

RENDEZ-VOUS DE L'ÉVALUATION 2025



*Société Française
de l'Évaluation*



SOMMAIRE	2
INTRODUCTION AUX ACTES DES RENDEZ-VOUS DE L'ÉVALUATION 2025	4
Le contexte : une crise de confiance démocratique	4
L'évaluation comme levier de restauration démocratique	4
1. COMMENT DÉCRIRE L'AVANCÉE DE L'IA DANS LA SOCIÉTÉ FRANÇAISE EN GÉNÉRAL ET AUPRÈS DES ACTEURS PUBLICS ?	6
Introduction et contexte institutionnel	6
Cadrage général : une technologie systémique non neutre	7
L'IA et l'environnement : entre risques et opportunités	7
L'IA et l'égalité femmes-hommes : encoder l'égalité ?	7
L'IA et l'emploi : analyse de controverse	8
Focus : l'ia dans les services publics	8
Échanges et débats	9
Conclusion de la table ronde	9
2. L'ÉVOLUTION DES MÉTIERS ET DES SAVOIRS MOBILISÉS PAR LES TRAVAUX D'ÉVALUATION. PRATIQUES DE PROFESSIONNALISATION EN FRANCE ET À L'ÉTRANGER	10
Introduction	10
Approche technique : exploitation de fonds documentaires et rag	12
Approche transformative : l'ia au service de l'humain et du participatif	12
Approche institutionnelle : protection des données et souveraineté	12
Débats et questions de la salle	13
3. QUELLES QUESTIONS L'IA POSE-T-ELLE À LA CONCEPTION ET L'ÉVALUATION DES POLITIQUES PUBLIQUES ?	14
Introduction	14
L'IA dans la fabrique des politiques publiques : le cas pôle emploi	15
Débats et interlude : la centralité du dialogue	15
Retour d'expérience : six ans d'ia à la banque mondiale	16
Débats : compétences, résistances et avenir du métier	16
Prospective : les trois temps de l'évaluation	17
Conclusion	17
REMERCIEMENTS ET ENGAGEMENTS	19
INTERVENANT·E·S ET MODÉRATION	20





INTRODUCTION AUX ACTES DES RENDEZ-VOUS DE L'ÉVALUATION 2025

Intelligence Artificielle et Évaluation des Politiques Publiques : défis éthiques, techniques et démocratiques

UN CONTEXTE D'URGENCE ET DE TRANSFORMATION

Le 9 janvier 2025, le Palais d'Iéna, siège du Conseil économique, social et environnemental (CESE), a ouvert ses portes pour accueillir la deuxième édition des « *Rendez-vous de l'évaluation* ». Organisée conjointement par la Société française de l'évaluation (SFE) et le CESE, cette journée d'étude s'est tenue dans un contexte de transformation technologique accélérée, marquant une étape décisive pour la communauté des professionnel-le-s de l'action publique.

Le choix de la thématique, « *Intelligence Artificielle et Évaluation des Politiques Publiques* », ne relève pas du hasard mais d'une nécessité impérieuse. Comme l'a souligné l'ouverture de cette journée, nous vivons une époque charnière où l'intelligence artificielle (IA) ne se contente plus d'être un sujet technique réservé aux spécialistes de l'informatique (IT) ou du marketing. Elle s'impose désormais comme un acteur clé bouleversant nos modes de vie, de travail, et in fine, la fabrique même de l'action publique.

Ces actes ont pour vocation de restituer la richesse des échanges tenus dans l'hémicycle, en croisant les regards institutionnels, techniques et éthiques. Ils s'ouvrent sur les orientations stratégiques partagées lors de la séance inaugurale par la présidence du CESE et la présidence de la SFE, posant les jalons d'une réflexion collective sur l'avenir de nos métiers et de nos démocraties.

L'IA : UNE DUALITÉ ENTRE PROGRÈS INOUIÏS ET ZONES D'OMBRE

Retenu par un déplacement en Outre-mer, Thierry Beaudet, Président du CESE, a tenu à adresser un message vidéo inaugural pour cadrer les enjeux macro-politiques de cette journée. Son intervention rappelle que l'intelligence artificielle est une thématique qui occupe considérablement les conseillers du CESE, nécessitant une approche nuancée qui refuse tant l'angélisme technologique que le rejet systématique.

D'une part, l'IA ouvre des « *horizons de progrès inouïs* ». En automatisant des décisions complexes, elle promet des avancées spectaculaires dans des domaines vitaux pour le bien commun : la médecine prédictive, l'accompagnement éducatif, la sécurité des transports ou encore le traitement massif des données.

D'autre part, cette technologie porte en elle des « *zones d'ombre* » qu'il convient d'aborder avec lucidité.

- Sur le plan économique et social : le risque de remplacement de l'humain par la machine dans de nombreux métiers est une réalité tangible qui appelle à la vigilance.
- Sur le plan politique et démocratique : l'IA possède une capacité inédite à analyser massivement les comportements et à fabriquer des messages susceptibles de façonner l'opinion publique. Ce pouvoir de manipulation représente une menace potentielle pour nos démocraties et nos libertés individuelles.

C'est pour embrasser cette complexité que le CESE a institué une commission temporaire, présidée par Thierry Cadart. Celle-ci a pour mandat d'analyser les impacts de l'IA sur l'action publique, l'emploi, le travail, l'éthique et l'environnement — ce dernier point ayant déjà fait l'objet d'un avis publié.

LA NÉCESSITÉ D'UNE RÉGULATION ET D'UNE ÉVALUATION CITOYENNE

L'un des messages forts portés par cette introduction réside dans le rappel qu'« aucune technologie n'est neutre ». L'IA est indissociable des conditions de sa naissance et des acteurs qui la financent. Thierry Beaudet souligne un paradoxe inquiétant : cette innovation de rupture tend à conforter des acteurs économiques déjà surpuissants, principalement les géants du numérique américains. Dans un contexte géopolitique incertain, marqué par la crainte d'une dérégulation accrue outre-Atlantique, la question de la souveraineté et de la régulation devient centrale.

Face à une technologie « d'emblée dominée par les grands acteurs », l'évaluation ne peut se limiter à une expertise technique. Elle requiert une « expertise d'usage » et une mobilisation de la société civile. Il s'agit d'une mission démocratique essentielle : proposer des « boussoles éthiques » et des régulations anticipatrices face aux développements foudroyants de la technologie.

La France et l'Europe ont déjà ouvert la voie avec le Règlement général sur la protection des données (RGPD) et, plus récemment fin 2023, avec le règlement sur l'IA (AI Act). Ces cadres réglementaires, bien que cruciaux, ne suffisent pas à clore le débat. Le chantier de l'évaluation reste immense et continu, notamment dans la perspective du sommet international sur l'IA prévu à Paris en février.

L'ÉVALUATION DES POLITIQUES PUBLIQUES À L'ÉPREUVE DE L'IA

Si le CESE pose le cadre sociétal, la Société Française de l'Évaluation (SFE), par la voix de sa présidente, Virginie Besrest, a recentré le débat sur les implications professionnelles pour les évaluateurs et évaluatrices.

L'intelligence artificielle s'invite désormais dans l'appréciation de la valeur, de la pertinence, de l'efficacité, de l'efficience et de la cohérence des interventions publiques. Les outils comme le *machine learning*, l'analyse prédictive et surtout l'IA générative offrent des opportunités inédites pour renouveler nos méthodes. Cependant, ces avancées ne sont pas exemptes de défis techniques, éthiques et humains.

