Réaliser la première évaluation des risques liés à l'IA de votre OSC

Un guide pratique et un recueil d'études de cas destinés aux organisations de la société civile

Introduction : Le travail d'enquête commence

Si vous avez suivi cette série, vous avez déjà accompli un travail considérable. Vous avez défini vos principes en matière d'IA responsable, c'est-à-dire les valeurs que défend votre organisation. Vous avez constitué votre Arbre, c'est-à-dire le groupe de personnes chargées d'animer le débat sur la gouvernance. Vous avez rédigé votre politique en matière d'IA, ces accords collaboratifs qui guident les décisions quotidiennes, avec leur système de feux verts, jaunes et rouges.

Dans la méthodologie présentée à l'article 1, cela correspond à la phase **« CONSTRUIRE »** , au cours de laquelle les principes se concrétisent dans la pratique grâce à l'évaluation des risques, aux stratégies d'atténuation et aux contrôles de gouvernance continus. Vous avez terminé les étapes « REGARDER » (mettre en lumière les tensions, rassembler les connaissances) et « CRÉER » (définir les principes, établir votre « Arbre », rédiger votre politique). C'est maintenant l'heure de mener l'enquête.

L'IA n'est pas seulement une question technique ; c'est une question sociotechnique. Son impact dépend entièrement du contexte dans lequel elle s'inscrit : les personnes, les inégalités existantes et les décisions sur lesquelles elle exerce une influence. Une évaluation des risques consiste avant tout à prendre le temps de s'arrêter et de réfléchir aux conséquences potentielles. Il ne s'agit pas de trouver une seule « bonne » réponse. Il s'agit d'un processus continu et itératif consistant à se poser, en équipe, des questions critiques afin de garantir une action responsable.

L'évaluation des risques n'est pas une invention de l'ère de l'IA. Ses origines remontent au secteur financier, où les banques et les assureurs ont mis au point des méthodes systématiques pour évaluer l'incertitude avant d'engager des ressources. Au fil des décennies, cette pratique s'est étendue aux domaines de la santé, de la protection de l'environnement et de l'ingénierie, partout où les décisions ont des conséquences pour autrui. Ce qui relie tous ces éléments, c'est une attitude fondamentalement *prudente face au risque* : la prise de conscience qu'avant d'agir, nous avons la responsabilité de réfléchir à l'impact de nos actes sur les autres.

Cette approche trouve ses origines dans des traditions bien antérieures aux cadres proposés par l'industrie. Par exemple, *le « destur »* : une pratique issue de la culture turque qui consiste à s'arrêter un instant, à prendre conscience de son environnement et à demander la permission avant d'entrer dans un espace qui ne nous appartient pas. Dans le meilleur des cas, l'évaluation

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

des risques est une forme de « destur » : une véritable pause pour se demander : *« Devons-nous pénétrer dans cet espace ? » Qu'est-ce qu'on apporte ? Qu'est-ce que nous risquons de perturber ?* Les cadres et les grilles d'évaluation présentés ci-après ne sont rien d'autre que la structure qui confère une certaine rigueur à cette intuition

Ce guide vous accompagne pas à pas dans la réalisation d'une analyse des risques. À l'issue de ce processus, une évaluation complète portant sur un cas d'utilisation de l'IA sera prête, avec la liste des parties prenantes identifiées, une évaluation des préjudices, des stratégies d'atténuation et un registre des risques mis à jour en continu.

Quel scénario s'applique ?

- **Adoption d'un nouvel outil :** Vous envisagez de mettre en place un système d'IA et souhaitez l'évaluer avant son déploiement.
- **Vous utilisez déjà l'IA :** Des outils ont été adoptés depuis des mois sans avoir fait l'objet d'une évaluation officielle. Une analyse rétrospective n'est pas seulement légitime, c'est aussi le point de départ le plus honnête.
- **Système externe ayant une incidence sur la communauté :** Une agence ou une institution gouvernementale utilise un algorithme qui a des répercussions sur les personnes prises en charge par une OSC. Il faut l'examiner de manière critique depuis l'extérieur.

