exécution, puis la confiance dans l'IA augmente avec l'usage, la vigilance diminue, et le délestage cognitif peut se transformer insensiblement en délégation du raisonnement. L'utilisateur en arrive à accepter la réponse de l'IA sans l'examiner, parce qu'elle arrive vite, qu'elle est bien formulée et qu'elle *semble* juste. Il ne s'agit plus seulement d'une délégation de l'exécution, mais d'une délégation du raisonnement lui-même.

Or l'esprit critique repose précisément sur l'exercice d'un raisonnement délibéré : comparer, douter, vérifier, confronter, reformuler. Si ces opérations sont moins exercées, la capacité de jugement peut s'émousser.

Un **cercle vicieux** peut alors s'installer : une réponse jugée satisfaisante réduit l'effort de vérification ; la délégation à l'IA s'élargit ; l'exercice du jugement critique se raréfie ; la dépendance à l'outil augmente ; la capacité de jugement s'affaiblit en retour.

L'IA agentique amplifie ce risque de délégation du raisonnement.

Là où l'IA générative classique répond à une question ponctuelle, un agent IA prend en charge un processus complet - recherche, analyse, rédaction, mise en forme - et produit un livrable complet, fini. Le passage du « *je délègue une question* » au « *je délègue un raisonnement entier* » devient alors encore plus naturel et moins visible pour l'utilisateur.

¹ On distingue trois phénomènes : le délestage cognitif (délégation de l'exécution), la reddition cognitive (délégation du raisonnement), et la dette cognitive (affaiblissement progressif des capacités d'apprentissage).

Un affaiblissement de l'esprit critique qui dépasse la relation à l'IA elle-même

Les études récentes suggèrent que **l'usage de l'IA peut affaiblir l'exercice du jugement critique** en réduisant la vigilance, l'effort de vérification et la part de raisonnement effectivement assumée par l'utilisateur (Carnegie Mellon & Microsoft, 2025, Shaw & Nave, 2026).

Ce constat établit un lien profond avec l'effort cognitif (cf. Guide #9) : l'esprit critique exige un effort de réflexion soutenu, et un étudiant qui a pris l'habitude de moins fournir cet effort exerce moins bien son esprit critique, même s'il en connaît les principes. Effort cognitif et esprit critique sont les deux faces, complémentaires, d'une même pièce.

1.2. Les enjeux : ce que les étudiants et les établissements ont à perdre - et à gagner

Pour les étudiants, un esprit critique affaibli a des **conséquences professionnelles** directes. Un ingénieur qui ne vérifie pas un bilan carbone produit par l'IA, un urbaniste qui accepte sans recul une analyse d'impact générée automatiquement, un prévisionniste qui ne questionne pas une modélisation assistée : dans chaque cas, l'insuffisance d'esprit critique peut conduire à des décisions erronées aux conséquences environnementales, sanitaires ou économiques réelles.

Pour les établissements, l'enjeu du développement de l'esprit critique des étudiants touche à leur **mission d'enseignement** elle-même et à la **valeur du diplôme** qu'ils délivrent.

Mais **l'enjeu n'est pas seulement défensif**. L'IA permet désormais à des profils non techniques de produire des analyses ou des modélisations qui ont l'apparence du travail d'ingénieur. Seul un professionnel doté d'une expertise disciplinaire et d'un esprit critique exercé peut évaluer si une modélisation est physiquement plausible, si un bilan est complet, si une analyse de risque tient compte de tous les paramètres pertinents.

La valeur distinctive des professionnels de demain ne réside pas dans leur capacité à produire - l'IA risque de le faire de plus en plus vite - mais dans leur capacité à **juger**.

L'esprit critique doit également être envisagé comme une compétence transversale majeure. Dans des environnements de travail de plus en plus assistés par l'IA, il constitue une aptitude clé pour interagir avec des systèmes automatisés, en comprendre les limites et en encadrer les usages. À ce titre il s'impose comme une *soft skill* stratégique, appelée à être très valorisée sur le marché du travail.