L'impératif est clair : garder l'humain au cœur du dispositif. L'IA a été conçue « par l'humain pour l'humain ». Elle doit demeurer un outil au service de nos valeurs fondamentales — éthique, équité, transparence, respect des droits — et viser à renforcer les compétences des professionnel·le·s plutôt qu'à les substituer.

Une préoccupation majeure concerne l'avenir de la profession et la formation des nouvelles générations. Comment les évaluateurs juniors acquerront-ils les compétences critiques nécessaires s'ils délèguent l'analyse première à des algorithmes ? Comment apprendront-ils à identifier les biais inhérents aux analyses produites par l'IA s'ils n'ont pas traversé les années de pratique nécessaires à la construction de ce jugement professionnel ? C'est autour de ces interrogations que les travaux de cette journée ont été structurés.

Comment décrire l'avancée de l'IA dans la société française en général et auprès des acteurs publics ?

MODÉRATEUR :

Thierry Cadart (Conseiller au CESE,
Président de la commission temporaire sur l'IA)

INTERVENANT-E-S :

Erik Meyer (Co-rapporteur de l'avis général sur l'IA au CESE)

Gilles Bernaud-Desroches (Groupe Entreprises,
co-rapporteur de l'avis IA et Environnement)

Fabienne Tatot (Délégation aux droits des femmes
et à l'égalité du CESE)

Jean-Marie Truffat (Commission Travail-Emploi du CESE)

INTRODUCTION ET CONTEXTE INSTITUTIONNEL

Thierry Cadart ouvre la session en rappelant la double légitimité de sa présence à cette tribune : d'une part en tant que responsable de la mission d'évaluation des politiques publiques au sein du CESE, et d'autre part en tant que président de la commission temporaire dédiée à l'intelligence artificielle.

Il réaffirme le positionnement du Conseil économique, social et environnemental dans l'écosystème de l'évaluation. Si le CESE n'a ni l'intention ni la prétention de devenir un organisme d'évaluation technique, il entend y prendre toute sa place en apportant sa valeur ajoutée spécifique : la dimension délibérative et la participation citoyenne, inscrites dans son ADN et renforcées par la loi organique. La coopération avec la SFE est jugée cruciale pour intégrer pleinement le Conseil dans cet écosystème.

Cette table ronde s'inscrit dans une séquence plus large, qualifiée de « *mois de l'IA au CESE* », articulée autour du sommet mondial

de l'IA prévu en février. Thierry Cadart annonce plusieurs événements clés : la présentation en séance plénière de l'avis général sur l'insertion de l'IA dans la société, une « grande conversation » avec le Comité national du numérique le 7 février, ainsi que des événements sur la régulation le 13 février.

Pour cette session, l'objectif est de présenter la production intellectuelle du CESE à travers trois prismes spécifiques : l'environnement, l'égalité femmes-hommes et l'emploi, ainsi qu'un focus sur les services publics.

CADRAGE GÉNÉRAL : UNE TECHNOLOGIE SYSTÉMIQUE NON NEUTRE

Érik Meyer introduit les travaux de la commission temporaire en qualifiant l'IA de «troisième révolution industrielle». Contrairement aux précédentes, celle-ci se distingue par sa vitesse de déploiement et sa capacité à toucher simultanément tous les pans de la société : santé, éducation, handicap, services publics, information et démocratie.

L'approche retenue par le CESE pour évaluer l'IA repose sur la mesure de son utilité sociale, analysée à travers quatre piliers fondamentaux définis collectivement :

1. **La compatibilité avec les limites planétaires** : reconnaissant que l'IA n'est pas neutre en termes de consommation de ressources.
2. **Le progrès social** : l'innovation doit systématiquement améliorer la condition humaine.
3. **Les enjeux économiques.**
4. **Le respect des droits fondamentaux, des libertés individuelles et de l'égalité réelle.**

Érik Meyer insiste sur un point central : les systèmes d'IA ne sont pas neutres. Ils sont créés par des humains, basés sur des algorithmes conçus par des humains et entraînés sur des données choisies par des humains. De ce fait, ils sont intrinsèquement biaisés et peuvent être manipulés, menaçant potentiellement le fonctionnement démocratique.

L'IA ET L'ENVIRONNEMENT : ENTRE RISQUES ET OPPORTUNITÉS

Gilles Bernaud-Desroches présente l'avis sur les effets de l'IA sur l'environnement, voté à l'unanimité des 19 groupes du CESE. Il refuse l'approche binaire qui opposerait radicalement risque et opportunité, préférant une analyse nuancée des grandeurs.

Opportunités : l'IA au service de la transition («IA for Green»)

Gilles Bernaud-Desroches détaille plusieurs domaines où l'IA apporte une contribution positive avérée :

- **Recherche climatique** : accélération de la modélisation pour les chercheurs du GIEC.
- **Surveillance** : monitoring des incendies, des pollutions maritimes et de l'évolution des forêts.
- **Sciences participatives** : exemple de l'application *Pl@ntNet*.
- **Agriculture** : optimisation des semis et gestion météorologique, cruciale pour la sécurité alimentaire.
- **Bâtiment** : gestion intelligente de l'énergie permettant de réduire les consommations de 20 à 40 %, ce qui se traduit directement par une baisse des émissions carbone.

Risques et consommation énergétique

Le rapporteur alerte sur l'empreinte matérielle du numérique. La consommation mondiale d'électricité représente environ 20 % de l'énergie totale. Le numérique consomme 10 % de cette électricité, et l'IA représente déjà 10 % de la consommation du numérique. Les projections indiquent un risque de doublement de cette consommation, ce qui est incompatible avec les objectifs de sobriété.

Il souligne également la difficulté de mesurer l'impact réel («l'incertitude de l'impact») par rapport à la facilité de mesurer la consommation carbone. Il cite l'exemple de la météo : sa précision accrue sauve des vies, mais comment chiffrer cet impact positif face au coût énergétique des calculateurs ?

Lutte contre la désinformation

Gilles Bernaud-Desroches illustre le besoin de rigueur scientifique face aux idées reçues, y compris dans les médias traditionnels. Il cite un reportage du journal de 20h liant faussement les «terres rares» aux batteries, alors que ce problème concerne d'autres composants informatiques. L'IA doit être évaluée sur sa capacité à informer correctement sans dégrader la santé mentale des utilisateurs.

Préconisations

L'avis formule neuf recommandations, dont :

- **L'éducation** : apprendre à utiliser l'IA «quand c'est nécessaire», sur le modèle de l'éducation à l'eau ou à l'énergie, pour favoriser une sobriété numérique.
- **La conception** : favoriser dès l'origine des algorithmes frugaux.
- **L'information citoyenne** : mise en place d'un «éco-score» de l'IA pour éclairer les choix des utilisateurs.

L'IA ET L'ÉGALITÉ FEMMES-HOMMES : ENCODER L'ÉGALITÉ ?

Fabienne Tatot présente la contribution de la Délégation aux droits des femmes et à l'égalité (DDFE). Elle contextualise son propos en rappelant que la société actuelle demeure patriarcale, marquée par des inégalités coûteuses (estimées à 118 milliards d'euros par an, incluant le coût des violences, des inégalités salariales et de santé),.

