

Supports à l'évaluation certificative

Giulia GLENISSON, Technopédagogue en Faculté de Gembloux Agro-Bio Tech

Manon MASSIN, Technopédagogue en Faculté d'Architecture

Laurence MICHIELS, Technopédagogue en Faculté de Médecine Vétérinaire

Pierre-Yves THIELEN, Technopédagogue en Faculté de Psychologie, Logopédie et Sciences de l'Education



Plan de la séance

- ▶ Introduction
- ▶ Enjeux et pédagogie
- ▶ Présentation d'outils d'évaluation disponibles à l'Uliège
- ▶ Présentation d'outils d'évaluation mobilisant l'I.A.
- ▶ Moment de questions-réponses

Présentation des intervenants

► Pascal Detroz

- Professeur Ordinaire
- Responsable académique du Système Méthodologique d'Aide à la Réalisation de Tests (SMART)
<http://smart.uliege.be>
- Directeur de l'Unité de recherche Didactifen
<http://didactifen.uliege.be>
- Institut de Formation et de Recherche en Enseignement Supérieur
<http://lfres.uliege.be>

► Alicia Herteler

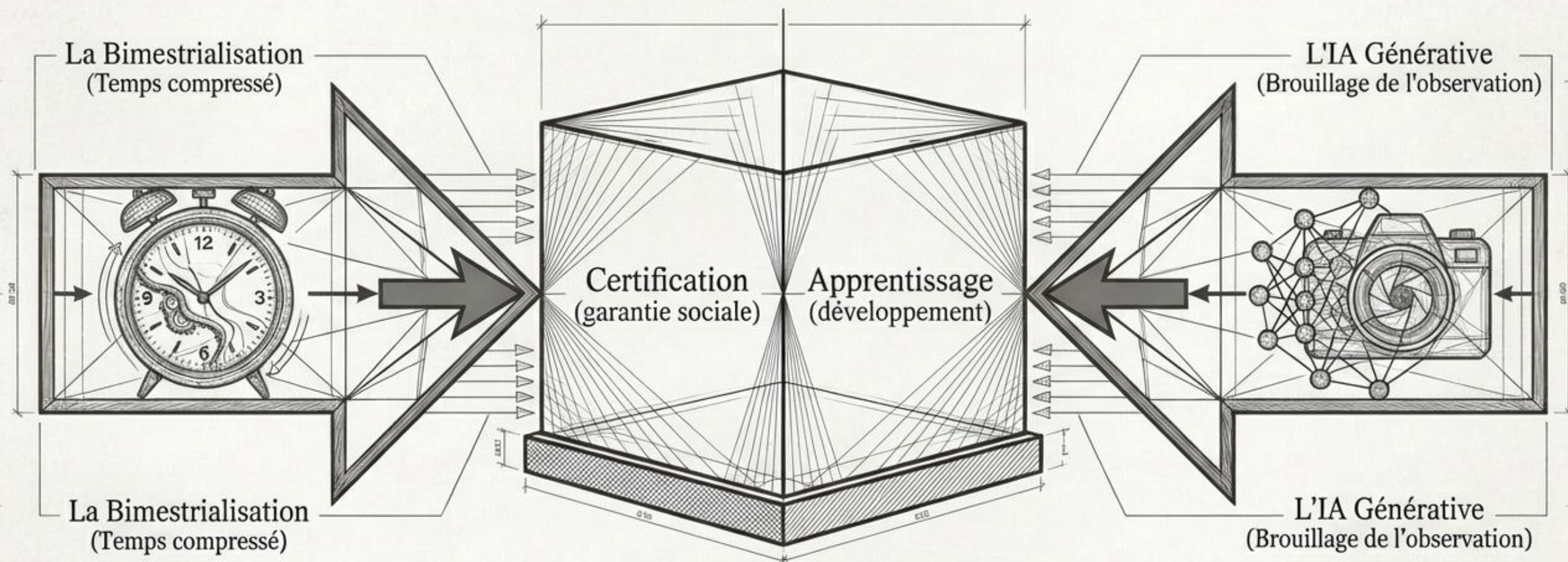
- Assistante Techno-pédagogue Faculté de Philosophie et Lettres
- Assistante Techno-pédagogue Faculté HEC Liège - Digital Expérience
- Accompagnement IA - CARE Outils Numériques Uliège
<https://www.digital.uliege.be>

REPENSER NOS ÉVALUATIONS (DÉCRET PARCOURS & BIMESTRIALISATION)

Vers une écologie de l'évaluation universitaire

Principes directeurs et analyse structurelle de l'évaluation continue.

L'impératif de réinvention face aux mutations de l'enseignement supérieur



L'université ne peut plus se reposer sur une évaluation terminale unique. L'accélération du rythme académique et le brouillage technologique de l'observation exigent un arbitrage systémique entre la robustesse de la décision certificative, le développement de compétences complexes et la soutenabilité logistique.

Volet Évaluation : Concevoir une architecture évaluative robuste et soutenable

Vers une certification crédible et des apprentissages de haut niveau



Analyser l'évaluation continue face aux contraintes structurelles

Concept central

Collecte d'informations distribuée tout au long du temps d'enseignement pour fonder la décision certificative, fonctionnant comme alternative à l'épreuve terminale unique.

[+] Atouts pédagogiques

Triangulation — Multiplication des occasions d'observation dans des contextes variés, fiabilisant le jugement sur le long terme en évitant la dépendance à une mesure ponctuelle.

Feedback & Régulation — Rapprochement temporel entre l'évaluation et l'apprentissage, permettant des retours utiles en cours d'acquisition (évaluation formative intégrée).

[!] Points de vigilance docimologiques

Soutenabilité (Qualité de la donnée) — La multiplication de petites tâches dans des contextes non contrôlés fragilise grandement la validité et la stabilité de la décision académique finale.

Intelligence Artificielle — Risque de biais de validation. Il devient complexe d'inférer la compétence réelle de l'étudiant à partir de productions répétées hors surveillance.

Charge cognitive & Stratégies — Saturation attentionnelle face à une pression diffuse (assessment-driven learning). Les étudiants surinvestissent le court terme (ce qui est noté immédiatement) au détriment de l'apprentissage profond.

Quatre principes fondateurs d'une architecture évaluative



Optimisation (Complexité / Qualité)

Résoudre la tension entre la robustesse de la mesure (indispensable à la certification) et la complexité des performances évaluées. L'exigence docimologique doit s'aligner sur la nature de la tâche.



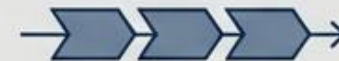
Proportionnalité (Enjeux / Exigences)

Adapter le niveau d'exigence méthodologique à l'enjeu de la décision. Un fort enjeu certificatif exige une procédure très robuste (calibration, recorection) ; un faible enjeu permet l'agilité et le formatif.



Variation structurée (Complémentarité)

Aucune modalité n'est parfaite isolément. Combiner des dispositifs aux propriétés complémentaires permet d'observer différentes dimensions des apprentissages sans créer de redondance.



Approche programme (Temporalité & Répartition)

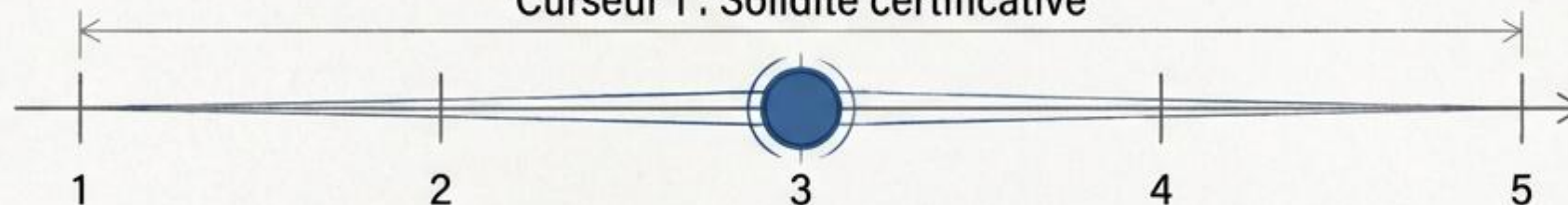
Dépasser l'échelle du cours isolé. Répartir les fonctions (certificatives, formatives, complexes) dans le temps et entre les unités d'enseignement pour assurer la soutenabilité globale.