Les étapes ci-dessous s'appliquent aux trois cas de figure. Lorsque la procédure diffère, notamment en matière de plaidoyer et de contrôle externe, les recommandations spécifiques sont signalées par la mention **« Pour les évaluations en matière de plaidoyer ».** **nts."**

Comment utiliser cette boîte à outils

Ce document contient :

1. **Un modèle d'évaluation des risques et des impacts (parties A, B, C) :** Tableaux vierges accompagnés de questions guides à remplir pour un système d'IA spécifique.
2. **Trois exemples concrets :** À la fin de chaque section, vous trouverez des exemples illustrant trois cas concrets : un déploiement interne au sein d'une OSC, un système éducatif et une action de plaidoyer externe, afin de vous aider à vous lancer.
3. **Annexes :** Guides de référence permettant d'identifier les préjudices, leurs causes et les principales ressources disponibles.

Suivez les étapes, répondez aux questions relatives au système concerné et inspirez-vous des exemples proposés. L'objectif est de commencer modestement, de construire la compréhension et d'en faire un dialogue permanent.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

techsoup Digital Activism Program with support from Google.org

Modèle d'évaluation des risques et des impacts

Partie A : Contexte et champ d'application (fondements)

Étape 1 : Constituer l'équipe et examiner la documentation

L'évaluation des risques n'est jamais une démarche individuelle. Lorsque 70 responsables issus de 30 ONG spécialisées dans l'éducation ont étudié le système d'alerte précoce contre le décrochage scolaire du Wisconsin lors d'un atelier animé par Tech To The Rescue, ils ont élaboré une refonte complète du système en 30 minutes. La magie ne résidait pas dans l'expertise, mais dans la diversité des points de vue.

1. Date de l'évaluation et responsable : (Indiquez la date du jour et le nom de la personne à l'origine de cette évaluation, généralement un membre de l'« Arbre » ou la personne qui a signalé le feu orange ayant déclenché cet examen.)

2. Équipe d'évaluation (interfonctionnelle) : (Énumérez les noms et les fonctions. Visez une diversité de compétences : des membres de l'équipe du programme travaillant directement avec les bénéficiaires, des représentants de la communauté ou des personnes ayant une expérience de terrain, une personne disposant de connaissances techniques (même élémentaires), d'une expertise juridique ou en matière de conformité, ainsi qu'un décideur issu de la direction.)

Pour les évaluations des actions de plaidoyer : Associer des personnes directement concernées par le système externe : elles perçoivent ce qui échappe aux développeurs et aux responsables gouvernementaux.

Équipe minimale viable : 3 à 5 personnes. Même à deux, il est possible de réaliser une première évaluation pertinente à l'aide du guide de démarrage rapide figurant à la fin de ce manuel.

Processus de travail recommandé : Un membre de l'équipe, généralement celui qui a lancé l'évaluation, prépare à l'avance le contexte CIDA (étape 2) et une première cartographie des parties prenantes (étape 3). Cette personne présente ses conclusions à l'ensemble de l'équipe, qui procède ensuite à une réflexion collective sur les préjudices potentiels, évalue les risques et élabore ensemble des mesures d'atténuation (étapes 4 à 7). Le travail préparatoire permet de s'assurer que la discussion repose sur des faits ; la séance collaborative permet de bénéficier de points de vue variés. Le travail d'enquête comporte deux phases : *l'investigation* (qui peut être menée par une ou deux personnes) et *l'interprétation* (qui doit être effectuée à plusieurs).

3. Documents examinés : (Énumérer tous les documents de référence consultés : documentation du système, articles de presse, études de validation, politiques de confidentialité, fiches de modèle, rapports d'incident, témoignages de la communauté.)

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Étape 2 : Définir le contexte sociotechnique (C-I-D-A)

Les scénarios « feu orange » prévus par la politique en matière d'IA constituent précisément le point de départ. La politique indique *quand* il faut faire une pause. L'évaluation des risques vous dit *comment* approfondir votre analyse. Le cadre CIDA : contexte, entrée, décision et action, donne une structure à l'enquête.