Par ailleurs, l'esprit critique est indissociable de la capacité à apprendre. Or, dans un contexte où l'apprentissage tout au long de la vie tend à supplanter la seule formation initiale, il devient une condition essentielle de l'adaptabilité et de l'employabilité durable.

Ainsi, le profil des diplômés des établissements du RESDD à l'ère de l'IA repose sur un triptyque de compétences indissociables : l'expertise technique disciplinaire, la maîtrise raisonnée de l'IA (fluency IA), et l'esprit critique, le sens du jugement et la créativité. **À mesure que l'IA démocratise la capacité à produire, la capacité à juger apparaît comme le véritable facteur différenciant des professionnels de demain. Ce triptyque constitue le fil directeur de l'ensemble des recommandations du GT RESDD IA (cf. Guides #4, #8, #9, #11).**

Une opportunité : vers une culture émergente de la vigilance face à l'IA

L'essor de l'IA constitue aussi une opportunité pédagogique majeure. Sa démocratisation s'accompagne déjà d'un début de réflexe de méfiance chez les utilisateurs, notamment sur les réseaux sociaux où l'on observe des comportements émergents : volonté de demandes explicites de sources, doute face à des contenus trop spectaculaires, ou remise en question de la véracité de certaines vidéos.

Cette forme de scepticisme, imparfaite mais en progression, peut servir de point d'appui stratégique. Elle ouvre une fenêtre pour structurer et renforcer ces intuitions en véritables compétences :

apprentissage de la vérification, analyse des sources, compréhension des biais. Plutôt que de lutter contre l'usage de l'IA, il devient alors possible de capitaliser sur cette vigilance naissante pour développer un esprit critique outillé et durable.

1.3. Trois leviers pour développer l'esprit critique à l'ère de l'IA

L'esprit critique ne se développe pas spontanément par l'usage de l'IA ; au contraire, comme analysé précédemment, un usage non encadré tend à l'affaiblir (Carnegie Mellon & Microsoft, 2025). **Il doit donc être explicitement enseigné et exercé** - et il doit l'être dans les bonnes conditions.

Dans cette perspective, trois constats méritent d'être mis en évidence :

1. **Les étudiants novices dans un domaine sont les plus vulnérables** : ils n'ont pas l'expertise pour détecter les erreurs de l'IA dans leur discipline (Thesen & Park, 2025). L'esprit critique ne peut donc pas être enseigné dans l'abstrait - c'est en critiquant une réponse IA sur un sujet que l'on maîtrise progressivement qu'il se développe efficacement. Cela implique d'ancrer la formation à l'esprit critique dans les enseignements disciplinaires, pas dans un module séparé.
2. **Les exercices d'« audit critique » -analyser, corriger et commenter une production IA- sont parmi les activités les plus efficaces pour développer l'esprit critique disciplinaire** (Vangrunderbeeck, 2024). Le croisement avec des sources primaires est le réflexe le plus protecteur, mais c'est aussi le moins spontané chez les étudiants. Ces résultats plaident pour des activités pratiques et régulières plutôt que pour un enseignement déclaratif sur les biais de l'IA.
3. **L'esprit critique est cognitivement coûteux** : vérifier chaque réponse est chronophage et exigeant. L'objectif n'est donc pas d'exiger une vigilance absolue en toute circonstance, mais d'apprendre aux étudiants à calibrer leur niveau de vérification en fonction des enjeux, des risques et du contexte d'usage.

Ces constats conduisent à structurer l'action autour de trois dimensions complémentaires et progressives :

- **Dimension 1 - Critiquer les réponses** : détecter les erreurs factuelles, les hallucinations, les biais, les approximations, les manques. C'est le prérequis immédiat : chaque étudiant doit acquérir le réflexe de vérification.
- **Dimension 2 - Critiquer les sources** : tracer la provenance des informations, distinguer les faits des opinions, croiser avec des sources primaires. Cette dimension développe la rigueur intellectuelle au-delà de la seule interaction avec l'IA.
- **Dimension 3 - Critiquer le système** : comprendre les mécanismes de production de l'IA, ses biais systémiques, ses « *chains of thoughts* », ses limites structurelles et les enjeux sociétaux de son déploiement. Cette dimension construit une compréhension transférable à tout outil futur.