Une technologie conçue par et pour les hommes

Le constat est celui d'une exclusion massive des femmes du secteur : elles ne représentent que 12 % des chercheurs en IA et 6 % des développeurs. Conséquence directe : les algorithmes sont majoritairement conçus par des hommes, imprégnés de codes virils, ce qui se reflète dans les jeux vidéo et les applications.



Reproduction et amplification des biais

Loin d'être neutres, les algorithmes amplifient les stéréotypes existants :

- **Objets connectés** : les assistants vocaux (Alexa, etc.) ont par défaut des voix et des prénoms féminins, renforçant l'image de la femme au service de l'utilisateur.
- **Sites de rencontres** : les algorithmes favorisent l'appariement de femmes jeunes et moins aisées avec des hommes plus âgés et riches, standardisant une norme relationnelle inégalitaire.
- **Robotique sexuelle** : développement de robots conçus pour assouvir les besoins masculins, heurtant les efforts d'éducation au consentement.
- **Cybercriminalité** : les *deepfakes* (trucages vidéo) à caractère pornographique ciblent les femmes à plus de 90%, dans un vide juridique inquiétant.

Propositions pour une IA égalitaire

La DDFE propose deux axes majeurs :

- 1. Créer les conditions d'une conception égalitaire** : cela passe par l'orientation des filles vers le numérique dès le CP, la constitution d'équipes de conception mixtes, et la sensibilisation des développeurs aux biais. L'utilisation de données synthétiques est suggérée pour corriger les déséquilibres des bases de données historiques.
- 2. Encoder l'égalité** : utiliser la puissance de calcul de l'IA pour mesurer objectivement les inégalités et évaluer l'impact des politiques publiques en faveur de l'égalité. L'IA pourrait ainsi servir à justifier et piloter l'action publique, à condition d'une volonté politique forte.

L'IA ET L'EMPLOI : ANALYSE DE CONTROVERSE

Jean-Marie Truffat rapporte les travaux de la commission Travail-Emploi, réalisés sous forme d'analyse de controverse pour refléter la complexité du sujet sans prétendre à une conclusion définitive.

Une diffusion fulgurante

Il souligne la vitesse d'adoption inédite de l'IA générative : ChatGPT a atteint un million d'utilisateurs en seulement cinq jours, contre deux ans pour Netflix et deux mois pour Instagram. Cette technologie pénètre l'entreprise non seulement par décision stratégique, mais aussi «par le bas», via l'usage personnel des salariés.

Trois axes de tension

L'analyse s'est concentrée sur trois questions :

- 1. Inégalités au travail** : risque d'accroissement des écarts de revenus et de compétences, et fossé grandissant entre les entreprises maîtrisant les outils et les autres.

2. Conditions de travail et santé : l'impact est ambivalent. Si l'IA peut automatiser les tâches répétitives (ce qui est positif), elle risque d'intensifier le travail en supprimant les « *temps de respiration* » interstitiels, menant à une densification mentale et au burn-out.

3. Volume de l'emploi : La destruction d'emplois par substitution s'oppose à la création de nouvelles activités. L'enjeu central devient le partage des gains de productivité (salaire, temps de travail).

L'impératif du dialogue social

Jean-Marie Truffat conclut sur la nécessité absolue d'anticiper. L'introduction de l'IA ne doit pas être subie. Il appelle à un dialogue social renouvelé pour comprendre le « *pourquoi* » de l'IA, définir son usage et éviter le recours à des « *IA d'étagère* » inadaptées. Il alerte également sur les risques liés à la sous-traitance, qui peut rendre les entreprises et administrations prisonnières de solutions externes (« *pieds et poings liés* »).

FOCUS : L'IA DANS LES SERVICES PUBLICS

Érik Meyer reprend la parole pour illustrer l'impact de l'IA sur l'action publique, en se basant notamment sur un rapport de la Cour des comptes concernant la Direction générale des Finances publiques (DGFiP).

Efficience vs substitution

Le constat est contrasté :

- **Réussite** : l'IA est efficace quand elle vient en aide à l'agent, par exemple pour repérer des décharges sauvages via l'analyse d'images.
- **Échec et tensions** : les difficultés surgissent quand l'IA vise à remplacer l'agent. L'exemple de la détection des piscines non déclarées est cité : le système a généré plus de 15% d'erreurs (confondant bâches bleues et piscines), augmentant paradoxalement la charge de travail des agents pour traiter les réclamations.

Perte de souveraineté et hallucinations

Érik Meyer dénonce une stratégie perçue comme visant la suppression d'emplois plutôt que l'amélioration du service. Il pointe un problème majeur de compétence : le recours massif à la sous-traitance et aux cabinets de conseil, protégés par le secret des affaires, empêche la montée en compétence des informaticiens de l'État et crée une dépendance technologique.

Enfin, il aborde les limites des IA génératives (comme le système « *Albert* »). Ces outils fonctionnent par probabilité et non par vérité. Si la requête (le prompt) n'est pas extrêmement précise, l'IA peut « *halluciner* », c'est-à-dire inventer des réponses juridiques ou réglementaires fausses, ce qui est inacceptable dans un contexte administratif.

ÉCHANGES ET DÉBATS

La session de questions-réponses a permis d'approfondir plusieurs thématiques clés avec la salle.

Le temps du dialogue social

Alexandre Ribeiro-Dumas (Conseiller secteur public) a interrogé la pertinence des délais légaux d'information-consultation (3 mois) face à la rapidité de déploiement de l'IA.

→ **Réponses** : Jean-Marie Truffat a plaidé pour une intervention bien en amont, au niveau des conseils d'administration, pour co-construire le projet plutôt que de simplement consulter sur sa mise en œuvre. Érik Meyer a rappelé que le délai légal n'empêche pas une phase de pré-information longue. Fabienne Tatot a mentionné l'existence d'une charte et d'un manifeste produits par l'IRES et les syndicats pour encadrer ces pratiques. Gilles Bernaud-Desroches a nuancé en rappelant que l'IA ne s'installe pas du jour au lendemain (logique non binaire 0/1) et s'inscrit dans une continuité d'outils (comme Excel en son temps).

Confidentialité et « Shadow IT »

Une participante a soulevé le risque de fuite de données via l'usage d'outils grand public comme ChatGPT dans les services publics.

→ **Réponses** : Les intervenants ont confirmé que les outils internes sont soumis au RGPD et sécurisés. Cependant, le risque majeur réside dans l'usage personnel et non contrôlé par les agents (Shadow IT), qui alimentent des modèles publics avec des données sensibles sans en avoir conscience.

L'IA dans le travail social

Laura Jouvart a questionné l'usage de l'IA pour l'évaluation des besoins sociaux (ex : placement d'un enfant), craignant la perte de l'analyse humaine et les biais cognitifs.

→ **Réponses** : Thierry Cadart a reconnu le risque de confier des décisions humaines complexes à des systèmes manquant de transparence. La réponse collective a réaffirmé que la responsabilité juridique reste celle du décideur humain (chef de service), même si l'IA est parfois utilisée comme un « échappatoire » pour diluer cette responsabilité.

Monopoles et régulation sectorielle

Scott Silbon (Consultant) a évoqué le risque de formation de cartels via l'IA, citant l'exemple d'un logiciel immobilier aux USA utilisé par quasi tous les propriétaires pour fixer les prix, créant un monopole de fait.