Répondez aux questions suivantes pour vous faire une idée globale de la situation :

CONTEXTE (« pourquoi » et « où »)

- Objectif : Pourquoi l'IA est-elle nécessaire dans ce cas de figure ? Quels sont les avantages escomptés ?
- Environnement : Quel est le contexte social et politique ? Avec quels déséquilibres de pouvoir ou quelles inégalités existants ce système pourrait-il interagir ?
- Parties prenantes : Qui sont toutes les parties prenantes, directes et indirectes ? (Ce point est abordé plus en détail à l'étape 3.)
- Préoccupations : Quelles sont les principales préoccupations ou les problèmes connus liés au système, tels qu'ils sont signalés par les utilisateurs ?

ENTRÉES (« quoi »)

- Sources des données : Comment les données d'entraînement ont-elles été collectées ? Ces données proviennent-elles d'un tiers ou sont-elles collectées en interne ?
- Contenu des données : Le modèle utilise-t-il directement des données démographiques ? Quelles sont les limites connues de ces données ?
- Documentation : Existe-t-il une documentation relative à cet ensemble de données (par ex. une fiche technique) ?

DÉCISION (le « comment » de l'algorithme)

- Type d'algorithme : S'agit-il d'un apprentissage supervisé, d'un apprentissage profond ou d'un système basé sur des règles ? Utilise-t-on des composants tiers ?
- Niveau d'autonomie : Le système est-il entièrement automatisé ou sert-il d'aide à un opérateur humain ? Un être humain peut-il facilement passer outre ?
- Interprétabilité : Est-ce difficile d'expliquer une décision ? Une explication est-elle fournie à l'utilisateur ?
- Enjeux : Quelles sont les conséquences d'une mauvaise décision ? Dans quelle mesure est-il difficile d'en inverser les effets ?

ACTION (le « so what? »)

- Interaction avec l'utilisateur : Qui sont les utilisateurs directs ? Quelle formation reçoivent-ils ?

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

- Sortie du système : Quels sont les résultats concrets que les utilisateurs voient, et comment les utilisent-ils ?
- Conséquences : Quelles mesures concrètes sont prises à partir des résultats obtenus, et quels en sont les effets en aval ?

Pour les évaluations des actions de plaidoyer : Ajoutez ces questions :

- Qui a commandé ou financé ce système ?
- Quelles sont les dynamiques de pouvoir à l'œuvre entre l'organisme chargé du déploiement et les communautés concernées ?
- Quelles informations sont accessibles au public concernant le fonctionnement du système ? Qu'est-ce qui est opaque ?

Le CIDA en action : Trois exemples illustratifs

Ces exemples sont donnés à titre indicatif et ne sont pas exhaustifs. Répondre aux questions relatives à un système spécifique permettra d'acquérir une compréhension approfondie. Lancez-vous dans l'aventure !

Recommandations et questions	**Exemple interne à une OSC : Chatbot d'aide juridique basé sur l'IA (cas fictif)**	**Exemple dans le domaine de l'éducation : Système d'alerte précoce contre le décrochage scolaire (DEWS)**	**Exemple d'action de plaidoyer menée par une OSC : Le système espagnol VioGén**
CONTEXTE (pourquoi/où) : Pourquoi utilise-t-on l'IA ? Quel est l'environnement social ?	Objectif : Pour trier les demandes reçues de personnes sollicitant une aide juridique et traiter plus rapidement les dossiers urgents. Environnement : OSC spécialisée dans la prestation de services, intervenant au sein d'une communauté multilingue où le niveau de maîtrise du numérique varie.	Objectif : Identifier les élèves exposés au risque de décrochage scolaire afin de leur apporter un soutien en temps utile. Environnement : District scolaire où l'on constate un écart de taux de réussite au baccalauréat entre les différents groupes ethniques.	Objectif : Évaluer le risque de récidive en matière de violences sexistes et déterminer les niveaux de protection policière. Environnement : L'Espagne, où la violence à caractère sexiste constitue une crise nationale et où les moyens dont dispose la police sont limités.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