Ces trois dimensions structurent les recommandations de la partie II.

Pour que ces leviers fonctionnent, plusieurs conditions doivent être réunies :

1. **Former les enseignants** : ceux-ci doivent eux-mêmes connaître l'IA et savoir détecter ses erreurs dans leur discipline, faute de quoi ils ne pourront pas intégrer la dimension critique dans leurs enseignements ;
2. **Préserver les conditions cognitives d'exercice du jugement** : un étudiant qui a perdu l'habitude de réfléchir par lui-même ne peut pas exercer son esprit critique, même formé (cf. Guide #9). Trois leviers pédagogiques peuvent être mobilisés :
 - **apprendre à utiliser l'IA de façon à obtenir les sources et le raisonnement** en plus de la réponse afin de pouvoir se faire un avis sur la qualité de ceux-ci,

- **montrer l'IA qui se trompe** (les démonstrations où l'IA produit des réponses convaincantes mais fausses sont pédagogiquement puissantes),
 - **valoriser le doute** (un étudiant qui questionne une réponse IA et identifie une erreur démontre une compétence de haut niveau) ;
3. **Apprendre à interroger l'IA de manière critique** : les étudiants doivent être formés à ne pas se limiter à la réponse produite par l'IA, mais à exiger des éléments de transparence permettant son évaluation. Cela implique de savoir demander les sources, les hypothèses, les étapes de raisonnement ou les limites d'une analyse. Par exemple, plutôt que de solliciter uniquement une réponse, l'étudiant peut formuler une requête du type : « *Donne ta réponse en explicitant les hypothèses retenues, les étapes de raisonnement et les sources mobilisées, puis indique les principales limites ou incertitudes.* » Cette posture transforme l'usage de l'IA en un exercice d'analyse critique plutôt qu'en simple délégation ;
 4. **Mesurer pour piloter** : l'érosion de l'esprit critique est un phénomène **invisible à court terme**. Les établissements doivent donc se doter d'outils de mesure spécifiques (exercices d'audit, comparaisons avec et sans IA, enquêtes de perception...) pour évaluer l'ampleur du problème et l'efficacité des solutions mises en œuvre (cf. recommandation #6).

Deux points de vigilance doivent enfin être gardés à l'esprit :

- le cercle vicieux de l'expertise (il faut des connaissances pour critiquer l'IA, mais l'IA tend à court-circuiter l'acquisition de ces connaissances),
- et la course à la plausibilité (l'amélioration continue des modèles rend les erreurs plus subtiles et plus difficiles à détecter).

II. Six recommandations d'actions opérationnelles à mener

Principes d'action et vue d'ensemble

Trois principes guident les recommandations :

- **Ancrer dans la discipline** : l'esprit critique se développe d'abord dans des contextes disciplinaires situés.
- **Pratiquer régulièrement** : l'esprit critique se construit par des exercices répétés de vérification, de comparaison et de correction.
- **Calibrer la confiance** : il ne s'agit pas de tout vérifier en permanence, mais d'ajuster le niveau de vigilance à l'importance des enjeux.

Six recommandations d'actions opérationnelles sont proposées :

Priorité	#	Recommandation	Qui	Impact en jeu	Effort	Outils
Indispensable	1	Former les étudiants à détecter les erreurs et hallucinations de l'IA	Enseignants + réf. IA	Fiabilité du travail académique	1-2j	A
Indispensable	2	Intégrer systématiquement la vérification des sources et du raisonnement dans les travaux utilisant l'IA	Enseignants	Rigueur intellectuelle	Faible	B

