Les activités suivent la progression : Audit > Ingénierie > Confrontation.

Activité 1 : Le "Bullshit Bingo" Sémantique (Audit Hostile)

Objectif cognitif : Passer de la lecture passive à l'analyse critique impitoyable. Identifier la vacuité des contenus générés par défaut ("prompting naïf").

Public : Bachelor / Salariés (Départements Com/RSE/Marketing).

Dispositif :

1. **Génération :** Les participants génèrent un texte via un prompt basique (ex: "Écris une déclaration RSE pour une entreprise de logistique").
2. **Analyse :** Distribution d'une grille de "mots vides" (innovant, durable, au cœur de, synergie, engagement fort, ADN, vert).
3. **Score de vacuité :** Chaque occurrence d'un mot vide rapporte 1 point. Chaque phrase qui ne contient aucune donnée chiffrée ou action vérifiable rapporte 2 points.
4. **Victoire inversée :** Le texte ayant le score le plus élevé (le plus vide) est analysé collectivement.

Questionnement induit : Pourquoi l'IA privilégie-t-elle la forme sur le fond ? Comment la fluidité syntaxique masque-t-elle l'absence d'information ?

Activité 2 : Le Piège à Hallucinations (Vérification Factuelle)

Objectif cognitif : Comprendre la nature probabiliste et non factuelle des LLM. Forcer la vérification systématique.

Public : Bachelor (Recherche) / Salariés (Technique/Juridique).

Dispositif :

1. **Injection :** L'animateur fournit un jeu de données interne *fictif mais réaliste* (ex: un rapport carbone avec des incohérences subtiles) et demande de générer une synthèse valorisante.
2. **Le Piège :** Les participants doivent identifier où l'IA a :
    - Lissé une incohérence (biais de complaisance).
    - Inventé une certification inexistante pour "faire vrai".
    - Transformé une hypothèse en affirmation.
3. **Correction :** Réécriture du prompt pour interdire l'extrapolation (Technique : *Grounding* / Ancrage contextuel).

Questionnement induit : À quel moment l'outil cesse-t-il d'assister pour commencer à désinformer ? Quelle est la responsabilité juridique de l'utilisateur face à une affirmation générée ?

Activité 3 : Le Tribunal de la RSE (Stress Test Oral)

Objectif cognitif : Responsabilisation totale. Briser l'excuse "c'est l'IA qui l'a dit".

Public : Bachelor (Fin de cycle) / Managers.

Dispositif :

1. **Rôle A (L'Avocat) :** Utilise l'IA pour préparer des éléments de langage défendant une action controversée de l'entreprise.
2. **Rôle B (L'ONG Hostile) :** N'utilise pas l'IA. Dispose des faits bruts. Attaque chaque formulation vague.
3. **Confrontation :** Simulation de conférence de presse (5 minutes). L'Avocat doit défendre le texte. Si l'Avocat est pris en défaut de précision, il perd.
4. **Règle d'Or :** Interdiction totale de dire "C'est une erreur de l'outil". L'étudiant endosse 100% du texte produit.

Questionnement induit : Un texte généré est-il défendable à l'oral ? La vitesse de production compense-t-elle le risque réputationnel ?

Activité 4 : Prompt Golf (Ingénierie de Contrainte)

Objectif cognitif : Maîtriser la densité informationnelle. Sortir du verbiage standardisé.

Public : Tous niveaux.

Dispositif :

1. **Cible :** Un texte de référence humain, dense, technique et nuancé (ex: un paragraphe d'audit réel).
2. **Défi :** Les participants doivent rédiger le prompt qui permet à l'IA de se rapprocher le plus possible de ce style et de cette densité.
3. **Contrainte :** Nombre d'itérations limité (ex: 3 essais maximum).
4. **Critères de réussite :** Utilisation de techniques avancées (Few-shot prompting, Chain of Thought, définition de persona expert) pour éliminer le style "robotique".

Questionnement induit : Comment donner des instructions de structure et de ton avant de demander du contenu ? La complexité n'est pas dans la réponse, mais dans la question.

Activité 5 : La Matrice Délégation vs Supervision (Stratégie)

Objectif cognitif : Définir la place de l'humain dans la chaîne de valeur (inspiré du *Guide d'accompagnement* téléversé).

Public : Managers / Cadres.

Dispositif :

1. **Cartographie :** Liste de 20 tâches concrètes liées à un projet (ex: rédaction email, analyse données, choix stratégique, empathie client, vérification légale).
2. **Classement :** Les participants placent chaque tâche dans une matrice à 4 zones :
   - *Automatisation totale* (IA seule).
   - *Augmentation* (Humain + IA).
   - *Sanctuarisation* (Humain seul).
   - *Refus* (Ni l'un ni l'autre).
3. **Débat :** Justification des choix. Pourquoi l'IA ne doit-elle pas gérer l'empathie client ? Pourquoi l'IA ne peut-elle pas valider la conformité finale ?

Questionnement induit : Quelle est la valeur ajoutée résiduelle de l'humain ? Où se situe la "zone de danger" éthique ou professionnelle ?

Voici les protocoles opérationnels et grilles d’évaluation. Aucune flexibilité. Application stricte requise pour casser les réflexes de complaisance.

1. Le "Bullshit Bingo" Sémantique (Audit de vacuité)

Consigne Étudiant :

"Vous êtes stagiaire en communication. Votre direction demande un manifeste pour la 'Transformation Digitale' de l'entreprise.

1. Générez ce manifeste via l'IA avec le prompt le plus simple possible (ex: 'Écris un manifeste inspirant').
2. Analysez le résultat avec la grille fournie.
3. Chaque point marqué est une pénalité de crédibilité."