→ **Réponses** : Thierry Cadart et Érik Meyer ont identifié cela comme un problème politique majeur de concentration de pouvoir et de « boîte noire » (secret des affaires). La transparence des algorithmes est présentée comme la seule solution viable. Gilles Bernaud-Desroches a ajouté que le secteur privé lui-même craint ces monopoles qui menacent la liberté d'entreprendre, créant une alliance objective entre public et privé pour une régulation.

CONCLUSION DE LA TABLE RONDE

Thierry Cadart clôture la séance en soulignant que l'IA est éminemment un « sujet politique » qui nécessite une régulation et une évaluation continues, confirmant l'apport des travaux du CESE pour éclairer ces enjeux.



L'évolution des métiers et des savoirs mobilisés par les travaux d'évaluation

Pratiques de professionnalisation en France et à l'étranger

MODÉRATEUR :

Alix de Saint-Albin (Société Française de l'Évaluation)

INTERVENANT-E-S :

Guillaume Soto (Expert data/IA, en direct de Yaoundé)

Silva Ferretti (Consultante indépendante en évaluation de programmes de développement)

Sandrine Beaujean (Directrice département Évaluation, Agence Artemis, Luxembourg)

INTRODUCTION

Alix de Saint-Albin ouvre cette seconde table ronde en la situant dans la continuité logique des travaux présentés précédemment par le CESE. Si la première partie de la matinée a permis de dresser un panorama sociétal et institutionnel des impacts de l'intelligence artificielle, cette seconde séquence se veut résolument pragmatique et opérationnelle.

L'objectif affiché est de « mettre les mains dans le moteur ». Il ne s'agit plus ici de théoriser sur les impacts macroscopiques de l'IA, mais d'examiner comment cette technologie pénètre, transforme et percute le quotidien des professionnel-le-s de l'évaluation des politiques publiques. L'évaluation, profession spécialisée dans l'appréciation des impacts sur les groupes sociaux et l'action publique, se trouve elle-même objet de transformation.

La session est structurée selon une logique de gradation dans la complexité et l'appropriation technique :

- 1. L'usage immédiat et transformatif** : utiliser les outils « sur étagère » (type ChatGPT) pour repenser le métier (intervention de Sylvain Ferretti).
- 2. L'usage technique et structuré** : développer des systèmes spécifiques pour interroger des bases documentaires via des méthodes de RAG (intervention de Guillaume Soto).
- 3. L'usage institutionnel et prudentiel** : la stratégie d'une organisation (agence Artemis) pour intégrer l'IA tout en garantissant la protection des données et la conformité réglementaire (intervention de Sandrine Beaujean).



APPROCHE TECHNIQUE : EXPLOITATION DE FONDS DOCUMENTAIRES ET RAG

Guillaume Soto, ingénieur télécom de formation et expert en analyse de données pour l'évaluation, intervient à distance depuis Yaoundé (Cameroun). Il propose une plongée technique dans l'utilisation de l'IA pour gérer l'infobésité caractéristique des métiers de l'évaluation.

Le défi du volume de données

Le point de départ est un constat partagé par toute la profession : l'évaluation produit et utilise une quantité massive de données (rapports, retranscriptions d'entretiens, bibliographies, bases de connaissances). Le défi n'est plus de trouver l'information, mais d'y accéder rapidement et d'en extraire une synthèse pertinente. Pour illustrer son propos, Guillaume Soto s'appuie sur un cas d'usage réel mené avec l'OFAC (Observatoire des forêts d'Afrique centrale). Cet organisme dispose d'un fonds documentaire de plus de 3000 documents (articles scientifiques, rapports politiques, documents gouvernementaux). L'objectif du projet était de créer un chatbot capable de répondre aux questions des utilisateurs en puisant dans cette base immense.

Comprendre la «boîte noire» et les hallucinations

Avant de détailler la solution, l'intervenant tient à clarifier le fonctionnement intrinsèque des modèles de langage (LLM). Il utilise l'analogie de l'apprentissage de l'enfant : on ne programme pas une IA comme un logiciel classique, on l'entraîne en lui montrant des millions d'exemples.

- L'effet boîte noire : tout comme l'IRM permet de voir l'activité du cerveau sans pour autant décoder le contenu exact de l'information stockée, on peut observer une IA sans savoir précisément comment elle code l'information.
- Le risque d'hallucination : sans accès à une base de données structurée externe, l'IA fonctionne comme un «brillant étudiant» qui a beaucoup lu mais qui, au moment de l'examen, peut inventer des faits précis ou dire des absurdités avec beaucoup d'aplomb. Pour corriger ce défaut, il est nécessaire de lui fournir les documents de référence au moment de la question, transformant l'IA en un assistant qui synthétise des sources vérifiées plutôt que d'inventer des réponses.

La solution technique : Le RAG

(Retrieval Augmented Generation)

Guillaume Soto détaille l'architecture technique mise en place, connue sous l'acronyme RAG (Génération Augmentée par la Récupération). Ce système repose sur trois piliers :

1. **La base documentaire vectorisée** : les documents ne sont pas stockés sous forme de texte simple, mais transformés en vecteurs numériques représentant les concepts et les idées.
2. **Le système de recherche** : il récupère les fragments de texte pertinents.
3. **Le LLM (Large Language Model)** : il rédige la réponse finale à partir des fragments trouvés.

Fonctionnement concret : lorsqu'un utilisateur pose une question (ex : «Quels sont les pays membres de l'OFAC?»), le système ne cherche pas des mots-clés exacts. Il «vectorise» la question, c'est-à-dire qu'il la convertit en concepts mathématiques (concept de «pays», de «membre», d'«organisation»). Il compare ensuite ce vecteur avec ceux de la base documentaire pour identifier les passages traitant des mêmes idées.

Dans l'exemple cité, le système a identifié trois passages potentiels :

- Un passage listant explicitement les membres.
- Un passage mentionnant le Tchad et la Centrafrique comme parties prenantes.
- Un passage évoquant Madagascar ayant «consulté» l'OFAC (faux positif pour la notion de membre). Le LLM analyse ces extraits, écarte l'information non pertinente (Madagascar) et synthétise une réponse précise.

Avantages et limites

Les avantages majeurs pour l'évaluation :

- **Recherche sémantique** : pas besoin de connaître le mot-clé exact ou le jargon technique ; l'IA comprend l'intention et les concepts.
 - **Multilinguisme** : le système fonctionne indifféremment en français ou en anglais, favorisant l'inclusion des différents acteurs (cas de l'OFAC).
 - **Fiabilité et Sourçage** : contrairement à ChatGPT utilisé «à vide», ce système cite ses sources. L'utilisateur peut vérifier d'où vient l'information, ce qui est crucial pour la rigueur évaluative.
- Les limites actuelles : Le système excelle sur des questions factuelles ou de synthèse simple (ex : «Quelle est la réglementation sur les aires protégées?»). En revanche, il montre ses limites sur des questions analytiques complexes nécessitant une méta-analyse transversale (ex : «Quelles sont les principales solutions pour lutter contre la déforestation?»). Si la réponse n'est pas explicitement formulée dans les textes, l'IA peine encore à construire un raisonnement déductif complexe à partir de sources éparpillées.