Priorité	#	Recommandation	Qui	Impact en jeu	Effort	Outils
Recommandé	3	Concevoir des activités d'« audit critique » de l'IA dans chaque discipline	Enseignants + réf. IA	Culture de la vérification	Variable	C
Recommandé	4	Former les étudiants aux biais systémiques de l'IA et à leurs conséquences disciplinaires	Enseignants + réf. IA	Conscience des limites structurelles	3-6h	D
Souhaitable	5	Développer l'esprit critique à l'ère de l'IA agentique : auditer un processus, pas seulement un résultat	Enseignants + réf. IA + DSI	Préparation aux usages avancés	3-8h	E
Souhaitable	6	Évaluer et piloter le développement de l'esprit critique à l'ère de l'IA	Dir. études + réf. IA	Amélioration continue	intégré au pilotage	F

#1. Former les étudiants à détecter les erreurs et hallucinations de l'IA

Quoi : dès le module de sensibilisation (Guide #8), puis de façon récurrente dans les enseignements disciplinaires, les étudiants apprennent à identifier les types d'erreurs que l'IA commet dans leur domaine.

Cinq types d'erreurs à faire reconnaître et analyser :

- **Hallucinations factuelles :** l'IA invente des faits, des chiffres, des dates, des noms.
- **Fausse/mauvaises références :** l'IA génère des citations bibliographiques inexistantes ou peu fiables mais plausibles.
- **Erreurs de raisonnement :** la logique est apparemment correcte mais contient un saut ou une erreur cachée.
- **Biais de perspective :** la réponse privilégie un angle au détriment d'autres, sans le signaler.
- **Approximations disciplinaires :** la réponse est « à peu près correcte » pour un non-spécialiste mais imprécise ou erronée pour un expert du domaine.

Impact attendu : les étudiants acquièrent le réflexe de vérification. La confiance aveugle dans l'IA est remplacée par une confiance calibrée.

Qui : enseignants + référent IA.

Base de travail disponible : voir exercices de détection d'erreurs en Boîte à outils A.

#2. Intégrer systématiquement la vérification des sources et du raisonnement dans les travaux utilisant l'IA

Quoi : tout travail utilisant l'IA inclut une obligation de vérification à la fois des sources et du raisonnement. L'étudiant doit démontrer qu'il a vérifié les informations clés auprès de sources primaires (articles scientifiques, données officielles, textes réglementaires), mais aussi qu'il a examiné la cohérence du raisonnement proposé (hypothèses, enchaînement logique, validité des inférences, prise en compte des limites).

Règles de base à communiquer aux étudiants :

- Ne jamais citer une référence bibliographique fournie par l'IA sans l'avoir vérifiée : elle peut être inexacte, incomplète ou totalement inexistante.
- Vérifier au moins les trois faits les plus importants de toute réponse IA utilisée dans un travail noté.
- Examiner le raisonnement proposé par l'IA : identifier les hypothèses implicites, vérifier la cohérence des étapes et repérer d'éventuelles simplifications abusives ou erreurs logiques.
- Distinguer explicitement dans le rendu ce qui provient de l'IA, ce qui a été vérifié et ce qui est un apport personnel.

Impact attendu : le réflexe de vérification devient un automatisme. La rigueur intellectuelle est renforcée.

Qui : enseignants (à intégrer dans les consignes de chaque travail utilisant l'IA).

Base de travail disponible : voir *checklist* de vérification en Boîte à outils B.

#3. Concevoir des activités d'« audit critique » de l'IA dans chaque discipline

Quoi : il est recommandé que chaque enseignant engagé dans la démarche conçoive, à terme, au moins une activité dans laquelle les étudiants évaluent de manière critique une production de l'IA dans le contexte de leur discipline. L'« *audit critique* » est l'exercice pédagogique le plus efficace pour développer l'esprit critique disciplinaire. Parmi les formats recommandés :

- **Audit de réponse :** l'enseignant soumet un problème disciplinaire à l'IA, puis les étudiants analysent la réponse : qu'est-ce qui est correct ? qu'est-ce qui est faux ou imprécis ? qu'est-ce qui manque ?
- **Correction commentée :** les étudiants reçoivent une réponse IA et doivent la corriger en justifiant chaque modification.
- **Comparaison entre modèles :** les étudiants soumettent la même question à deux modèles différents et analysent les différences (cohérence, profondeur, erreurs).
- **Audit d'un processus agentique :** pour les étudiants avancés, analyser le résultat d'un agent IA et retracer son processus de décision.