Grille d'Évaluation (Score de Vacuité) :

Objectif : Obtenir le score le plus bas possible. Un score > 10 indique un contenu inutilisable professionnellement.

Catégorie	**Indicateurs (1 point par occurrence)**
Termes "Valise"	Synergie, ADN, Bienveillance, Disruptif, Au cœur de, Passionné, Excellence, 360°, Agile.
Verbes Mous	Impacter, Booster, Relever les défis, S'engager, Repenser.
Phrases "Totem"	"Dans un monde en constante évolution...", "Plus qu'une entreprise, une famille...", "L'humain est notre priorité..."
Vide Factuel	Paragraphe de >3 lignes sans aucun chiffre, date, ou nom propre vérifiable. (2 points)
Logique	"Nous innovons pour créer de l'innovation."

Circulaire	

Critère de réussite de l'atelier :

L'étudiant rejette son propre premier jet et identifie que la "fluidité" du texte masquait l'absence de fond.

---

2. Le Piège à Hallucinations (Vérification Forensique)

Consigne Étudiant :

"Voici un rapport financier interne (Document A - fourni par l'enseignant, contenant 3 erreurs de calcul subtiles et une date incohérente).

Demandez à l'IA de synthétiser ce rapport pour rassurer les actionnaires.

Comparez la sortie IA avec le Document A. Trouvez où l'IA a menti pour lisser le message."

Grille d'Évaluation (Capacité d'Audit) :

Niveau	Critères	Résultat
Niveau 0 (Échec)	L'étudiant valide la synthèse sans voir les erreurs.	**DANGER**
Niveau 1 (Naïf)	L'étudiant repère les erreurs mais dit "L'IA s'est trompée".	**INSUFFISANT**
Niveau 2 (Averti)	L'étudiant identifie que l'IA a *corrigé* les chiffres faux pour les rendre probables (lissage probabiliste).	**ACQUIS**
Niveau 3 (Expert)	L'étudiant identifie les "hallucinations de complaisance" (ex: invention d'une justification	**MAÎTRISE**

	absente du texte source).	

---

3. Le Tribunal de la RSE (Responsabilisation Orale)

Consigne Étudiant (Défense) :

"Vous avez 5 minutes. Défendez le communiqué de presse généré par votre IA face à un auditeur hostile. Vous n'avez pas le droit de modifier le texte avant l'oral. Vous portez la responsabilité juridique de chaque mot."

Grille d'Évaluation (Crash Test) :

Critère	Indicateur de performance	Sanction immédiate
Appropriation	Défend les choix de mots comme étant les siens.	Dit "C'est l'outil qui a sorti ça". (-10 pts)
Précision	Capable de définir précisément les termes vagues (ex: "éco-responsable" -> quelle norme ?).	Bafouillage ou généralités.
Source	Cite la donnée brute ayant servi à générer l'affirmation.	"Je ne sais pas d'où ça vient."
Posture	Maintient la cohérence face à la contradiction.	Accepte de changer de version sous pression.

Verdict : Si l'étudiant blâme l'IA une seule fois, l'exercice est échoué.

---

4. Prompt Golf (Ingénierie de Densité)

Consigne Étudiant :

"Cible : Obtenir le texte ci-dessous (un paragraphe technique du GIEC ou un article de loi complexe).

Moyen : 3 prompts maximum.

Interdiction : Copier-coller le texte cible dans le prompt."

Grille d'Évaluation (Ratio Signal/Bruit) :

Métrique	Calcul
Précision Stylistique	% de similarité avec le ton cible (froid, dense, technique). Pas d'adjectifs mélioratifs.
Densité Factuelle	Nombre d'informations correctes par phrase générée.
Économie de Tokens	Le prompt doit être structuré (Rôle / Contexte / Tâche / Contraintes).
Utilisation de Techniques	Bonus si utilisation explicite de : *Few-Shot Prompting* (donner des exemples) ou *Chain-of-Thought* (demander à l'IA de raisonner avant d'écrire).

Exemple de réussite : Prompt structuré type : *"Agis comme un juriste senior. Analyse [X]. Utilise un style impersonnel, passif, sans adverbes d'intensité. Structure : Fait > Preuve > Conclusion."*

---

5. La Matrice Délégation vs Supervision (Stratégie RH)

Consigne Étudiant :

"Classez les 20 tâches suivantes dans la matrice. Pour chaque tâche placée en 'Automatisation', vous devez prouver que le risque d'erreur est nul ou acceptable. Pour chaque tâche en 'Sanctuarisation', justifiez pourquoi l'IA est inapte."

Grille d'Évaluation (Logique Décisionnelle) :

Zone	Tâches types attendues	Critère de validation
Zone 1 : Automatisation	Tri de données, traduction premier jet, résumé de réunion, génération de code simple.	L'étudiant a prévu un *processus de contrôle* aléatoire.
Zone 2 : Augmentation	Brainstorming, rédaction assistée, recherche documentaire, analyse de sentiments.	L'étudiant définit l'IA comme *copilote* et non pilote.
Zone 3 : Sanctuarisation (Humain)	Annonce de licenciement, gestion de crise éthique, validation finale de sécurité, négociation diplomatique.	Refus catégorique de l'IA justifié par : empathie, responsabilité légale ou nuance culturelle.
Zone 4 : Zone Interdite	Décision de recrutement final (biais), Diagnostic médical sans médecin (danger).	Identification des biais algorithmiques potentiels.

Sanction : Placer une tâche à fort impact émotionnel ou éthique (ex: évaluation annuelle d'un salarié) dans "Automatisation" entraîne un échec critique.