**Système
intégré**

Les continuums de la Mesure : Fiabilité et Sécurité

Curseur 1 : Solidité certificative

Décision fragile
(forte variabilité
docimologique,
interprétation ouverte)

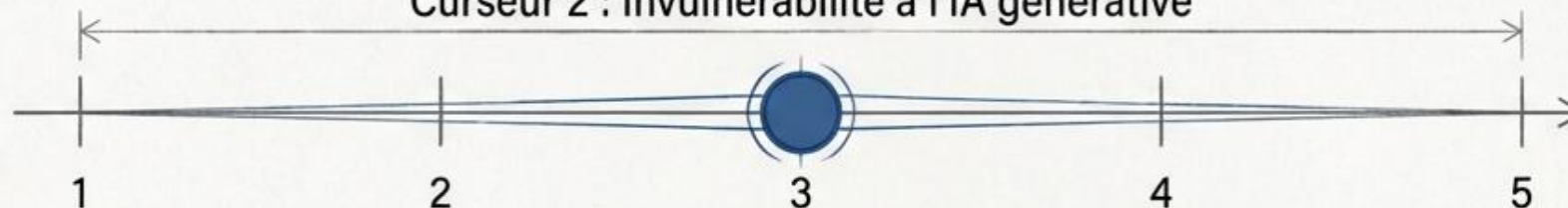


Décision robuste
(validité et fidélité
psychométrique élevées,
comparabilité stricte)

La capacité d'un dispositif à produire des décisions académiques incontestables et à garantir la comparabilité des performances (respect du contrat social de l'institution).

Curseur 2 : Invulnérabilité à l'IA générative

Vulnérable (délégation
technologique facile et
intraçable)



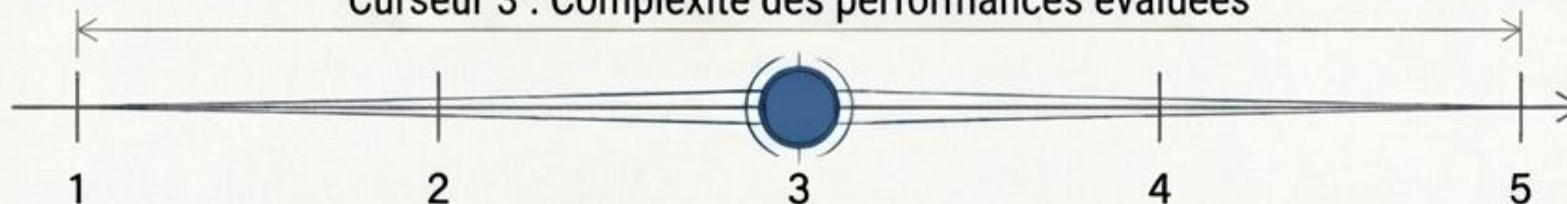
Sécurisé (performance
directement observable,
interaction en temps réel
ou environnement clos)

La capacité à s'assurer que la performance observée est bien le fruit exclusif du travail intellectuel de l'étudiant.

Les continuums de l'Apprentissage : Ambition et Développement

Curseur 3 : Complexité des performances évaluées

Restitution simple
(connaissances
déclaratives, application
de procédures)

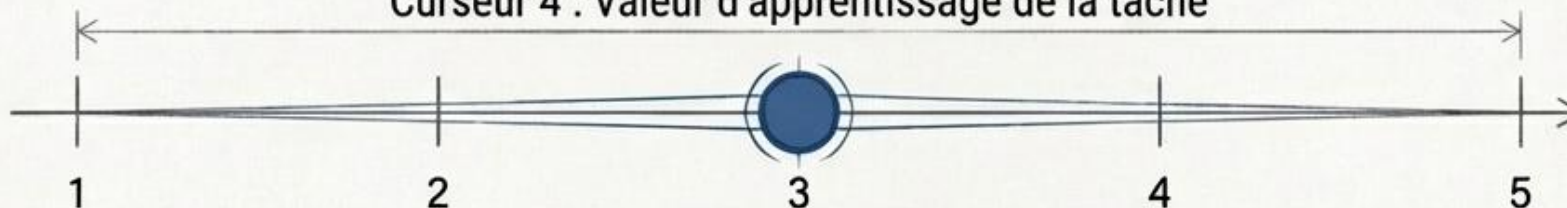


Intégration complexe
(résolution de problèmes
inédits, pensée critique,
mobilisation transversale)

Le niveau d'ambition intellectuelle de la tâche. Les compétences universitaires de haut niveau exigent des dispositifs ouverts, structurellement plus difficiles à fiabiliser.

Curseur 4 : Valeur d'apprentissage de la tâche

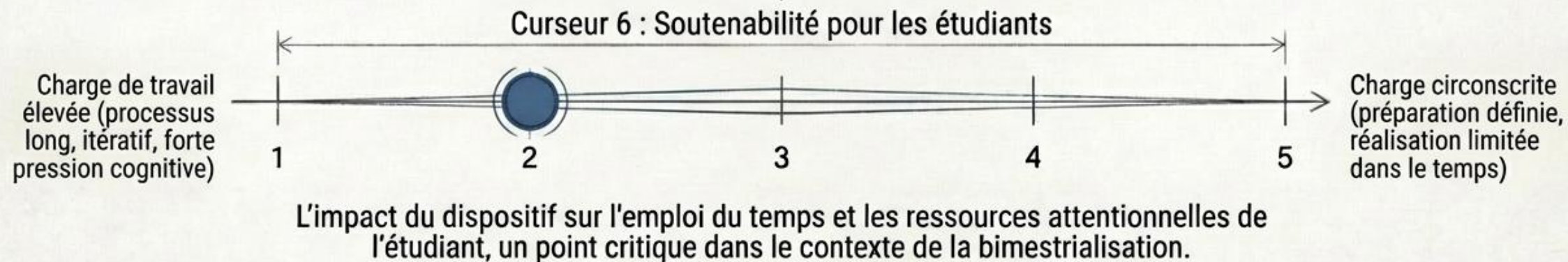
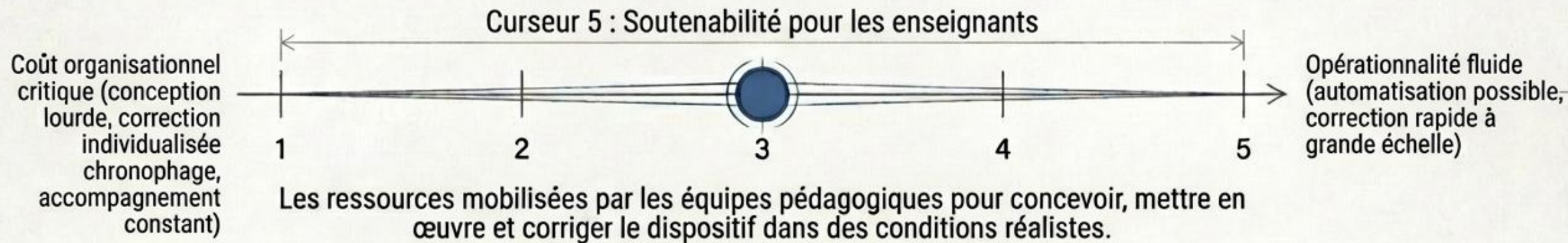
Constat ponctuel (la
tâche vérifie un état, sans
transformer l'étudiant)



Expérience de
développement
(la tâche est elle-même
une occasion structurante
d'apprentissage)

Le degré auquel la réalisation de l'évaluation contribue directement à la progression et à la réflexivité de l'étudiant.

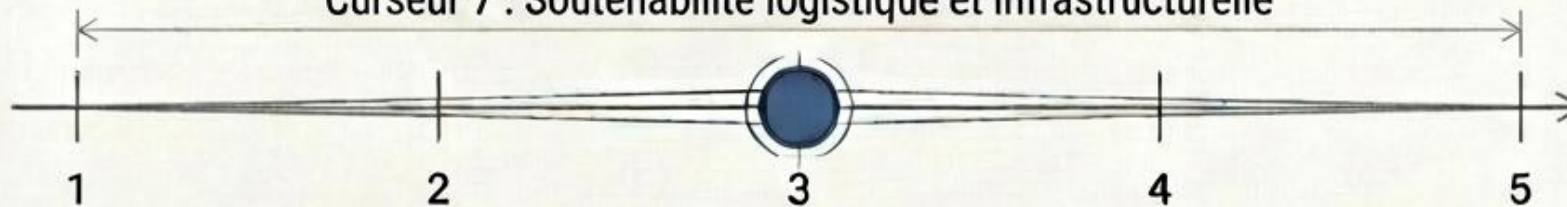
Les continuums de la Soutenabilité Humaine : La charge mentale



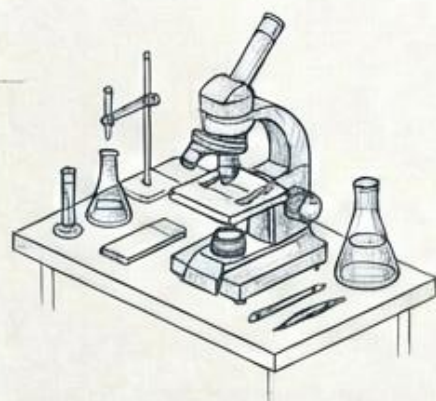
Le continuum de la Réalité Matérielle : Infrastructures et Logistique