Impact attendu : développement de l'esprit critique disciplinaire. Ces exercices ont aussi pour intérêt de rendre visibles, simultanément, certaines limites de l'IA et certaines lacunes de compréhension chez les étudiants. Au-delà de la formation, le réflexe d'audit critique est une compétence professionnelle permanente : à mesure que l'IA s'intègre dans les pratiques de travail, la capacité à évaluer critique ses productions deviendra un geste professionnel quotidien, au même titre que le contrôle qualité d'un livrable.

Qui : enseignants volontaires, accompagnés par le référent IA.

Base de travail disponible : voir la liste des activités d'audit en Boîte à outils C.

#4. Former les étudiants aux biais systémiques de l'IA et à leurs conséquences disciplinaires

Quoi : au-delà des erreurs ponctuelles, les étudiants doivent comprendre les biais structurels de l'IA et leurs implications dans leur discipline.

Biais à aborder :

- **Biais d'hallucination :** production d'informations fausses mais très convaincantes (références inexistantes, chiffres inventés, mécanismes erronés) sans signaler d'incertitude.

- **Biais de données d'entraînement** : surreprésentation de certaines langues, cultures, perspectives. Cela inclut également la présence de biais sociaux (stéréotypes sexistes, racistes, culturels, socio-économiques) que l'IA peut reproduire, voire amplifier. Sous-représentation des données locales, des contextes francophones, des spécificités du développement durable.
- **Biais d'autorité (ou effet de crédibilité)** : La formulation fluide, structurée et assurée donne une impression de fiabilité, même lorsque le contenu est discutable ou faux.
- **Biais de « lissage »** : l'IA tend à produire des réponses « moyennes » et consensuelles, évite certaines positions jugées sensibles ou propose des réponses « socialement acceptables », gommant les controverses, les positions minoritaires et les nuances, et reproduit des schémas de réponse standardisés, parfois inadaptés à des cas spécifiques ou locaux.
- **Biais de simplification excessive** : l'IA tend à simplifier des phénomènes complexes (notamment en environnement, économie, santé), en omettant des variables importantes, des effets systémiques ou des ordres de grandeur. Proche du biais de lissage, mais ici il s'agit d'une perte de complexité explicative.
- **Biais de complétude apparente** : L'IA donne une réponse qui semble complète et structurée, même lorsqu'elle est partielle ou qu'il manque des éléments essentiels.
- **Biais temporel** : les données d'entraînement ont une date de coupure. Les informations récentes, les réglementations en vigueur, les données actualisées peuvent être absentes ou obsolètes.
- **Biais de confirmation ou d'acquiescement** : l'IA tend souvent à suivre l'hypothèse implicite contenue dans la question, plutôt qu'à la remettre en cause ou à en tester les présupposés.
- **Biais de cadrage** : La réponse dépend fortement de la manière dont la question est posée. Certains aspects sont mis en avant, d'autres invisibilisés, sans que cela soit explicite. Complément du biais de confirmation, mais côté "construction de la réponse".
- **Biais de complaisance (ou d'approbation)** : L'IA tend à adopter un ton positif, encourageant et parfois excessivement validant vis-à-vis des propositions de l'utilisateur, même lorsque celles-ci sont discutables, incomplètes ou incorrectes. Elle peut ainsi sous-pondérer la critique, les objections ou les risques, avec pour effet le renforcement artificiel de la confiance de l'utilisateur dans ses propres idées, et diminution de la vigilance critique.

Impact attendu : les étudiants développent une conscience structurelle des limites de l'IA, transférable à tout outil futur.

Qui : enseignants + référent IA.

Base de travail disponible : voir fiche pédagogique sur les biais en Boîte à outils D.

