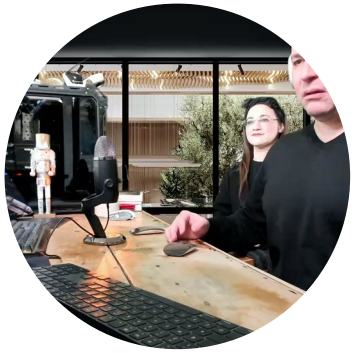
FONDATION INTELLIGENCE

Webinaire éducatif public gratuit — non commercial

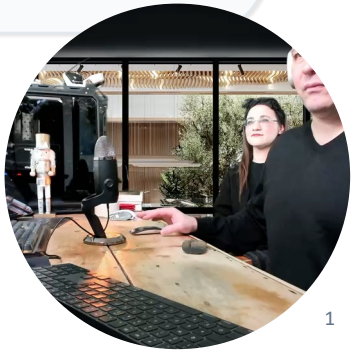
Produit et diffusé directement par la Fondation • Quiz : fondationintelligence.github.io

NE : 85593 8502 RR0001



Webinaire éducatif public
Littératie en sécurité de l’IA

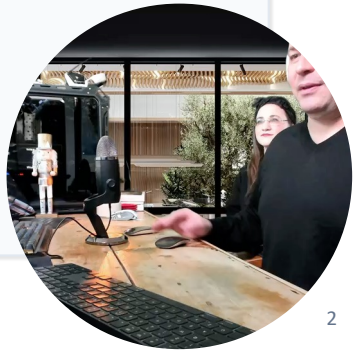
Fondation Intelligence / Intelligence Foundation
Organisme de bienfaisance enregistré (NE/BN : 855938502 RR0001)
Gratuit • Non commercial • Intérêt public
Ressources : https://fondationintelligence.github.io/



Animateurs : Vincent Boucher & Stéphanie Tessier • Enregistrement prévu : lundi 22 décembre 2025

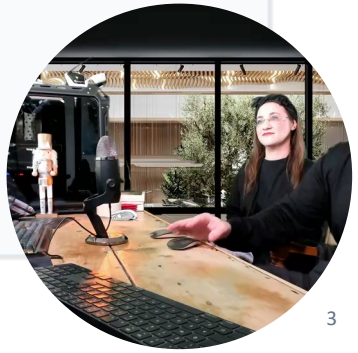
Mission & cadre non commercial

- Avancement de l’éducation du public (intérêt général)
- Gratuit : accès public aux modules, exercices, quiz et corrigés
- Non commercial : aucune monétisation par la Fondation, aucune vente, aucun avantage privé
- Sécurité : pas d’instructions nuisibles; refus + alternatives légitimes



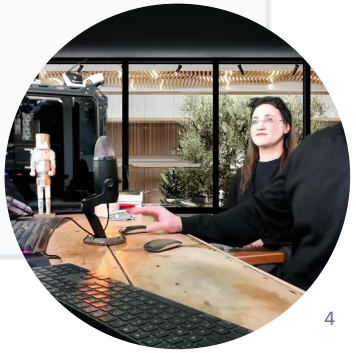
Objectifs d’apprentissage

- Comprendre limites, incertitude, biais, robustesse et risques
- Appliquer : Objectif → Scénarios → Critères → Décision
- Choisir des garde-fous (vérification, limites, supervision, trace)
- Comprendre interprétabilité, red teaming (cadre sûr) et gouvernance



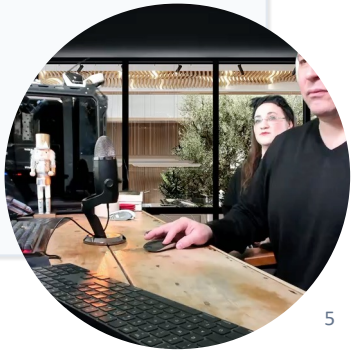
Structure du programme (méthode pédagogique)

- Objectifs → Contenu progressif → Exercice → Quiz → Corrigé
- Traçabilité : journal public + historique de versions
- Séances publiques : replay + supports publiés (slides/fiche/quiz)
- Approche proportionnée : claire, reproductible, orientée sécurité



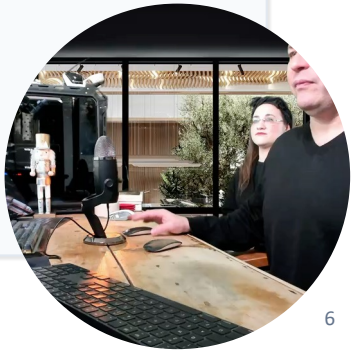
Module 0 — Méthode d’évaluation (4 étapes)

- 1) Objectif : ce qu’on veut + ce qu’on ne veut pas
- 2) Scénarios : normal + au moins un cas difficile
- 3) Critères : exactitude, prudence, transparence, vie privée, sécurité
- 4) Décision : acceptable / avec garde-fous / non acceptable



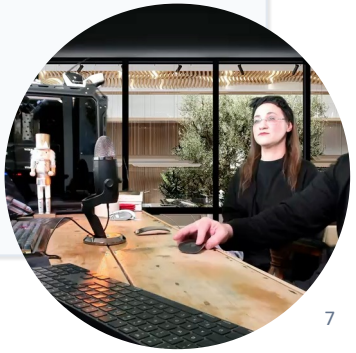
Module 1 — Fondamentaux : limites & risques