Curseur 7 : Soutenabilité logistique et infrastructurelle

Exigences lourdes
(laboratoires spécifiques,
plateformes complexes,
de partenariats externes)

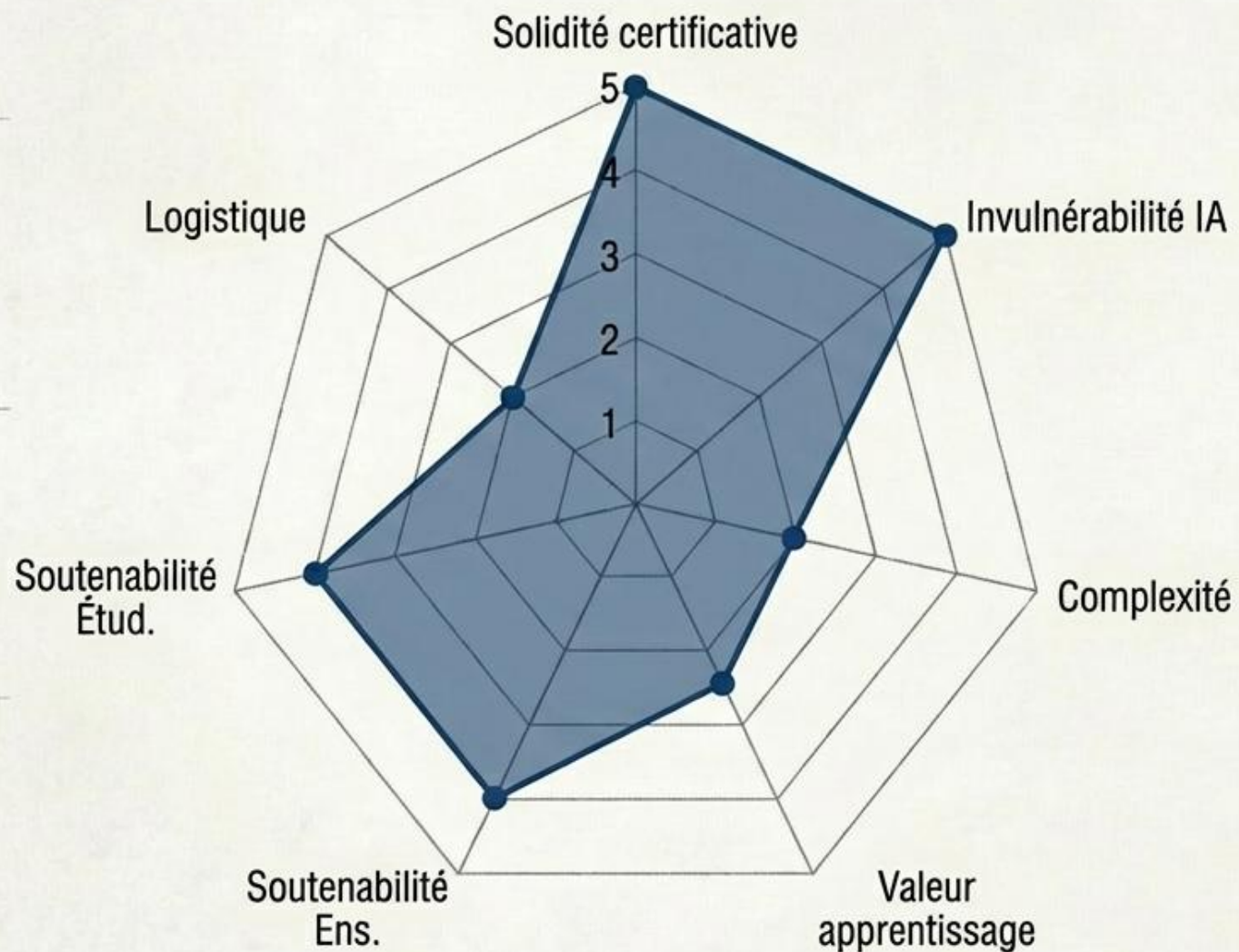


Déploiement standard
(infrastructures communes,
peu de ressources
matérielles atypiques)



La capacité de l'institution à déployer et maintenir le dispositif à grande échelle. La contraction du temps liée à la bimestrialisation transforme les contraintes classiques (disponibilité des auditoires, gestion des flux, surveillance) en véritables goulets d'étranglement organisationnels.

Examens écrits standardisés (QCM, exercices encadrés)



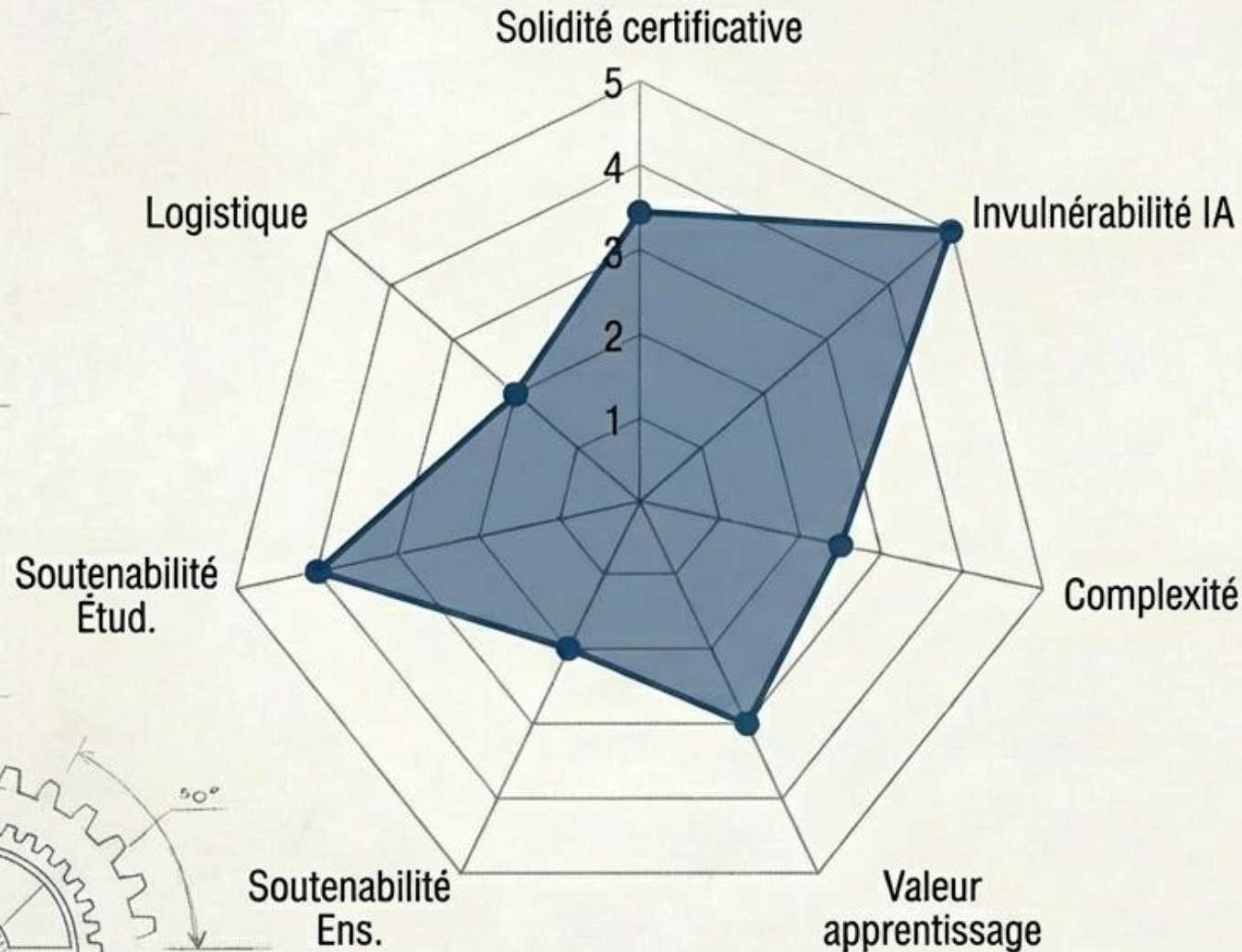
Profil analytique :
Le standard institutionnel de comparabilité.

Forces :
Haute fiabilité psychométrique, stabilisation des jugements, excellente soutenabilité enseignante pour les grandes cohortes. Totalement sécurisé face à l'IA en salle surveillée.

Limites :
Restreint à des savoirs circonscrits. Très faible valeur d'apprentissage (logique de constat pur).