#5. Développer l'esprit critique à l'ère de l'IA agentique

Cette recommandation concerne prioritairement les étudiants avancés et les formations pour lesquelles les usages agentiques présentent déjà une pertinence pédagogique ou professionnelle identifiable.

Quoi : pour les étudiants avancés (M1/A4+), développer la capacité à auditer non seulement le résultat d'un agent IA mais aussi le processus qu'il a suivi. Cela complète le Guide #8 (recommandation #5 sur la supervision) en insistant sur la dimension critique.

Exercice-type :

- L'enseignant fournit aux étudiants le résultat d'un processus agentique (rapport, analyse, synthèse) ainsi que le log des actions de l'agent.

- Les étudiants doivent : (1) identifier les sources consultées et évaluer leur fiabilité, (2) repérer les étapes où l'agent a pu dévier, (3) identifier les alternatives non explorées, (4) proposer un cadrage amélioré.

Impact attendu : préparation aux usages professionnels de l'IA agentique, où la capacité d'audit est une compétence critique.

Qui : enseignants + référent IA + DSI.

Base de travail disponible : voir exercice d'audit agentique en Boîte à outils E.

#6. Évaluer et piloter le développement de l'esprit critique à l'ère de l'IA

Quoi : mettre en place des indicateurs pour mesurer le développement de l'esprit critique des étudiants vis-à-vis de l'IA, et ajuster le dispositif en conséquence.

Indicateurs recommandés :

- Qualité de la détection d'erreurs dans les exercices d'audit critique (taux de faux positifs/négatifs).
- Systématicité de la vérification des sources dans les travaux rendus.
- Résultats aux activités de comparaison IA/sources primaires.
- Enquête de perception : « Faites-vous confiance aux réponses de l'IA sans vérification ? » (évolution année après année).

Impact attendu : pilotage fondé sur des données, amélioration continue du dispositif.

Qui : direction des études + référent IA.

Base de travail disponible : voir indicateurs de pilotage en Boîte à outils F.

III. Références utiles

Esprit critique : cadres fondateurs

Ennis, R. H. (2011). The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities. Définition de référence des dispositions et habiletés constitutives de l'esprit critique.

Facione, P. A. (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. American Philosophical Association. Rapport Delphi fondateur définissant par consensus d'experts les composantes de l'esprit critique.

Kahneman, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux. Distingue pensée rapide (système 1) et lente (système 2) ; cadre pour comprendre l'acceptation peu réfléchie des réponses de l'IA.

IA, cognition et esprit critique

Carnegie Mellon University & Microsoft Research (2025). [The Impact of Generative AI on Critical Thinking Skills](#). Plus la confiance dans l'IA est élevée, moins l'esprit critique est mobilisé.

Gerlich, M. (2025). [AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking](#). Societies, 15(1), 6. Corrélation négative entre usage fréquent de l'IA et esprit critique, médiée par le délestage cognitif.

Shaw, S. D. & Nave, G. (2026). Thinking - Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender. Working Paper, The Wharton School. Introduit le « désengagement cognitif » : les usagers suivent les réponses de l'IA, même erronées, près de 80 % du temps.

Thesen, T. & Park, H. (2025). [Students Lack the Expertise to Detect AI Hallucinations](#). Nature Machine Intelligence. Montre que les étudiants détectent mal les hallucinations de l'IA, surtout dans les domaines où ils sont novices.

Zhang, Y. et al. (2025). AI dependency and student outcomes: Critical thinking, self-confidence and problem-solving. Education and Information Technologies. Associe la dépendance à l'IA à une baisse de l'esprit critique, de la confiance en soi et de la résolution de problèmes.

Détection et fiabilité

Weber-Wulff, D. et al. (2023). [Testing of detection tools for AI-generated text](#). International Journal for Educational Integrity, 19(1), 26. Démontre le manque de fiabilité des outils de détection de texte généré par IA.

Cadres pédagogiques et institutionnels

Perkins, M. et al. (2024). [The AI Assessment Scale \(AIAS\) v2](#). CC BY-NC-SA. Échelle graduée des niveaux d'usage autorisé de l'IA dans une évaluation.