ENTRÉES (données « quoi ») : Quelles sont les données utilisées ? Sont-elles subjectives ou fondées sur des indicateurs indirects ?	Texte rédigé par un usager, décrivant sa situation juridique. La langue, le vocabulaire et la maîtrise du numérique peuvent influencer la manière dont le système interprète le caractère urgent d'une situation.	Résultats aux examens, assiduité, antécédents disciplinaires (subjectifs) et données démographiques, notamment origine ethnique (indicateur).	35 questions auxquelles le responsable du signalement a répondu concernant la victime et l'agresseur, notamment l'historique de leur relation, les incidents antérieurs et les indicateurs de risque.
DÉCISION (comment) : Quel est le résultat généré par l'IA ? Comment est-il interprété ?	Établit un niveau d'urgence (élevé/moyen/faible) et transmet la demande au personnel compétent. Le personnel peut se fonder sur cette classification sans procéder à un examen indépendant.	Génère un score de risque « élevé », « moyen » ou « faible ». Les enseignants peuvent interpréter l'expression « à haut risque » comme signifiant « élève à problèmes ».	Évalue le niveau de risque sur une échelle allant de 1 (négligeable) à 5 (extrême). Les agents de police acceptent la classification établie par l'algorithme dans 95 % des cas (audit de la Fondation Eticas).
ACTION (le "So what?") : Quelles mesures sont prises ? Quels en sont les effets ?	Les cas très urgents font l'objet d'un rappel immédiat ; les cas peu urgents sont placés en file d'attente. Impact : Une personne en situation de crise réelle qui ne parvient pas à s'exprimer clairement risque de ne pas être considérée comme une priorité.	Les élèves considérés comme « à haut risque » font l'objet d'une intervention ciblée. Impact : Stigmatisation, renforcement des stéréotypes, baisse de l'estime de soi.	Le niveau de risque détermine la fréquence des contrôles effectués par la police et permet de déterminer s'il convient de recommander la délivrance d'ordonnances de protection. Impact : Les femmes classées dans les catégories de risque les plus faibles bénéficient d'une protection moindre, ce qui peut avoir des conséquences potentiellement mortelles.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Étape 3 : Identifier les parties prenantes

Cartographie des parties prenantes : Réfléchissez ensemble à toutes les personnes et à tous les groupes concernés. Pensez d'abord à ceux qui ont le moins de pouvoir et tenez compte de l'intersectionnalité.

- **Utilisateurs directs :**
- **Personnes directement concernées :**
- **Groupes indirectement concernés :**
- **Parties prenantes institutionnelles :**

Pour les évaluations des actions de plaidoyer : La communauté *constitue* le principal groupe de parties prenantes. Les organisations de la société civile disposent souvent déjà de relations et d'un capital de confiance qui leur permettent de mettre en lumière ce que les systèmes automatisés ne détectent pas. S'appuyer sur les consultations, les enquêtes, les retours d'expérience de la communauté et les observations du personnel de terrain déjà existants.

Stakeholder Mapping Examples

Catégorie de parties prenantes	Interne à l'OSC : Chatbot d'aide juridique basé sur l'IA	Formation : DEWS	Plaidoyer des OSC : VioGén
Utilisateurs directs	Assistants sociaux, personnel chargé de l'accueil	Enseignants, responsables d'établissement, conseillers d'orientation	Agents de police, autorités judiciaires
Personnes directement concernées	Personnes sollicitant une aide juridique	Élèves, parents	Femmes reportant des violences sexistes
Groupes indirectement concernés	Familles des personnes sollicitant l'aide juridictionnelle, associations locales orientant les personnes vers ce service	Communauté locale, futurs employeurs	Enfants des victimes, familles, ensemble de la communauté des survivantes
Acteurs institutionnels	L'OSC elle-même, bailleurs de fonds, partenaires juridiques	Circonscription scolaire, ministère de l'Éducation, syndicat des enseignants	Ministère de l'Intérieur, pouvoir judiciaire, organisations de la société civile œuvrant dans le domaine de la violence sexiste

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Partie B : Identification et analyse des risques (le cœur du travail d'enquête)

Étape 4 : Réfléchissez aux préjudices potentiels et à leurs causes

Remarque sur le travail d'équipe : Cette étape est d'autant plus efficace qu'elle est réalisée de manière collaborative. La diversité des points de vue est essentielle pour mettre en lumière tout un éventail de préjudices potentiels qu'une seule personne pourrait ne pas percevoir. Encouragez les séances de brainstorming ouvertes, sans jugement.