Logistique :
Exige la mobilisation massive d'espaces et de surveillants, complexe en période de bimestrialisation resserrée.

Examens écrits ouverts (Dissertations, questions de synthèse)



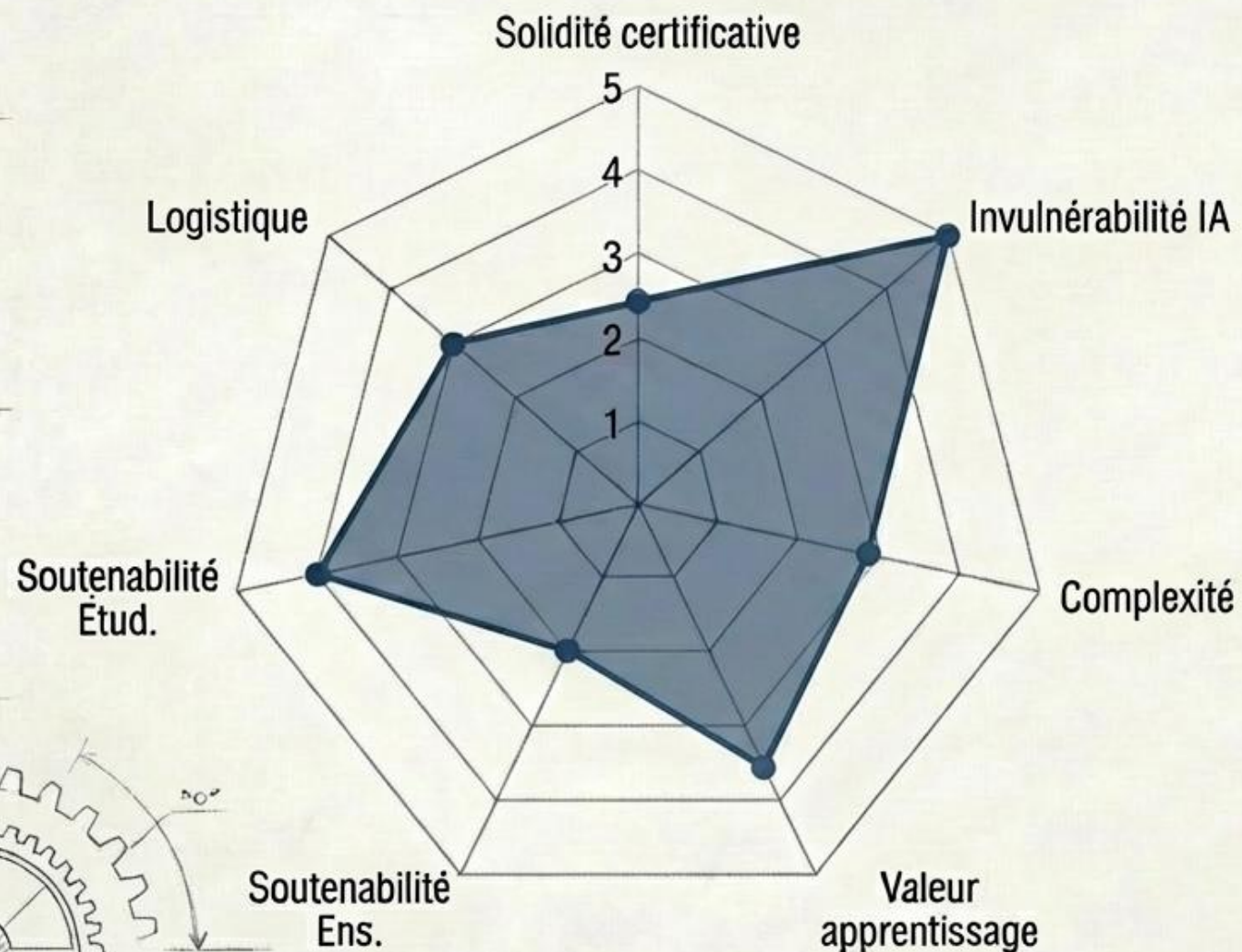
Profil analytique :
Le compromis classique de l'argumentation.

Forces :
Permet d'évaluer la structuration de la pensée et la mobilisation de connaissances. Protection garantie contre l'IA si organisé en auditoire.

Limites :
Vulnérabilité docimologique accrue (jugements interprétatifs). La correction devient coûteuse pour l'enseignant avec de gros effectifs.

Observation :
Si réalisé à distance (Take-home exam), l'invulnérabilité à l'IA chute drastiquement, nécessitant une réingénierie de la tâche.

Examens oraux (Interrogations, soutenances)

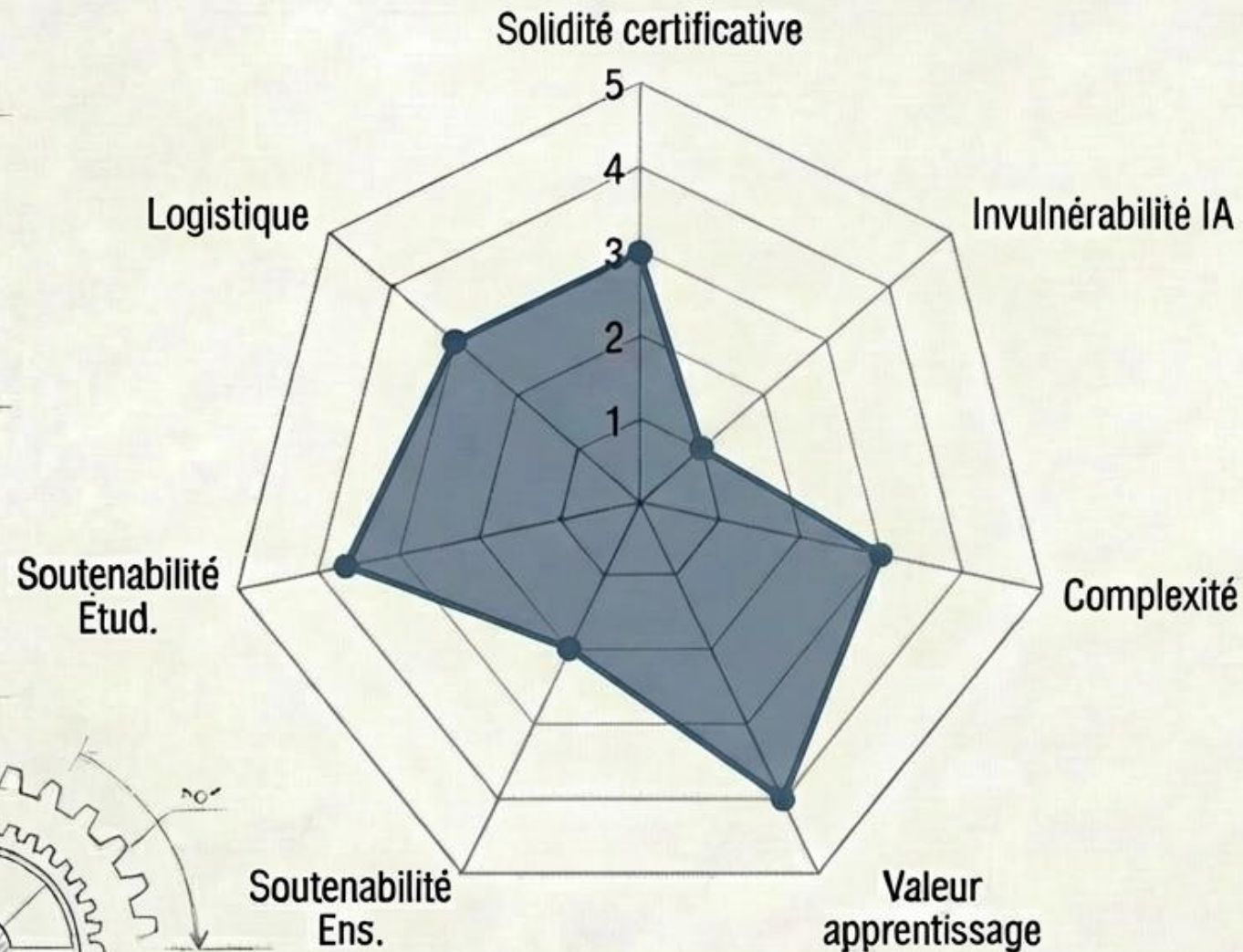


Profil analytique :
La haute interaction contextualisée.

Forces :
Permet de sonder en temps réel le raisonnement et l'adaptation. Bouclier absolu contre la fraude technologique générative.

Limites :
Risque docimologique élevé (interaction variable, biais de contexte).
Cauchemar de soutenabilité pour les enseignants face à de grandes cohortes (investissement individuel extrême).