UNESCO (2024). [Orientations pour l'IA générative dans l'éducation et la recherche](#). Cadre mondial, approche centrée sur l'humain.

Vangrunderbeeck, P. (2024). [L'IA générative dans l'enseignement supérieur](#). UCLouvain. Panorama des stratégies d'adaptation des enseignements.

BOITE A OUTILS

A - Exercices de détection d'erreurs et hallucinations de l'IA

Ces exercices-types peuvent être adaptés à toute discipline.

Exercice	Consigne	Compétence développée	Durée
Chasse aux hallucinations	L'enseignant soumet 5 questions de sa discipline à l'IA et fournit les réponses aux étudiants. Ceux-ci doivent identifier au moins 3 erreurs factuelles.	Détection d'erreurs factuelles	~30 min
Références fantômes	L'étudiant demande à l'IA une bibliographie sur un sujet et vérifie l'existence de chaque référence / évalue la qualité de la référence selon certains critères (rang de la revue, nombre de citations, h index de l'auteur...).	Vérification des sources	~1h
Correction disciplinaire	L'étudiant reçoit une réponse IA à un problème de sa discipline et doit rédiger une version corrigée et améliorée, en justifiant chaque modification.	Esprit critique disciplinaire	~1h
Comparaison de modèles	Même question posée à 2 modèles différents. L'étudiant analyse les différences et juge quelle réponse est la plus fiable ou rédige une nouvelle réponse fusionnant les éléments cités par les 2 modèles.	Calibration de la confiance	~45 min

B - Checklist de vérification des sources pour les travaux utilisant l'IA

Cette checklist a vocation à être jointe à tout travail utilisant l'IA.

- ☐ J'ai vérifié l'existence et la qualité de chaque référence bibliographique citée par l'IA.
- ☐ J'ai vérifié les 3 faits les plus importants de la réponse IA auprès de sources primaires.
- ☐ J'ai examiné le raisonnement proposé par l'IA (hypothèses, enchaînement logique, validité des conclusions).
- ☐ J'ai identifié les limites et les angles morts de la réponse ou du raisonnement de l'IA.
- ☐ J'ai distingué à la fin de mon rendu ce qui provient de l'IA, ce qui a été vérifié et ce qui est mon apport personnel.
- ☐ J'ai cherché au moins une source contradictoire ou complémentaire à la réponse IA.

C - Idées d'activités d'audit critique de l'IA par discipline

Ces fiches proposent des exemples d'activités d'audit critique adaptées aux disciplines du RESDD.

Discipline	Idées d'activité d'audit critique
Génie civil / BTP	• Demander à l'IA de dimensionner un élément de structure. • Les étudiants vérifient avec les normes et identifient les erreurs.
Urbanisme	• Demander à l'IA une analyse d'un quartier. • Les étudiants confrontent avec leurs relevés terrain.
Météorologie	• Demander à l'IA d'interpréter une carte synoptique. • Les étudiants comparent avec l'analyse de référence.
Environnement	• Demander à l'IA un diagnostic environnemental d'un site. • Les étudiants vérifient avec les bases de données officielles.
Énergie	• Demander à l'IA de calculer un bilan énergétique. • Les étudiants vérifient les ordres de grandeur et les hypothèses.