Pour chaque groupe de parties prenantes, identifiez les **préjudices** potentiels et établissez un lien entre ceux-ci et leurs **causes**. Voir **l'annexe A : Guides d'identification des préjudices et des causes** pour trouver des idées.

Identifiant du préjudice	Description des préjudices (quelles conséquences négatives pourraient se produire ?)	Partie(s) prenante(s) concernée(s)	Cause / Indicateur système (pourquoi cela se produirait-il ?)
H-01			
H-02			
H-03			

Étape 5 : Hiérarchiser les risques

Le risque est la combinaison de **la probabilité** qu'un préjudice se produise et de son **ampleur**. Cela permet de canaliser les efforts de manière structurée.

- **Probabilité (quelle est la probabilité ?) :** 1 = Peu probable, 2 = Possible, 3 = Probable, 4 = Presque certain
- **Ampleur (quelle est la gravité ?) :** 1 = mineur, 2 = modéré, 3 = grave, 4 = très grave

Probabilité : *Compte tenu de ce que nous savons de ce système, de ses utilisateurs et de son contexte, à quelle fréquence ce préjudice pourrait-il réellement se produire ?* **Ampleur** : *Si cela venait à se produire, dans quelle mesure cela affecterait-il les personnes concernées, compte tenu à la fois de la gravité des répercussions sur un individu et du nombre de personnes susceptibles d'être touchées ?*

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Identifiant du préjudice	Probabilité (1–4)	Ampleur (1–4)	Score de risque (P × A)	Justification / Informations manquantes
H-01				
H-02				

Légende du score de risque :

Score de risque	Niveau	Action requise
1–4	Faible	À suivre
5-8	Moyen	Atténuer les risques lorsque cela est possible
9-12	Élevé	Atténuer les risques avant de poursuivre
13-16	Critique	S'arrêter ou repenser le projet de fond en comble

Exemple illustratif : De l'évaluation des préjudices au score de risque (étapes 4 à 5)

Recommandations	Interne à l'OSC : Chatbot d'aide juridique basé sur l'IA	Formation : DEWS	Plaidoyer des OSC : VioGén
Partie prenante concernée	Personnes sollicitant une aide juridique (en particulier celles ayant une faible maîtrise du numérique ou des compétences linguistiques limitées)	Élèves (en particulier ceux issus de minorités)	Femmes reportant des violences sexistes
Description du préjudice	Des personnes en situation de crise réelle sont classées comme « faible urgence » car elles ne parviennent pas à décrire clairement leur situation par écrit, ce qui entraîne un retard dans l'aide qui leur est apportée.	Des élèves sont injustement qualifiés de « à haut risque », ce qui entraîne une stigmatisation et une baisse de l'estime de soi.	Des femmes sont classées dans la catégorie des personnes à faible risque, bien qu'elles soient exposées à un danger grave et ne bénéficient pas d'une protection policière suffisante.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Cause / Indicateur système	Biais : Le système offre de meilleures performances aux utilisateurs ayant un niveau d'alphabétisation plus élevé. Efficacité : Erreur de classification du degré d'urgence.	Biais : Le modèle utilise l'origine ethnique / le revenu comme variables. Transparence : Manque d'explicabilité.	Efficacité : L'algorithme sous-estime le risque dans certains profils de cas spécifiques. Interaction homme-machine : Dans 95 % des cas, la police se fie à ce résultat.
Probabilité (1–4)	3 (Probable) La diversité des niveaux d'alphabétisation est une réalité bien connue au sein de la communauté accompagnée.	3 (Probable) Confirmé par une analyse externe.	4 (Presque certain) D'après les informations fournies par la Fondation Eticas et la BBC. Entre 2003 et 2021, 71 femmes prises en charge par le système ont été tuées ; toutes avaient été classées dans la catégorie de risque « négligeable » ou « moyen ».
Ampleur (1–4)	3 (Grave) Un retard dans l'octroi de l'aide juridique peut entraîner la perte du logement, la perte de la garde des enfants ou une atteinte à la sécurité.	3 (Grave) Conséquences graves sur le bien-être des élèves et leurs perspectives d'avenir.	4 (Grave) Conséquences mortelles avérées.
Score de risque (P × A)	9 (Élevé)	9 (Élevé)	16 (Critique)