Travaux et projets longs individuels (Dossiers, mémoires)

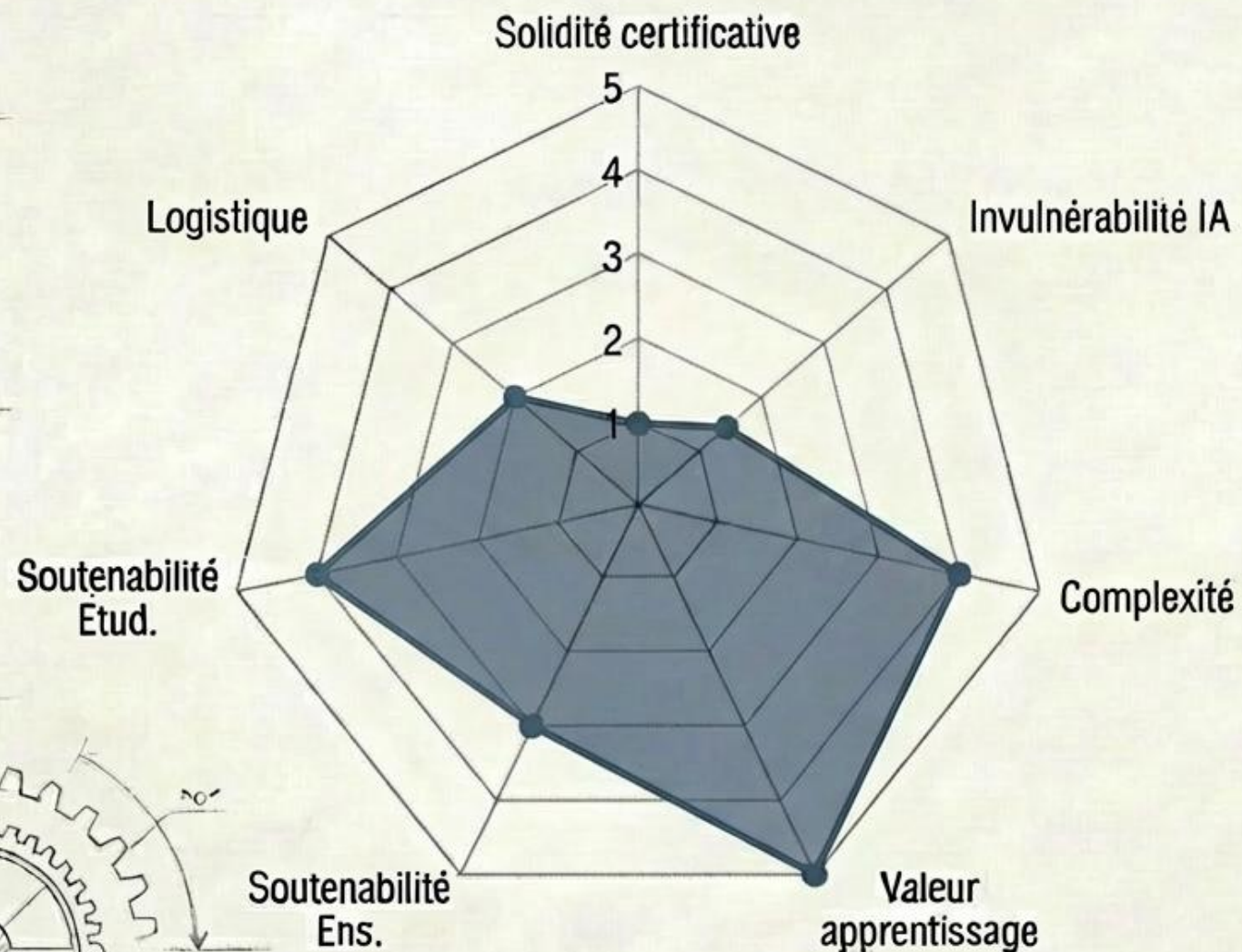


Profil analytique :
La construction asynchrone approfondie.

Forces :
Haute valeur formative. Observation de compétences intégratives (recherche, structuration, synthèse) difficiles à tester en auditoire.

Limites :
Forte vulnérabilité à l'IA générative dans la phase de rédaction asynchrone.
Chronophage pour l'étudiant (charge diffuse) et pour la correction enseignante.

Travaux et projets collectifs



Profil analytique :
L'expérience de la complexité collaborative.

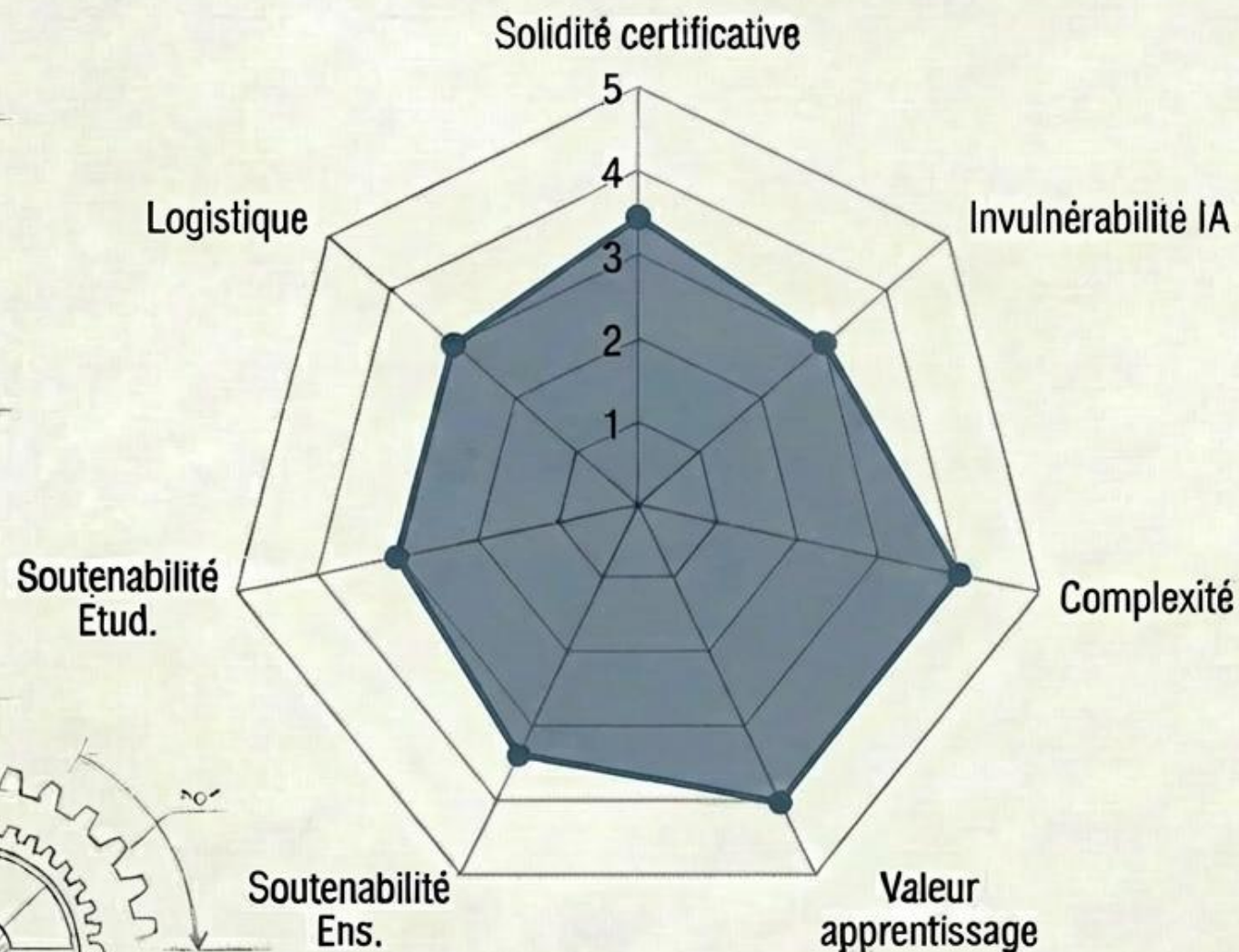
Forces :

Summum de l'apprentissage collaboratif et de l'organisation. Moins de copies à corriger pour l'enseignant.

Limites :

Incapacité à isoler la performance individuelle (faible validité certificative).
Risque de surcharge organisationnelle pour l'étudiant.
Déficit d'infrastructures adaptées au travail de groupe.