Outils recommandés et sécurité

Guillaume Soto conclut par des recommandations pratiques pour les évaluateurs souhaitant se lancer :

- Pour les petites équipes (besoin ponctuel) : utiliser NotebookLM (Google). C'est l'outil le plus accessible pour analyser un corpus de documents sans coder.
- Pour les données internes d'organisation : des solutions comme Copilot ou Dust (société française) permettent d'interroger des messageries et des fichiers internes.
- Pour l'évaluation de l'IA elle-même : il insiste sur la responsabilité de l'évaluateur vis-à-vis des outils qu'il déploie. Il recommande l'usage de Giskard, un outil permettant de simuler des attaques (questions de hackers) pour vérifier la robustesse et l'éthique du chatbot (biais racistes, sexistes, hallucinations).



APPROCHE TRANSFORMATIVE : L'IA AU SERVICE DE L'HUMAIN ET DU PARTICIPATIF

Silva Ferretti, consultante indépendante, propose une vision radicalement différente. Plutôt que de se focaliser sur l'archivage ou la recherche documentaire, il explore comment l'IA peut libérer l'évaluateur des tâches bureaucratiques pour le ramener vers l'essence de son métier : la relation humaine.

Démonstration en temps réel : la puissance de l'analyse instantanée

Pour prouver l'efficacité des outils «sur étagère» (ChatGPT, Claude), Silva Ferretti a réalisé deux expériences en direct durant la conférence :

1. **Analyse d'enquête** : elle a soumis les résultats bruts d'un questionnaire (auquel la salle avait répondu) à l'IA. En quelques secondes, l'outil a généré un rapport structuré, identifié les thèmes principaux et produit une présentation PowerPoint avec des graphiques.
2. **Analyse d'image** : elle a pris une photo de la salle et demandé à l'IA d'analyser l'audience. L'outil a fourni des statistiques démographiques (tranches d'âge, équilibre de genre, niveau d'engagement, contexte professionnel) à partir de la simple image. Cette démonstration soulève des enjeux majeurs sur la vie privée, notant que si un outil grand public peut le faire, des acteurs comme Meta disposent de capacités bien supérieures.

La remise en cause du modèle traditionnel

Le constat de Silva Ferretti est sans appel : l'IA sait déjà faire ce qui constitue 80 % du temps facturé par un évaluateur traditionnel :

- **Analyser des données.**
- **Rédiger des rapports.**
- **Créer des supports de communication (diaporama, podcasts).**
- **Préparer des outils méthodologiques et des guides d'entretien.**

Si l'évaluation se résume à la production de «rapports bureaucratiques», alors la profession est condamnée à être remplacée, car l'IA le fait plus vite et moins cher. L'administration publique pourrait théoriquement se passer des évaluateurs pour ces tâches.

Vers une évaluation relationnelle et participative (cas du Kenya)

Pour survivre et rester pertinent, l'évaluateur doit investir ce que l'IA ne peut pas faire : la relation, le sens et le dialogue politique. Silva Ferretti illustre cela par une mission réalisée au Kenya avec une ONG locale.

Au lieu de réaliser une évaluation extractive (l'expert vient, prend les données, fait son rapport), elle a utilisé l'IA pour co-construire l'évaluation avec les bénéficiaires :

- **Co-crédation de la méthode** : l'IA a aidé à générer rapidement des outils méthodologiques, libérant du temps pour discuter de leur pertinence avec les acteurs locaux.
- **Collecte par les pairs** : ce sont les clients du service (les bénéficiaires) qui ont mené les entretiens auprès d'autres clients. L'entretien est devenu un moment de relation et de renforcement communautaire plutôt qu'une simple collecte de données.

- **Analyse assistée** : l'analyse des données, habituellement le goulot d'étranglement (bottleneck) qui exclut les non-experts, a été confiée à l'IA. Cela a permis aux acteurs locaux d'accéder aux résultats instantanément et de les critiquer.

- **Dialogue politique** : l'évaluateur a utilisé les résultats pour faciliter un dialogue direct entre les citoyens et l'administration de l'hôpital.

La conclusion de Silva Ferretti est philosophique : « *L'IA connaît ce qu'on a écrit sur la réalité, mais pas la réalité elle-même* ». L'évaluateur devient un facilitateur qui utilise la machine pour gérer la technique, afin de consacrer son énergie à l'interprétation collective et à la négociation du sens entre les parties prenantes.

APPROCHE INSTITUTIONNELLE : PROTECTION DES DONNÉES ET SOUVERAINETÉ

Sandrine Beaujean, de l'agence Artemis (Luxembourg), apporte un contrepoint nécessaire en exposant les défis d'une structure établie face à l'innovation. Forte de son héritage dans la statistique officielle (Eurostat, Statec), l'agence Artemis place la protection des données au cœur de ses préoccupations, ce qui rend l'adoption de l'IA complexe.

Le parcours du combattant : de l'exploration à la conformité

Sandrine Beaujean décrit le cheminement de son agence, marqué par une prudence méthodologique :

- Mars 2023 : premier contact avec le MIT pour tester l'automatisation de tâches d'évaluation.
- Août 2023 : premier test concret sur des entretiens anonymisés d'une évaluation clôturée. Les résultats furent mitigés, ne répondant pas aux attentes de qualité finale.
- Janvier 2024 : prise de conscience des enjeux de souveraineté et de conformité (RGPD) suite à une conférence de la Chambre de commerce.
- Février 2024 : rencontre avec la CNPD (Commission Nationale pour la Protection des Données) du Luxembourg. L'agence participe au «DaproLab», une session de deux heures où des experts juridiques et techniques dissèquent les pratiques de l'entreprise. Cet épisode est vécu comme un «électrochoc».

Le dilemme de la confiance et la solution «Locale»

Les experts de la CNPD ont posé un ultimatum technique : soit faire confiance aveuglément aux déclarations de conformité des géants américains (ce que la culture de statisticien d'Artemis refuse), soit développer sa propre IA (impossible pour une PME), soit installer une IA en local.

C'est cette troisième voie qu'Artemis a choisie. L'agence investit dans des infrastructures lourdes pour faire tourner des modèles d'IA sur ses propres serveurs, garantissant qu'aucune donnée sensible ne transite par des serveurs tiers.

Résultats des tests et limites actuelles

Sandrine Beaujean présente un bilan honnête et nuancé des tests réalisés en interne par son groupe de travail « IA » (composé majoritairement de jeunes data scientists) :

- **Analyse d'entretiens** : limitée par la qualité des intrants (tous les entretiens ne sont pas transcrits intégralement).
- **Comptes-rendus de réunion** : dépendance forte à la structure de la réunion. Une réunion type brainstorming donne des résultats chaotiques, contrairement à une réunion formelle. Problème de gestion du multilinguisme.
- **Résumé de rapports** : expérience décevante avec l'IA locale qui a souffert d'hallucinations ou de blocages techniques (le système promettant un résumé qui n'arrive jamais). Elle conclut en interrogeant l'évolution des référentiels de compétences. Citant les 36 compétences du référentiel canadien, elle se demande si l'évaluateur de demain doit devenir un « super-évaluateur » maîtrisant à la fois son métier et la science des données, ou si l'IA permettra d'alléger ce fardeau.

DÉBATS ET QUESTIONS DE LA SALLE

La session de questions-réponses a permis de confronter les approches enthousiastes et prudentes, soulevant des questions fondamentales sur l'éthique et l'avenir du métier.

La question de la confiance et de la « Recette »

François Mouterde et d'autres participants ont posé la question centrale de la confiance : comment faire confiance à une « boîte noire » ? Comment valider le travail d'une IA ?