Partie C : Mesures d'atténuation et de suivi

Étapes 6 et 7 : Définir des stratégies d'atténuation et gérer le risque résiduel

Pour chaque risque identifié, définissez une **stratégie d'atténuation**, c'est-à-dire une mesure concrète visant à réduire la probabilité ou l'impact de ce risque. Désignez un **responsable** chargé de sa mise en œuvre, fixez une **date limite** et identifiez les **risques résiduels** qui pourraient subsister.

Ces catégories de mesures d'atténuation sont conformes aux exigences de la loi européenne sur l'IA relatives aux systèmes à haut risque : supervision humaine (art. 14), transparence (art. 13), qualité des données (art. 10) et non-discrimination.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Identifiant du préjudice	Stratégie d'atténuation (quelles mesures concrètes ?)	Propriétaire (qui est responsable ?)	Date limite	Risque résiduel (que reste-t-il après les mesures d'atténuation ?)
H-01				
H-02				

Le risque résiduel est une réalité, et non un échec. Même après la mise en œuvre des mesures d'atténuation, un certain risque subsiste. Pour chaque risque résiduel, définir :

- Quels mécanismes de gouvernance de l'IA seront mis en place ?
- Comment cela sera-t-il contrôlé ?
- Qu'est-ce qui déclenche une réévaluation ?

Pour les évaluations des actions de plaidoyer : Les mesures d'atténuation prennent différentes formes ; ce sont des exigences et des actions :

- Recommandations publiques à l'intention du responsable du déploiement du système
- Plaintes ou requêtes en matière de réglementation
- Mise en place de partenariats avec d'autres OSC, des journalistes et des chercheurs
- Mobilisation des médias pour sensibiliser le public
- Soutien direct à la communauté et sensibilisation
- Recours judiciaires (comme l'a démontré la coalition SyRI aux Pays-Bas, où six organisations de défense des droits de l'homme et un syndicat ont contesté avec succès devant les tribunaux un algorithme discriminatoire utilisé dans le domaine de l'aide sociale)

Exemple illustratif : Atténuation et risque résiduel (étapes 6 à 7)

Recommandations	Interne à l'OSC : Chatbot d'aide juridique basé sur l'IA	Formation : DEWS	Plaidoyer des OSC : VioGén
Stratégie d'atténuation	Prévoir une procédure manuelle d'enregistrement pour toute personne qui appelle ou se présente sur place. Prévoir un examen par un intervenant humain pour tous les cas avant la classification finale du degré d'urgence. Tester les performances du	Supprimer les données démographiques du modèle. Remplacer la mention « risque » par une suggestion indiquant « a besoin d'aide », accompagnée de ressources spécifiques.	Publier les résultats. Exiger un audit indépendant avec un accès complet aux données. Plaider en faveur de protocoles obligatoires de reprise en main par un opérateur humain. Soutenir les victimes par l'intermédiaire d'organisations

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

	chatbot pour différents groupes linguistiques.		partenaires telles que la Fondation Ana Bella.
Responsable et date limite	Le directeur de programme ; avant le déploiement	L'équipe de science des données ; 3e trimestre	L'équipe de plaidoyer de l'OSC et ses partenaires de coalition ; campagne en cours
Risque résiduel	Il se peut que le chatbot continue à mal interpréter le degré d'urgence pour certains utilisateurs. Nécessite un suivi continu de la précision de la classification pour l'ensemble des groupes démographiques.	Le modèle peut tout de même trouver des indicateurs indirects des données démographiques dans d'autres données. Nécessite un suivi continu afin de détecter d'éventuels impacts disproportionnés.	Le ministère de l'Intérieur peut refuser de coopérer. L'algorithme reste en service tant que le plaidoyer se poursuit. Il est essentiel que la communauté continue à apporter son soutien et à recenser les cas.