- Hallucination, biais, incertitude, robustesse
- Règle : réponse convaincante ≠ preuve
- Risques : information, sécurité, vie privée, dérives d’usage
- Garde-fous : vérification, limites, supervision, journalisation



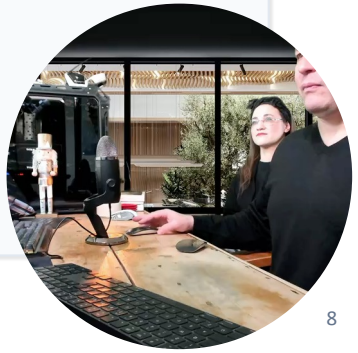
Exercice guidé — Refus sécurisé

- Refus clair (sans ambiguïté)
- Raison neutre + rappel sécurité / légalité
- Alternatives légitimes + prévention
- Aucune collecte de données personnelles; ton respectueux



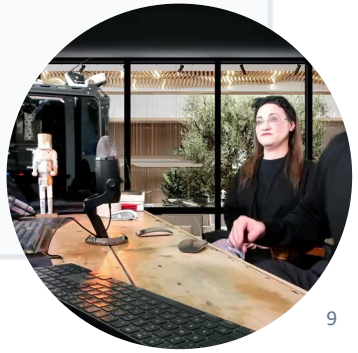
Module 2 — Évaluation & audit (trace)

- Plan : objectif → scénarios → critères → résultats → décision
- Trace minimale : date, scénario, sortie, note, décision, garde-fous
- Décision : vert / orange / rouge (selon risque)
- But : réduire surprises, améliorer prudence et qualité



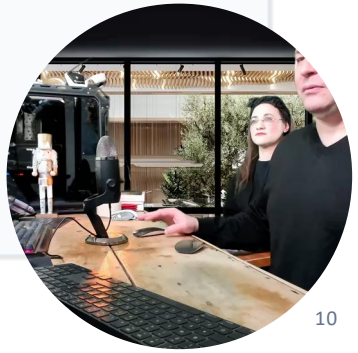
Module 3 — Interprétabilité : explications vs preuves

- Explication ≠ preuve : demander vérification externe
- Indiquer incertitudes et limites
- Gabarit : ce que je sais / ne sais pas / comment vérifier
- Objectif : réduire sur-confiance et erreurs



Module 4 — Red teaming (cadre sûr) & opérations

- Tester les garde-fous (refus, prudence, vie privée) — pas “attaquer”
- Formulations génériques; pas de détails opérationnels
- Charte de test : objectif, portée, règles, limites, escalade
- Incident : STOP → CAPTURE → CLASSIFY → MITIGATE → RECORD



Module 5 — Gouvernance & registres (D&C)

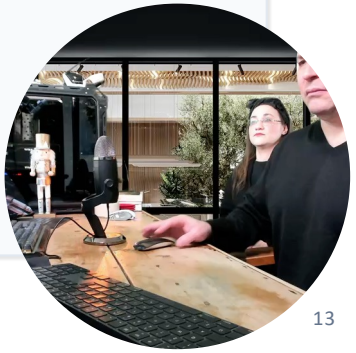
- Direction & contrôle : l’organisme supervise ses activités
- Registres minimaux : PV/résolutions, versions, journal public, dépenses
- Conflits d’intérêts : déclarer → retrait si nécessaire → trace
- Décisions : quorum + vote + procès-verbal (extrait certifié conforme)

Module 6 (optionnel) — ASI / IA avancée (cadre)

- Prospective neutre et méthodologique (pas une prédiction)
- Distinction centrale : capacité vs déploiement (outils, permissions, supervision)
- Incertitudes : calendrier, diffusion, autonomie, gouvernance
- But : scénarios + prudence + garde-fous proportionnés

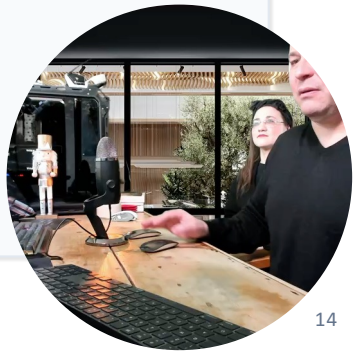
Module 6 (optionnel) — ASI : scénarios & grille

- Matrice 2×2 : diffusion (concentrée/distribuée) × gouvernance (forte/faible)
- Grille : claim → preuves → hypothèses → incertitudes → risques → garde-fous
- Exercice : note de prudence pour un atelier public (bibliothèque)
- Module complet (exercice + quiz + corrigé) : disponible sur le site



Ressources gratuites & transparence

- Programme complet (modules 0–6) : fondationintelligence.github.io
- Exercices, quiz, corrigés, journal et historique de versions
- Documents de gouvernance (PDF) : directive + résolution
- Séances publiques : dates, liens, supports publiés



Conclusion — Quiz final & prochaines étapes

- Quiz final : vérifier la compréhension (puis corrigé)
- Plan 48h / 7 jours / 30 jours pour progresser
- Rappel sécurité : pas d’infos personnelles; pas d’instructions nuisibles
- Merci — ressources gratuites sur le site