→ **Réponse technique (G. Soto)** : l'utilisation d'outils de test comme Giskard permet d'automatiser la vérification en posant des milliers de questions pièges pour identifier les biais de régression.

→ **Réponse pratique (S. Ferretti et G. Soto)** : la confiance se construit par l'usage itératif. Il ne faut jamais accepter le premier résultat brut. Il faut dialoguer avec l'IA, la challenger, et trianguler ses réponses. Sylvain Ferretti compare cela à la relation avec un collègue humain dont on apprend à connaître les forces et les faiblesses au fil du temps.

L'inflation des outils et le rôle de la SFE

Face au vertige des « 27 000 outils d'IA » existants mentionnés par un participant (Alexis Gallet), la question de la veille technologique pour les petites structures s'est posée.

→ **Réponse** : Sandrine Beaujean a admis l'impossibilité pour une seule structure de tout tester.

→ **Annonce** : François Mouterde a saisi l'occasion pour annoncer la création officielle d'un groupe de travail thématique « Technologies numériques, IA et Évaluation » au sein de la SFE, dont l'objectif sera justement de mutualiser cette veille et de partager les bonnes pratiques.

Anonymat et open data

Florent Loiseau a interpellé le panel sur l'excès d'anonymisation dans l'évaluation actuelle, plaidant pour une ouverture plus grande des données (open data) pour nourrir les modèles et la transparence démocratique.

→ **Réponse** : Sandrine Beaujean a confirmé que l'IA offre paradoxalement des solutions puissantes pour anonymiser massivement des bases de données, ce qui pourrait faciliter leur partage ultérieur sans risque pour les individus.

Désirabilité du métier et compétences

Caroline Massot (ANAP) a soulevé une question existentielle sur la « désirabilité » du métier. Si l'IA prend en charge l'analyse et la synthèse, que reste-t-il ? Elle exprime sa préférence pour former des juniors humains plutôt que d'entraîner des algorithmes, et s'inquiète de la perte de compétence critique si la « boîte noire » n'est jamais ouverte.

→ **Réponse** : le panel a réaffirmé que l'IA doit être vue comme un assistant qui gère la « douleur » (les tâches répétitives et ingrates) pour laisser place au plaisir de l'interprétation et de la stratégie. Guillaume Soto ajoute que l'IA permet parfois de mieux identifier nos propres biais humains en standardisant l'analyse.

Conclusion philosophique : ce que l'IA ne peut pas faire

Une dernière intervention de la salle a invité à lister ce que l'IA ne pourra jamais faire, pour rassurer sur l'avenir de la profession. Les participants ont identifié :

- **Créer de nouvelles connaissances (recherche-action).**
- **Comprendre le contexte politique sensible et non-dit.**
- **Interagir physiquement avec la réalité.**
- **Ressentir l'empathie nécessaire à la compréhension fine des situations sociales.**

Silva Ferretti a conclu la table ronde par cette formule marquante : l'IA ne traite que des représentations (ce qui est écrit), alors que l'évaluateur se frotte à la réalité vivante, contradictoire et humaine. C'est dans cet écart irréductible que réside la valeur ajoutée pérenne de l'évaluation humaine.

Quelles questions l'IA pose-t-elle à la conception et l'évaluation des politiques publiques ?

MODÉRATEUR :

François Mouterde (Société Française de l'Évaluation)

INTERVENANT·E·S :

Aurélie Simard (Directrice générale du centre d'expertise de l'INRIA pour la coopération internationale)

Estelle Raimondo (Responsable du service des méthodes d'évaluation, Groupe d'évaluation indépendant - IEG, Banque Mondiale)

Steve Jacob (Professeur au département de sciences politiques, Université Laval, Québec)

INTRODUCTION

François Mouterde ouvre cette troisième et dernière table ronde de la journée, marquant une progression logique dans la réflexion collective. Après une matinée consacrée aux enjeux sociétaux globaux (Table Ronde 1) puis aux applications techniques concrètes de terrain (Table Ronde 2), cette session de l'après-midi vise à se situer à l'intersection de ces deux mondes. L'objectif est de tirer des leçons générales et de se projeter dans l'avenir, en abordant les défis structurels posés par l'intelligence artificielle à l'évaluation des politiques publiques.

Le modérateur rappelle les cinq grands défis identifiés lors de la préparation de la session :

1. Le défi technique : Comment l'IA améliore-t-elle l'analyse des liens de cause à effet et l'exploitation des données ? Ce point a été largement traité en matinée.

2. Le défi humain : Il concerne l'impact sur les compétences professionnelles et les changements culturels, faisant le lien avec le nouveau référentiel de compétences de la SFE.

3. Le défi décisionnel : Il interroge l'insertion de l'IA dans le cycle des politiques publiques. L'IA peut-elle conduire à une évaluation en continu ? Les évaluations seront-elles déclenchées par des algorithmes ? Des IA dialogueront-elles entre elles pour formuler des recommandations ?

4. Le défi de gouvernance : Comment l'IA aide-t-elle à délimiter le périmètre des évaluations et à identifier les parties prenantes ?

5. Le défi éthique : Il englobe les risques de dépendance aux éditeurs mondiaux et les enjeux déontologiques.

Pour traiter ces questions complexes, le panel réunit des profils complémentaires naviguant entre les sphères scientifique, stratégique et opérationnelle.

- Aurélie Simard apporte son expertise issue de l'INRIA et de sa thèse sur l'adoption de l'IA dans le secteur public, avec une étude de cas approfondie sur France Travail.

- Estelle Raimondo, intervenant depuis Washington, partage l'expérience de la Banque Mondiale (IEG) où elle dirige les méthodes d'évaluation et a conseillé le gouvernement canadien.

- Steve Jacob, en direct de Québec, offre une perspective académique sur les impacts sociétaux de l'IA et l'administration publique, étant par ailleurs un compagnon de route de longue date de la SFE et du master d'évaluation de Lyon.

L'IA DANS LA FABRIQUE DES POLITIQUES PUBLIQUES : LE CAS PÔLE EMPLOI

Aurélié Simard ouvre les interventions en partageant les enseignements de son travail de recherche-intervention mené au sein de Pôle Emploi (aujourd'hui France Travail). Elle propose d'analyser comment l'IA transforme non seulement l'exécution, mais la conception même des politiques publiques.

De la transformation numérique à la rupture de l'IA

L'intervenante distingue d'abord l'IA des vagues précédentes de transformation numérique. Jusqu'en 2018, la logique dominante était celle du New Public Management : numériser pour rendre l'administration plus efficace, dématérialiser les processus et réorganiser la relation usager. Le rapport Villani de 2018 a marqué une rupture, présentant l'IA comme une révolution touchant tous les secteurs. Aurélié Simard confie que cette transition a fait « voler en éclats » ses propres référentiels de transformation numérique classique.

Elle a observé deux expérimentations majeures chez Pôle Emploi :

1. Un chatbot destiné aux demandeurs d'emploi pour leur préconiser des démarches basées sur leurs données personnelles.
2. Un outil d'assistance pour les conseillers, générant des brouillons de réponses aux emails des usagers.

Une nouvelle gestion de projet : l'incertitude au cœur

La différence fondamentale entre un projet informatique classique et un projet IA réside dans le point de départ.

- **Projet classique** : on part d'un cahier des charges précis, définissant des règles de gestion, des objectifs et des moyens.
- **Projet IA** : on démarre dans une « zone d'incertitude ». L'essentiel est d'avoir un problème à résoudre suffisamment critique pour justifier l'investissement, et de disposer de données potentielles pour le traiter. On ne code pas des règles, on explore des données pour trouver des corrélations et des solutions.