Étape 8 : Documenter et relier

Remplir le document d'évaluation des risques. Relier ensuite ses conclusions à l'écosystème de gouvernance :

- **Mises à jour de la politique :** De nouvelles « lignes rouges » ont-elles été découvertes ? Des feux orange peuvent-ils passer au vert, ou doivent-ils passer au rouge ? Y a-t-il de nouveaux engagements en matière de suivi ?
- **Priorités de l'Arbre :** Quels sont les éléments que le groupe de gouvernance doit surveiller, escalader ou examiner ?
- **Date de l'examen :** Fixer une date précise pour revenir sur cette évaluation. Les conditions changent ; les outils évoluent, les populations se déplacent, les réglementations sont modifiées.

Pour les évaluations des actions de plaidoyer : Le livrable pourrait être :

- Un rapport ou une note d'information destinés au grand public
- Une lettre adressée aux autorités de régulation ou aux décideurs politiques
- Des supports pédagogiques destinés à la communauté
- Des éléments probants à l'appui d'une campagne de coalition
- Une saisine dans le cadre des mécanismes de plainte prévus par la loi européenne sur l'IA

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Conclusion	Action requise	Responsable	Date limite

Adapter ce processus au plaidoyer externe

Le cadre et les méthodes d'évaluation CIDA décrits ci-dessus s'appliquent aussi bien lorsque l'OSC est responsable de la mise en œuvre que lorsqu'elle agit en tant qu'organisme de surveillance. Mais lorsqu'on examine de près un système qui n'est pas sous contrôle direct, trois éléments changent :

Accès à l'information : Il est peu probable qu'il y ait une transparence totale. Les preuves proviennent de demandes d'accès à l'information, de témoignages communautaires, de recherches académiques, de journalisme d'investigation et de bases de données d'incidents, et non de documents internes. C'est plus difficile, mais pas impossible. LightHouse Reports a mené une enquête sur l'algorithme suédois de détection des fraudes aux prestations sociales en analysant un ensemble de données datant de 2017 et en interrogeant les familles concernées. Eticas a procédé à un audit du système espagnol VioGén en s'appuyant sur des données judiciaires accessibles au public et sur des entretiens avec des survivantes, après que le ministère de l'Intérieur a refusé de coopérer

Dynamiques de pouvoir : L'objectif passe de « comment atténuer ce problème ? » à « que demandons-nous, et à qui ? » Le registre des risques sert de base factuelle pour les actions de plaidoyer, et non de plan d'action interne.

Résultats : Plutôt que des mesures de gouvernance, les résultats attendus prennent la forme d'évaluations publiques, de dossiers réglementaires, de lettres ouvertes, d'actions de sensibilisation de la population ou de recours juridiques. L'affaire SyRI aux Pays-Bas reste un exemple emblématique de ce que peut accomplir le travail de surveillance mené par la société civile.

Coalitions : Ce travail se fait rarement seul. La collaboration avec d'autres OSC, des journalistes, des chercheurs et des autorités de régulation permet d'amplifier l'impact. Les initiatives les plus efficaces en matière de contrôle externe de l'IA ont été le fruit d'une collaboration au sein de la société civile.

Mise à l'échelle : À quoi cela ressemble selon différents niveaux de ressources

Ressources	Format	Résultats produits

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

2 personnes, 1 heure	Discussion à deux en utilisant les questions du guide de démarrage rapide	Synthèse des risques d'une page pour un cas d'utilisation
3 à 5 personnes, 2 heures	Séance animée couvrant les étapes 1 à 7	Évaluation préliminaire avec notation des préjudices
5 à 8 personnes, une demi-journée	Atelier complet avec travail en petits groupes	Évaluation complète prête pour l'Arbre
Inter-organisations	Séance de collaboration entre plusieurs OSC	Évaluation conjointe et résultats des actions de plaidoyer menées par la coalition

Commencez là où vous en êtes. Une discussion d'une heure entre deux personnes a infiniment plus de valeur qu'un processus parfait qui ne se concrétise jamais.