Études de cas et résolutions de problèmes complexes

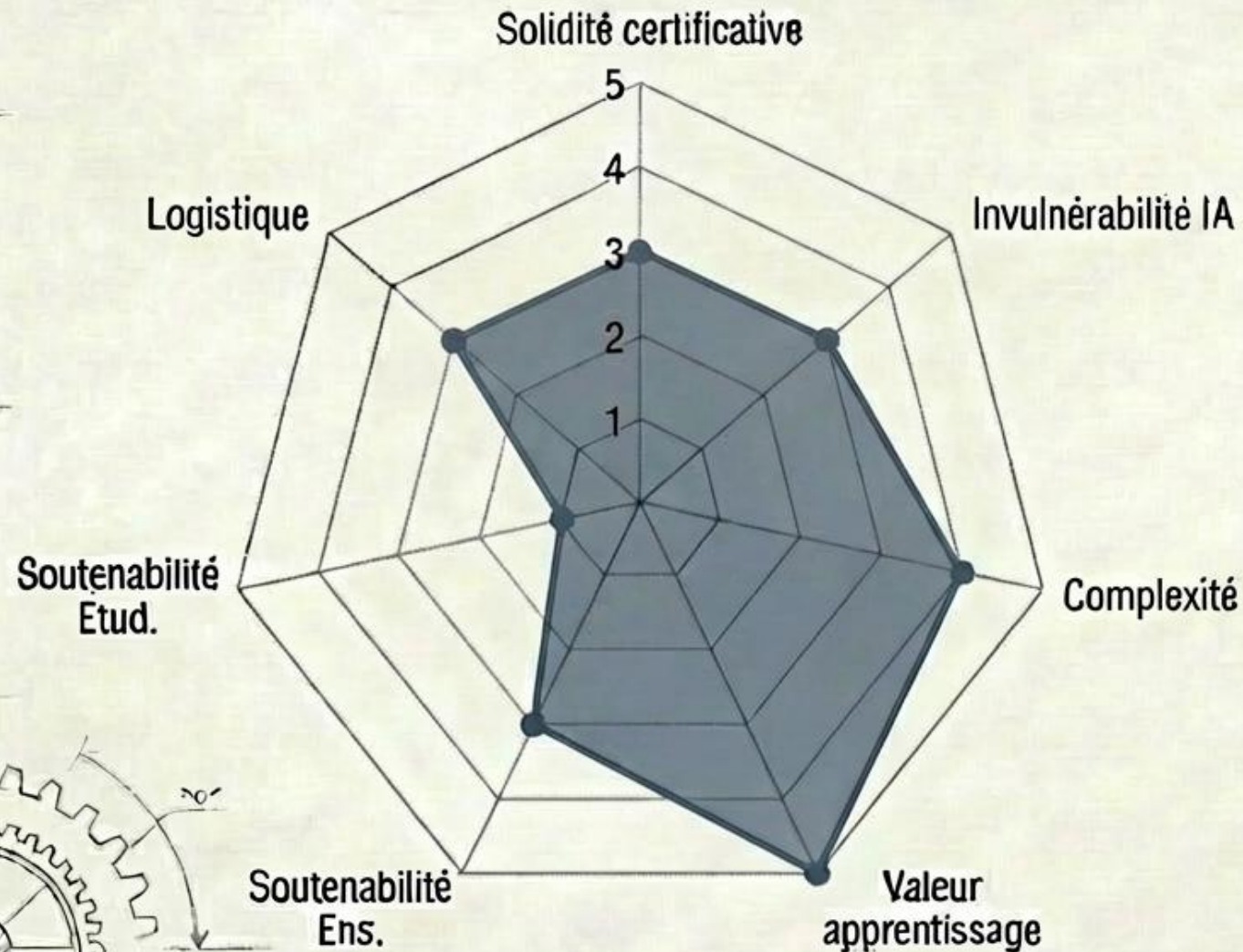


Profil analytique :
L'instrument d'intégration par excellence.

Forces :
Profil hautement équilibré. Permet d'observer l'application concrète des savoirs dans des situations réalistes. Bon compromis sur la soutenabilité enseignante et étudiante.

Limites :
L'invulnérabilité à l'IA dépend fortement de l'environnement. Exige une conception initiale rigoureuse de la part des équipes pédagogiques.

Portfolios (d'apprentissage, réflexifs, de compétences)

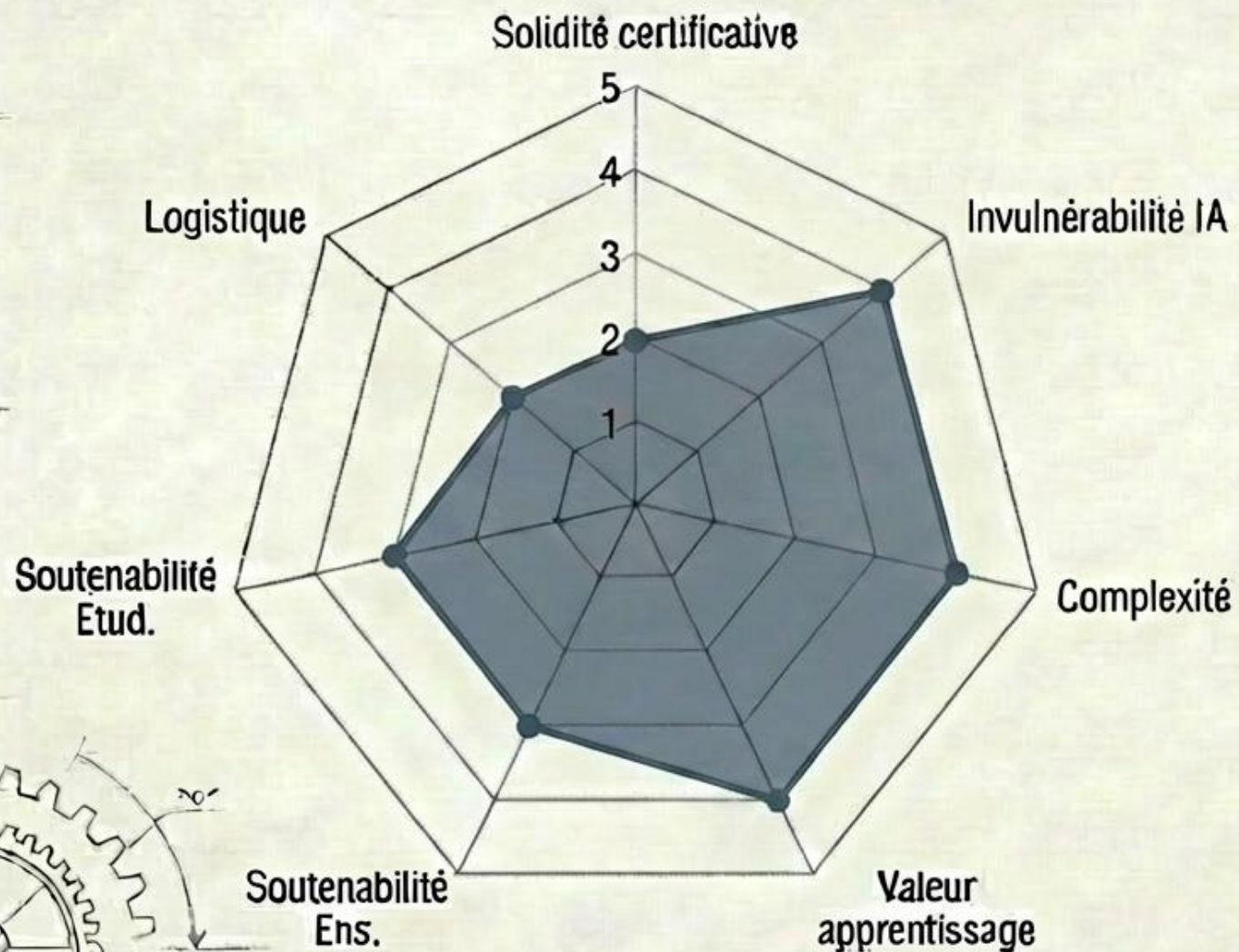


Profil analytique :
La documentation continue du développement.

Forces :
Excellence absolue pour l'analyse réflexive et le suivi longitudinal des compétences. Transforme l'évaluation en véritable processus d'apprentissage.

Limites :
Enorme investissement temps/énergie pour l'étudiant (collecte, sélection, analyse) et pour l'enseignant (lecture interprétative complexe). Repose sur des plateformes numériques.