Le défi du dialogue pluridisciplinaire

La réussite de ces projets repose sur la qualité du dialogue entre deux mondes souvent cloisonnés : les experts de la donnée (qui explorent) et les experts métiers (qui connaissent la réalité du terrain). Ce dialogue est complexe car il doit aligner des objectifs parfois divergents :

- **Facilitation du travail** : l'IA doit-elle alléger les tâches rébarbatives des agents face à la complexité croissante de la réglementation ?
- **Qualité de service** : s'agit-il d'améliorer des indicateurs de productivité (temps de réponse) ou de cibler les populations les plus fragiles (équité) ?

Selon l'objectif choisi, les données collectées et les algorithmes utilisés ne seront pas les mêmes. L'évaluateur a un rôle crucial à jouer dès cette phase amont pour aider à définir les indicateurs de succès et les écueils à éviter (protection des données personnelles, biais). Aurélié Simard souligne qu'on ne conçoit pas une IA pour orienter un demandeur d'emploi avec la même tolérance à l'erreur qu'une IA pour une centrale nucléaire ; ce compromis performance/risque doit être explicité.

Passage à l'échelle et « usines à IA »

Si l'expérimentation est la norme au départ, le véritable défi est le passage à l'échelle. Beaucoup d'administrations échouent à franchir ce cap. Le passage à l'échelle nécessite des infrastructures robustes, qualifiées d'« usines à IA », capables de traiter les données de manière massive, sécurisée et conforme.

Enfin, Aurélié Simard insiste sur la boucle de rétroaction. Les résultats des expérimentations IA doivent remonter aux concepteurs des politiques publiques pour ajuster non seulement les services, mais aussi les politiques de régulation et de soutien à l'innovation, créant une articulation nécessaire entre les niveaux locaux, nationaux et supranationaux.

DÉBATS ET INTERLUDE : LA CENTRALITÉ DU DIALOGUE

À la suite de cette intervention, François Mouterde interpelle le panel sur la notion de dialogue, qui semble paradoxale pour un utilisateur lambda percevant l'IA comme une « boîte noire » ou une solution miracle unilatérale.

Le dialogue des données et des équipes

Aurélié Simard précise que le dialogue est omniprésent, même s'il est invisible. L'utilisateur « dialogue » avec la machine via ses données comportementales (clics, navigation), ce qui permet au système de s'ajuster. Mais le dialogue le plus critique est humain : c'est celui qui doit avoir lieu au sein des équipes de conception pour définir ce qu'est un système « satisfaisant » et éthique.

Le besoin de médiateurs et de « courtiers de connaissances »

Estelle Raimondo rebondit vigoureusement sur ce point. Pour elle, le manque de profils hybrides est le principal frein actuel. Elle identifie un besoin criant de « courtiers de connaissances » (knowledge brokers), capables de naviguer entre trois mondes : l'informatique pur (IT), la science des données (data science) et le métier de l'évaluation. Sans ces traducteurs, les risques de malentendus et de « gadgetisation » sont élevés.

Aurélié Simard complète en citant l'exemple positif de la startup ayant accompagné Pôle Emploi, qui mobilisait des designers. Ces profils, dotés d'une culture générale en IA et de compétences en design de service, ont facilité la traduction des besoins entre techniques et métiers.

Distinguer les IA : gadget vs outils sérieux

Steve Jacob apporte une nuance technique fondamentale. Il appelle à ne pas confondre l'IA générative (type ChatGPT, qui crée du contenu probabiliste) et le Machine Learning (algorithmes prédictifs basés sur des données historiques), qui ne servent pas les mêmes objectifs. Il cite un rapport récent de l'organisation RAND montrant que 80% des projets d'IA dans le secteur privé échouent. Les causes principales sont une mauvaise compréhension par le leadership (attentes disproportionnées, vision « baguette magique ») et une inadéquation entre les équipes



techniques et les besoins réels. Pour réussir, l'informaticien doit être «derrière l'épaule» de l'employé lors des premières phases d'implémentation.

RETOUR D'EXPÉRIENCE : SIX ANS D'IA À LA BANQUE MONDIALE

Estelle Raimondo prend la parole pour détailler la trajectoire du Groupe d'évaluation indépendant (IEG) de la Banque mondiale. Son service, composé d'une petite équipe centrale (6 personnes dont 3 data scientists) au service de 120 évaluateurs, a mené des expérimentations systématiques depuis six ans.

Les cas d'usage : Texte et Image

L'IEG utilise l'IA principalement dans deux domaines :

1. Le Texte (Gain d'efficience) : dans une bureaucratie produisant des milliers de documents, l'IA (Machine Learning puis Générative) est utilisée pour classer les projets et analyser les portefeuilles. Cela permet de libérer les analystes des tâches ingrates de classification manuelle pour se concentrer sur l'interprétation.

2. L'Image (Nouvelles questions) : l'utilisation de données géospatiales et d'images (ex: Google Street View) permet d'évaluer l'impact de projets d'infrastructure ou de développement urbain là où les données traditionnelles manquent.

Estelle Raimondo évoque une «triple boucle d'apprentissage» : au-delà de l'efficience et des nouvelles données, l'IA pousse l'organisation à remettre en question ses propres critères de rigueur et ses hypothèses méthodologiques. Parfois, la machine s'avère plus fiable que l'humain.

Une gouvernance «tâche par tâche»

L'approche de la Banque mondiale est pragmatique. Plutôt que de s'appuyer uniquement sur des cadres éthiques généraux (comme ceux de l'UE, jugés utiles mais insuffisants pour l'opérationnel), l'IEG a mis en place une gouvernance «tâche par tâche». Pour chaque application, on évalue les risques spécifiques : a-t-on besoin d'une précision absolue (ex: chiffres financiers pour le Financial Times) ou d'une aide créative (ex: revue de littérature) ? Cette approche nécessite des mécanismes de décision agiles, capables de dire «non» à un projet si le risque est trop élevé dans un contexte donné.

Ressources humaines et stratégie

Estelle Raimondo insiste sur trois piliers RH :

- 1. Alphabétisation minimale** : tout le personnel, y compris la direction, doit acquérir un socle commun de compréhension de l'IA.
- 2. Internalisation des compétences** : la meilleure décision de l'IEG a été de recruter des data scientists en interne plutôt que d'externaliser (outsourcing). Cela garantit la maîtrise et la pertinence des outils.
- 3. Profils juniors** : les descriptions de poste des nouveaux analystes incluent désormais explicitement des compétences en code et visualisation de données.

DÉBATS : COMPÉTENCES, RÉSISTANCES ET AVENIR DU MÉTIER

La discussion s'engage sur la transformation des métiers et la gestion du changement.

Le cycle d'acceptation de l'IA

Steve Jacob partage les résultats de ses recherches sur la psychologie des employés face à l'IA. Il observe un cycle en quatre temps :

- 1. Crainte** : peur d'être remplacé.
- 2. Contestation** : rejet des résultats de l'IA, l'employé estimant que son expertise est supérieure.
- 3. Acceptation** : utilisation progressive.
- 4. Protection (biais d'automatisation)** : l'employé finit par sur-protéger l'outil. Steve Jacob cite l'exemple d'une organisation où l'IA, surnommée «Bill», est défendue par les collègues comme l'un des leurs.