D - Fiche pédagogique : biais systémiques de l'IA

Type de biais	Explication	Conséquence pour les étudiants	Exercice pratique
Biais d'hallucination	Production d'informations fausses mais très convaincantes sans signaler d'incertitude.	Acceptation d'informations erronées (faits, chiffres, références).	Identifier et vérifier des affirmations précises dans une réponse IA (chasse aux erreurs).
Biais de données	L'IA a été entraînée sur des données majoritairement anglophones et occidentales.	Surreprésentation de certaines langues, cultures et perspectives, mais aussi reproduction de biais sociaux présents dans les données (stéréotypes sexistes, racistes, culturels, socio-économiques).	Comparer la réponse IA en français et en anglais sur le même sujet, ou analyser la présence de stéréotypes dans une réponse.
Biais d'autorité	La formulation donne une impression de fiabilité, même si le contenu est faux.	Baisse de la vigilance critique face à des réponses bien rédigées.	Comparer deux réponses (une correcte, une fausse mais bien écrite) et justifier laquelle est fiable.
Biais de lissage	L'IA tend à produire des réponses consensuelles, socialement acceptables, standardisées, gommant controverses et nuances.	Disparition des débats, des positions minoritaires, des nuances. Réponses génériques peu adaptées.	Demander à l'IA les arguments « pour et contre » puis comparer avec des sources académiques.
Biais de simplification excessive	Simplification de phénomènes complexes avec omission de variables clés.	Compréhension partielle ou erronée des systèmes complexes.	Identifier les variables ou facteurs absents dans une réponse IA.
Biais de complétude apparente	Réponse structurée mais incomplète ou partielle.	Non-détection des lacunes importantes.	Lister ce qui manque pour rendre la réponse réellement complète.
Biais temporel	Données d'entraînement datées.	Réglementations obsolètes, données périmées, absences d'événements récents.	Vérifier la date d'entraînement du modèle et comparer par rapport aux nouveautés de l'actualité (attention à la possibilité de recherche sur internet)
Biais de confirmation ou d'acquiescement	L'IA suit l'hypothèse implicite de la question.	Validation d'hypothèses erronées.	Poser des questions non orientées, séparer les faits/opinions dans les questions, et donner moi les arguments pour/contre. Puis comparer avec une question sans ces éléments d'attention.
Biais de cadrage	La réponse dépend de la formulation de la question.	Vision partielle, angles morts.	Reformuler une même question de plusieurs façons et comparer les réponses. Tester les mêmes questions dans une autre conversation, posées dans un ordre différent.
Biais de complaisance	Ton validant et peu critique envers les idées de l'utilisateur.	Renforcement artificiel de la confiance, baisse de l'esprit critique.	Proposer une idée discutable et analyser le niveau de réponse critique de l'IA.

E - Exercice d'audit d'un processus agentique

Cet exercice est destiné aux étudiants avancés (M1+).

Scénario : un agent IA a été chargé de produire une note de synthèse sur les impacts environnementaux d'un projet d'aménagement. Le résultat (note de 5 pages) et le log des actions de l'agent (20 étapes) sont fournis aux étudiants.

Consignes pour les étudiants :

1. Analyser les sources consultées par l'agent : sont-elles fiables ? actualisées ? diversifiées ? Y a-t-il des sources clés manquantes ?
2. Identifier les étapes où l'agent a pu dévier : raccourcis, sélection biaisée, hallucinations intermédiaires.
3. Vérifier les conclusions de la note au regard des sources identifiées.
4. Proposer un cadrage amélioré : que feriez-vous différemment si vous deviez redéfinir la mission de l'agent ?
5. Rédiger un avis motivé de 1 page : cette note est-elle utilisable en l'état ? avec quelles réserves ?

F - Indicateurs de pilotage de l'esprit critique à l'ère de l'IA

Indicateur	Mode de recueil	Cible	Fréquence
Taux de détection des hallucinations dans les exercices d'audit	Résultats aux exercices d'audit critique	> 70 % des erreurs détectées	Semestriel
Taux de vérification des sources dans les travaux rendus	Analyse d'un échantillon de travaux	> 90 % des travaux incluent une vérification	Semestriel
Confiance calibrée dans l'IA	Enquête étudiants : « Vérifiez-vous les réponses de l'IA ? »	> 70 % « toujours ou souvent »	Annuel
Qualité des audits agentiques (M1+)	Grille d'évaluation spécifique	Progression mesurée	Annuel
Retours enseignants sur l'esprit critique observé	Enquête qualitative enseignants	Tendance positive	Annuel

Les cibles proposées ci-dessous ont une valeur indicative et doivent être adaptées à la taille de l'établissement, à son niveau de maturité et au degré de déploiement du dispositif