Évaluations pratiques et simulations

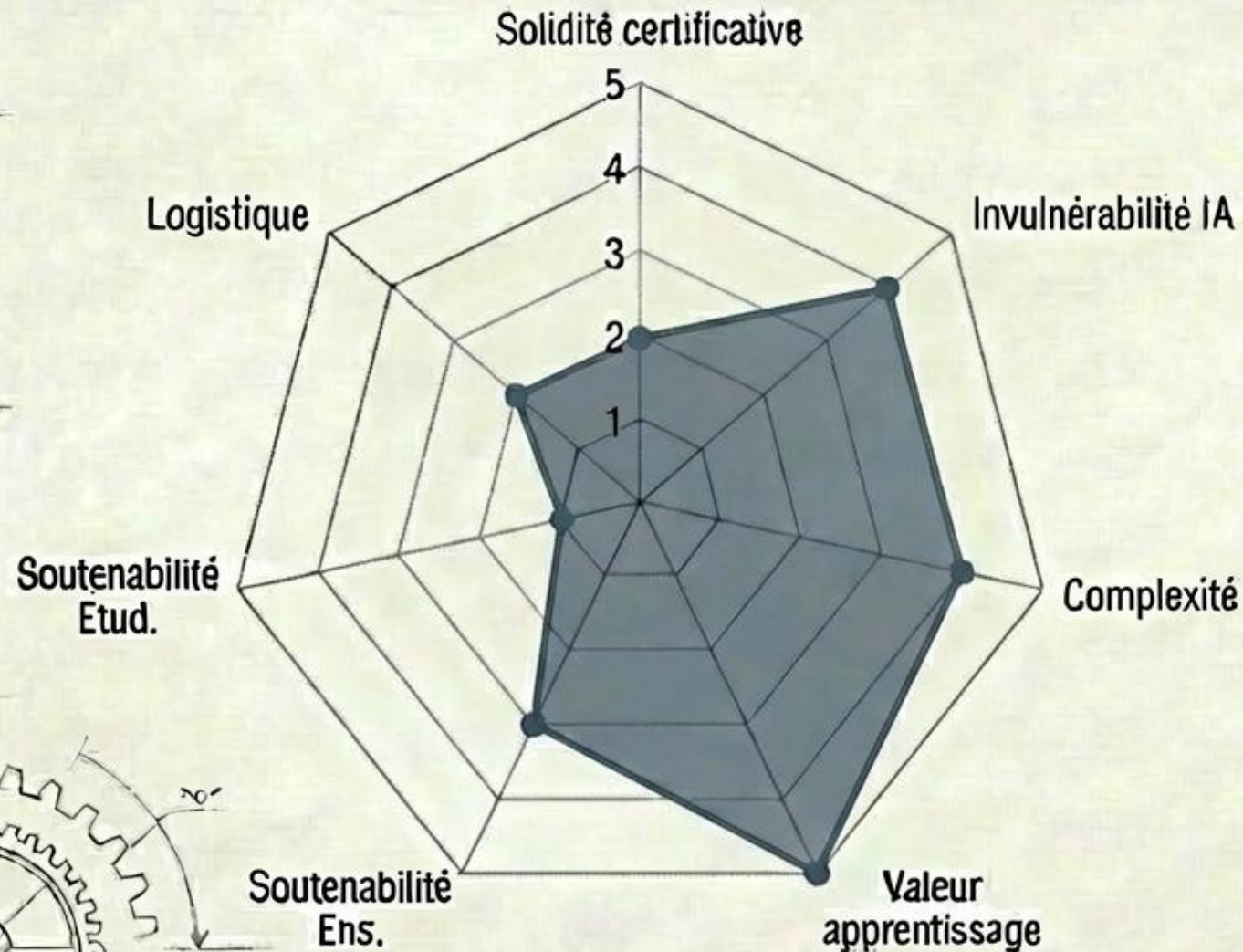


Profil analytique :
L'observation directe de l'action.

Forces :
Indispensable pour les compétences techniques, relationnelles et décisionnelles en temps réel. Quasiment imperméable à l'IA générative.

Limites :
Coût logistique et infrastructurel très lourd (matériel spécifique, locaux adaptés). Mobilisation enseignante intense pendant toute la durée des simulations.

Évaluations en situation authentique (Stages / Service learning)

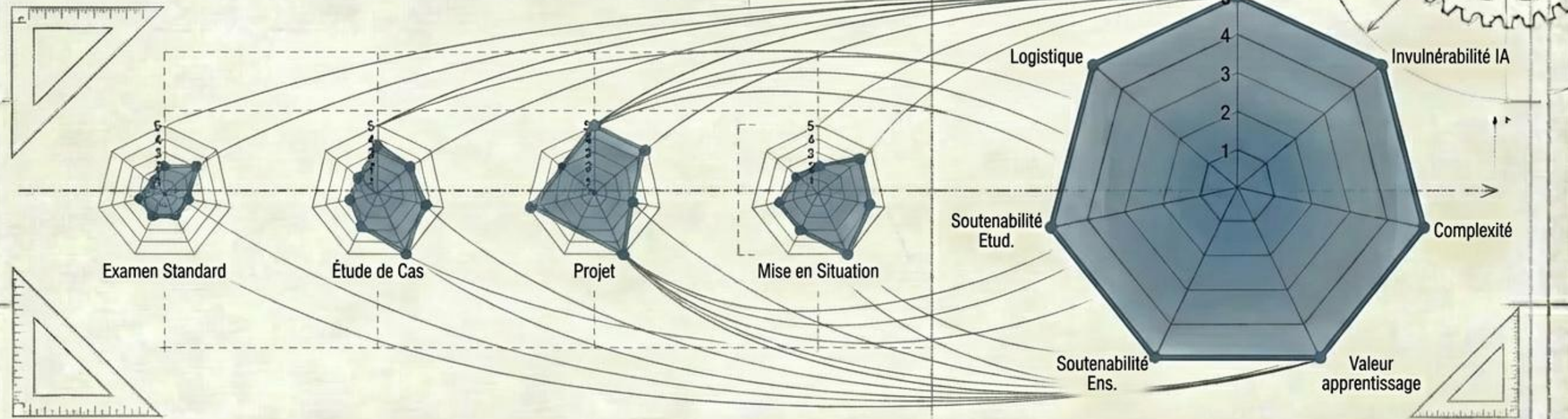


Profil analytique :
L'immersion dans l'imprévisibilité du réel.

Forces :
L'expérience d'apprentissage la plus riche, forçant l'adaptation à des variables non contrôlées. Protégé de l'IA par sa nature même.

Limites :
Très faible comparabilité certificative (chaque terrain de stage est unique). Logistique très complexe. Charge étudiante maximale (immersion totale).

L'Écologie d'Évaluation : L'Architecture Programmatique



Aucune modalité n'est parfaite isolément. Le défi est de concevoir une architecture à l'échelle du programme :

1. Triangulation des observations

Combiner des mesures standardisées (robustes) avec des productions complexes (formatrices).

2. Proportionnalité des enjeux

Aligner l'exigence docimologique sur le poids certificatif de la décision.

3. Régulation de la charge

Espacer intelligemment les évaluations pour garantir la soutenabilité humaine dans un cycle bimestriel.

L'évaluation ne se pense plus au niveau du cours, mais comme un système coordonné garantissant la crédibilité du diplôme.

L'évaluation à l'ère de l'IA

Alicia Herteler



L'évaluation à l'ère de l'IA

- ▶ Deux problématiques
 - ▶ Les travaux écrits à domicile
 - ▶ L'étude à l'aide de l'IA
- ▶ Evaluer avec l'IA
 - ▶ IA Act



LE PIÈGE DE LA LOGIQUE DE RÉSULTAT

- résultats mesurables (examens, rendus, mémoires), ce qui crée une pression poussant les étudiant·e·s vers l'usage détourné de l'IA. Deux effets connexes ont été identifiés :
- L'IA produit des travaux de meilleure apparence formelle, mais cela ne reflète pas nécessairement un meilleur apprentissage.
- Il faut revaloriser l'imperfection comme signe du processus d'apprentissage : un travail d'étudiant·e ne doit pas avoir la qualité d'un article publiable.

Il faut former à penser, et non à faire du propre. Il faut redonner confiance aux étudiant·e·s dans la lenteur du processus d'apprentissage.

RECONNECTER L'ÉVALUATION AUX PROCESSUS

Les textes rendus ne sont pas nécessairement le reflet des compétences réelles des étudiant·e·s — ce phénomène n'est pas nouveau, mais l'IA le massifie. Il faut donc recentrer l'évaluation sur les processus plutôt que sur les produits finis.

Plusieurs pistes ont été identifiées :

- Écrits sur table (sans IA) : valide des compétences de base, mais limité pour les travaux complexes nécessitant durée et accès aux sources.
- Soutenance orale de travaux écrits : l'écrit devient formatif, l'oral est sommatif — la performance orale valide les acquis du travail à domicile.
- Défense écrite sur table : rapport remis à domicile, suivi d'une expertise critique rédigée sur table.
- Pédagogie par projet : implique les étudiant·e·s dans une démarche créative avec une dimension expérientielle difficile à déléguer à l'IA.
- Contrôle différé sur un échantillon (ex. : 10 % de la cohorte) : vise non à sanctionner la fraude mais à vérifier l'acquisition réelle des compétences via un bref oral.