Ce dernier stade présente un risque majeur de dépendance et de perte de sens critique. Il cite des cas inquiétants aux États-Unis (système Compass pour les libérations conditionnelles) et en Espagne (violences conjugales), où juges et policiers suivent les scores de risque de l'IA dans 95% des cas, renonçant à leur propre jugement situationnel.

Quelles compétences pour les jeunes évaluateurs ?

À la question de Steve Jacob sur les compétences «résiduelles» indispensables à transmettre aux étudiants, Estelle Raimondo identifie quatre piliers que l'IA ne remplace pas :

- 1. La facilitation du dialogue entre les parties prenantes.**
- 2. L'analyse critique** : adopter une posture «enthousiaste mais sceptique».
- 3. La communication d'influence** : savoir porter des messages politiques complexes.
- 4. La compréhension technique** : savoir ce qu'est un modèle pour ne pas le voir comme de la magie.



Questions de la salle : environnement et management

Un expert en évaluation environnementale interroge l'application de l'IA aux audits de risques environnementaux à la Banque Mondiale. Estelle Raimondo précise que si l'audit systématique des risques reste une procédure standard, l'IA est surtout utilisée pour des évaluations thématiques (biodiversité, énergies renouvelables) grâce aux images satellites validées par la littérature scientifique.

Vincent Tison (CD41/SFE) interroge la gestion des résistances internes («On aimait bien l'approche conventionnelle»). Estelle Raimondo répond par la preuve de l'utilité : l'IA a permis de supprimer les tâches les plus pénibles (classification de milliers de projets), ce qui a créé un consensus positif. De plus, elle a permis de répondre à des questions jusqu'alors inaccessibles, comme l'analyse du contrat social via les réseaux sociaux. La stratégie est celle de la complémentarité : on ne change pas ce qui fonctionne très bien avec les méthodes traditionnelles.

PROSPECTIVE : LES TROIS TEMPS DE L'ÉVALUATION

En guise de conclusion et d'ouverture, Steve Jacob présente une grille de lecture issue d'un article récent publié dans Digital Government. Il décrit l'évolution de l'évaluation en trois temps.

Temps 1 : l'évaluation augmentée (le présent)

C'est l'ère actuelle, dominée par l'IA générative. Elle permet d'améliorer 72% des étapes de l'évaluation (définition d'indicateurs, analyse, restitution adaptée aux publics) et de générer des gains d'efficacité d'environ 35%. Le défi majeur est la découvrabilité des outils (savoir qu'ils existent) et l'apprentissage par l'essai-erreur. Steve Jacob recommande des outils fiables comme Consensus, qui interroge uniquement la littérature scientifique pour éviter les hallucinations des modèles grand public.

Temps 2 : l'évaluation intégrée (le futur proche)

L'objectif est de réduire la boucle de rétroaction. Souvent critiquée pour arriver «trop tard» (des années après la mise en œuvre), l'évaluation pourrait devenir quasi-continue grâce à la collecte massive de données en temps réel (commentaires usagers, données opérationnelles). Il cite une phrase fascinante et effrayante : «L'abondance des données permet aux ordinateurs de comprendre le monde d'une manière inaccessible à l'humain». Cela implique pour les évaluateurs d'apprendre à traiter des données non conventionnelles.

Temps 3 : l'évaluation automatisée (le futuriste)

Dans un scénario où les administrations déploieraient des systèmes d'IA autonomes pour délivrer des services publics (robots, feux de signalisation intelligents, systèmes d'octroi de droits), l'évaluation serait intégrée directement dans l'algorithme. Les machines apprendraient les unes des autres et s'ajusteraient en temps réel sans intervention humaine. Bien que cela semble farfelu aujourd'hui, c'est une réflexion nécessaire sur l'avenir de la fonction évaluative.

CONCLUSION

Steve Jacob conclut son propos par une mise en garde solennelle sur les biais algorithmiques. Contrairement à une calculatrice qui donne un résultat certifié à 100%, aucun système d'IA n'est infallible. À chaque étape du cycle de vie de l'IA, des décisions humaines sont prises qui influencent le résultat. Sans une culture technique minimale pour auditer et questionner ces systèmes, les décideurs et évaluateurs risquent de naviguer à l'aveugle.

François Mouterde clôture la table ronde en soulignant la clarté de cette trilogie prospective (Augmentée, Intégrée, Automatisée), qui offre une feuille de route structurante pour les travaux à venir de la communauté de l'évaluation.





REMERCIEMENTS ET ENGAGEMENTS

La publication de ces actes est l'occasion de renouveler les remerciements exprimés lors de l'ouverture envers celles et ceux qui ont rendu cet événement possible.

Nous saluons l'engagement du CESE, en particulier de Thierry Cadart et Fabrice Bergeron, pour l'accueil au Palais d'Iéna et la qualité du partenariat avec la SFE. Nous remercions chaleureusement les bénévoles et organisateurs, notamment François Mouterde, Alix de Saint-Albin, ainsi que les administrateurs de la SFE qui donnent de leur temps personnel pour faire vivre notre communauté professionnelle.

Une reconnaissance particulière est adressée aux dix intervenant·e·s des tables rondes, dont certain·e·s ont fait le déplacement depuis l'Italie, le Canada ou les États-Unis pour partager leur expertise internationale. Nous remercions également le jury du prix de master (Anne Revillard, Agathe Devaux, Gaëlle Barron) et le réseau des jeunes évaluateurs pour leur investissement.

Enfin, ces actes sont dédiés à l'ensemble des participant·e·s présent·e·s ce jour-là. Notre responsabilité collective, réaffirmée en conclusion de l'ouverture, est de garantir que l'intelligence artificielle serve le bien commun. Il nous appartient de veiller à ce qu'elle contribue à un avenir plus juste et plus équitable, en renforçant la précision, la transparence et l'utilité sociale de l'évaluation des politiques publiques.

Puisse la lecture de ces actes nourrir vos réflexions, inspirer vos pratiques et renforcer votre vigilance éthique face aux transformations en cours.



INTERVENANT·E-S ET MODÉRATION

Thierry Beaudet, Président du Cese

Sandrine Beaujean, Responsable du département « Renforcement de capacités et Évaluation » au sein de l'agence Artemis Information Management

Virginie Besrest, Présidente de la SFE

Marianne Tordeux Bitker, Rapporteuse CESE

Thierry Cadart, Questeur du CESE

Silva Ferretti, Consultante indépendante en évaluation des programmes de développement

Steve Jacob, Professeur au Département de science politique de l'Université Laval

François Mouterde, Président d'honneur de la SFE

Estelle Raimondo, Responsable du service des méthodes d'évaluation à la Banque mondiale

Alix de Saint-Albin, Administrateur de la SFE

Aurélié Simard, INRIA, Executive Director of the Paris expert center of the GPAI

Guillaume Soto, Ingénieur télécom et expert indépendant en analyse de données et IA pour l'évaluation

Fabienne Tatot, Rapporteuse CESE

Jean-Marie Truffat, Rapporteur CESE

Gilles Vermot Desroches, Rapporteur CESE

Cette synthèse a été produite par un outil d'intelligence artificielle - NotebookLM, à partir des enregistrements vidéo des différents panels qui se sont tenus dans le cadre des Rendez-vous de l'évaluation 2025, organisés par le CESE et la SFE le 9 janvier 2025.