Source:

 Oregon State University
Ecampus 

Bloom's Taxonomy Revisited

RECOMMANDATION

CAPACITÉS DE L'IA

COMPÉTENCES HUMAINES
DISTINCTIVES

CRÉER	À réviser	Proposer une série d'alternatives, énumérer les inconvénients et les avantages potentiels, décrire des cas concrets réussis	Formuler des solutions créatives faisant appel au jugement humain, collaborer spontanément
ÉVALUER	À réviser	Identifier les avantages et les inconvénients de différentes lignes de conduite, élaborer des grilles d'évaluation	S'engager dans une réflexion métacognitive, évaluer de manière holistique les conséquences éthiques des actions alternatives
ANALYSER	À modifier	Comparer et différencier les données, déduire des tendances et des thèmes, calculer, prédire	Penser et raisonner de manière critique dans les domaines cognitif et affectif, interpréter des problèmes, des décisions et des choix authentiques et s'y référer

APPLIQUER	À réviser	Utiliser un processus, modèle ou méthode pour illustrer comment résoudre une question quantitative	Opérer, mettre en œuvre, conduire, exécuter, expérimenter et tester en situation réelle; appliquer la créativité et l'imagination au développement d'idées et de solutions
COMPRENDRE	À réviser	Décrire un concept avec des mots différents, reconnaître un exemple apparenté, traduire	Contextualiser les réponses en tenant compte des considérations émotionnelles, morales ou éthiques
MÉMORISER	À modifier	Se rappeler d'informations factuelles, énumérer des réponses possibles, définir un terme, construire une chronologie de base	Se rappeler d'informations dans des situations où la technologie n'est pas accessible

- Utilisez ce tableau comme référence pour évaluer et apporter des modifications aux activités et évaluations de cours alignées sur les objectifs pédagogiques en tenant compte des capacités des outils de l'IA générative et des compétences humaines distinctives.
- Toutes les activités et évaluations de cours bénéficieront d'une révision au regard des possibles outils alimentés par l'IA. Notez que les activités et évaluations de type MÉMORISER et ANALYSER sont plus susceptibles de nécessiter des modifications.

Cadrer l'usage

- ▶ Charte Uliège
https://www.student.uliege.be/cms/c_19230399/fr/faq-student-charte-uliege-d-utilisation-des-intelligences-artificielles-generatives-dans-les-travaux-universitaires
- ▶ Usage réfléchi et responsable de l'IA
 - ▶ Usages permis
 - ▶ Assistant linguistique
 - ▶ Assistant recherche d'information
 - ▶ Usages proscrits
 - ▶ Présenter une production d'une IA comme sienne
- ▶ Usage critique de l'IA
 - ▶ Responsable de la qualité des informations, des idées et des analyses fournies dans les travaux
- ▶ Usage transparent de l'IA
 - ▶ Référencer l'usage
 - ▶ Détailler dans la section méthode
- ▶ Usage de surpassement de l'IA
 - ▶ [...]tout ce qui fait un travail personnel de qualité reste indiscutablement humain et seul porteur d'apprentissages durables.

L'ÉCHELLE DE NIVEAUX D'USAGE DE L'IA

Transparence

Une échelle à cinq niveaux, à croiser avec les compétences attendues dans chaque cours ou travail.

- ▶ Niveau Description
- ▶ Niveau 0 Interdiction totale
- ▶ Niveau 1 Correction orthographique et reformulation légère — avec vigilance sur la perte de voix de l'étudiant·e
- ▶ Niveau 2 (recommandé) L'étudiant·e fournit une idée, l'IA questionne et challenge (IA comme Socrate) — l'apport intellectuel reste entièrement celui ou celle de l'étudiant·e
- ▶ Niveau 3 Co-écriture : l'IA participe activement à la rédaction
- ▶ Niveau 4 Production déléguée (ex. : code) — l'étudiant·e doit analyser le résultat et ne jamais le reprendre tel quel

Cette échelle est conçue pour être appliquée différemment selon les compétences visées : une même matière peut autoriser le niveau 4 pour la mise en forme et le niveau 0 pour l'argumentation.

L'utilisation de l'IA en 5 niveaux

Niveau 0 : Aucune utilisation

Attestation de non-recours – déclaration sur l'honneur

Niveau 1 : Assistance minimale

Correction linguistique

Traduction

Normes APA

Niveau 2 : Assistance encadrée

Exploration conceptuelle
-
Brainstorming

Structuration
-
Planification

Assistance bibliographique

Gestion de projet

Niveau 3 : Collaboration significative

Production de texte

Analyse

Visualisation

Support méthodologique

Feedback critique

Niveau 4 : Usage étendue

Programmation et code

ÉVALUER LE RAISONNEMENT ET LA PROGRESSION

Quatre déplacements évaluatifs :

- De la restitution de contenu vers l'analyse critique et créative : questions de recherche originales, propositions d'expérience, revue critique par les pair·e·s.
- Du produit vers le processus de travail : construction par étapes, outils de structuration du raisonnement (cartes conceptuelles, plans structurés), auto-évaluation, grilles critériées.
- Du produit final vers la capacité à progresser en réponse à la critique : plusieurs versions, intégration de feedbacks, révision itérative.
- De l'écrit vers l'interaction directe : présentations orales, questions en classe, réponses spontanées aux évaluateur·rice·s.

LE DÉFI DES GRANDES COHORTES

Ces approches processuelles sont facilement applicables en petits groupes ou en master, mais difficiles à mettre en œuvre pour des cohortes de plusieurs centaines d'étudiant·e·s.

Des adaptations ont été évoquées : QCM intelligents, épreuves à réponses courtes, travaux de groupe, ou contrôle différé sur échantillon.

L'étude à l'ère de l'IA

- ▶ Les étudiants étudient des erreurs, inexactitudes, approximation
- ▶ Sensibiliser --> littératie IA --> MinIAtures
- ▶ Tester des RAG
 - ▶ NotebookLM
- ▶ Etudiants moniteurs – méthodologie et contrôle qualité

Evaluer avec l'IA – IA ACT

L'objectif : une IA digne de confiance en Europe

- ▶ Passage 56
 - ▶ promouvoir une éducation et une formation numériques de haute qualité et permettre à tous les apprenants et enseignants d'acquérir et de partager les compétences et compétences numériques
- ▶ Chapter 1, Article 4
 - ▶ prendre des mesures pour garantir, dans la meilleure mesure possible, un niveau suffisant de littératie en IA de leur personnel et des autres personnes
- ▶ Systèmes d'IA à haut risque : « Annexe III, passage 3
 - ▶ les systèmes d'IA destinés à être utilisés pour évaluer les résultats d'apprentissage, y compris lorsque ces résultats sont utilisés pour orienter le processus d'apprentissage
 - ▶ des systèmes d'IA destinés à être utilisés dans le but d'évaluer le niveau d'éducation approprié auquel un individu recevra ou pourra accéder
 - ▶ des systèmes d'IA destinés à être utilisés pour surveiller et détecter les comportements interdits des élèves lors des examens

L'IA peut aider mais ne doit jamais avoir le dernier mot

Exemple outils et pratiques

- ▶ Gradescope
- ▶ Exemple d'Hélène Soyeur
- ▶ Grille d'évaluation très détaillée à donner à l'IA
 - ▶ Exemple paliers CECRL
- ▶ Création d'un Agent dans Copilot
- ▶ Test RAG correcteur
 - ▶ https://m365.cloud.microsoft/chat/?titleId=T_fa6b5c05-8efb-9956-c786-41e449700c6a&source=embedded-builder

Contactez-nous



Chloé BERARDIS

Faculté de Droit, Sciences politiques et Criminologie

cberardis@uliege.be



Olivier GASON

Faculté des Sciences Sociales

olivier.gason@uliege.be



Maxime HARMEGNIES

Faculté de Médecine vétérinaire

mharmegnies@uliege.be



Alicia HERTELER

HEC – École de gestion
Faculté de Philosophie et Lettres

a.herteler@uliege.be



Alessia HUBY

Faculté des Sciences

alessia.huby@uliege.be



Manon MASSIN

Faculté d'Architecture

mmassin@uliege.be



Laurence MICHIELS

Faculté de Médecine Vétérinaire

lmichiels@uliege.be



Thérèse REGGERS

Faculté de Médecine

Therese.Reggers@uliege.be



Pierre-Yves THIELEN

Faculté de Psychologie, Logopédie et Sciences de l'éducation

py.thielen@uliege.be



Giulia GLENISSON

Faculté de Gembloux Agro-Bio Tech

giulia.glenisson@uliege.be



Merci de votre participation !

Avant de partir... votre avis sur ce MTP6 nous intéresse !



OU

Cliquez sur ce lien :

[Feedback MTP6- Supports à l'évaluation certificative – Remplir le formulaire](#)