Démarrage rapide : Si vous ne disposez que de 30 minutes

1. **Choisissez un scénario de « feu orange** » figurant dans la politique — ou un système d'IA externe ayant un impact sur la communauté.
2. **Listez les personnes concernées** (3 à 5 groupes de parties prenantes). .
3. **Citez trois problèmes qui pourraient survenir** pour le groupe le plus vulnérable.
4. **Notez-les :** Quelle est la probabilité ? Quelle est la gravité ?
5. **Notez une mesure d'atténuation** pour l'élément présentant le risque le plus élevé — ou **une demande** à adresser au responsable du système.

Voilà la première évaluation des risques. Tout le reste découle de là.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Annexe A : Guides d'identification des préjudices et de leurs causes

Cette annexe propose deux guides destinés à faciliter le brainstorming. Tout d'abord, une liste des **préjudices** potentiels (les effets négatifs). Deuxièmement, une liste des **causes** potentielles (les caractéristiques ou indicateurs du système susceptibles d'entraîner ces préjudices).

Guide n° 1 : Préjudices potentiels

(Adapté du cadre de modélisation des préjudices de Microsoft et du Future of Privacy Forum)

Catégorie de préjudice	Description et questions guides
Risque de préjudice (émotionnel/psychologique)	La technologie peut être source de détresse émotionnelle, nuire à l'estime de soi ou porter atteinte à la réputation. En quoi ce système pourrait-il amener quelqu'un à se sentir stigmatisé, anxieux ou mal représenté ?
Refus de services essentiels (perte d'opportunités et pertes économiques)	Les décisions automatisées peuvent limiter de manière injuste l'accès aux ressources et le bien-être économique. Ce système pourrait-il priver quelqu'un d'un emploi, d'un logement, d'une formation, d'allocations ou d'une autre opportunité essentielle ?
Atteintes aux droits de l'homme (dignité, vie privée, liberté, impact environnemental)	La technologie peut porter atteinte aux droits fondamentaux. En quoi ce système pourrait-il divulguer des informations privées, contrôler les comportements ou restreindre l'autonomie des individus ?
Érosion des structures sociales et démocratiques (manipulation, préjudice social)	À grande échelle, la technologie peut renforcer les stéréotypes et aggraver les inégalités. Ce système pourrait-il perpétuer des préjugés, diffuser de fausses informations ou affaiblir la cohésion sociale ?

Pour un brainstorming plus approfondi, consultez cette liste détaillée des types de préjudices :

- **Perte d'opportunité :** Se voir refuser l'accès à des services, à des prestations, à l'éducation ou à l'emploi
- **Perte de dignité :** Être traité comme un numéro, être étiqueté, stigmatisé ou déshumanisé

 Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

- **Perte de liberté :** Restrictions à la liberté de circulation, d'association ou d'expression
- **Préjudice économique :** Pertes financières, prestations refusées, tarification abusive
- **Atteinte à la vie privée :** Collecte de données non autorisée, surveillance, divulgation d'informations personnelles
- **Érosion de la confiance :** Perte de confiance dans les institutions, les services ou la technologie
- **Renforcement des stéréotypes :** Renforcement des préjugés existants liés au genre, à l'origine ethnique, à la classe sociale ou à d'autres caractéristiques
- **Amplification des inégalités de pouvoir :** Élargissement du fossé entre ceux qui contrôlent les systèmes d'IA et ceux qui y sont soumis
- **Déshumanisation :** Suppression ou réduction du jugement humain, de l'empathie et de la discrétion dans les décisions ayant une incidence sur la vie des gens
- **Perte de compétences :** Dépendance excessive à l'automatisation entraînant une perte du jugement professionnel
- **Absence de recours effectif :** Impossibilité de savoir si vous êtes concerné, impossibilité de contester, impossibilité de faire appel

Guide n° 2 : Causes possibles / Indicateurs système

(Adapté du Cadre de gestion des risques liés à l'IA du NIST et des recherches de BABL AI Audit des algorithmes : évaluer les algorithmes qui nous évaluent)

Catégorie de cause	Description et questions guides
Biais	Le système aboutit systématiquement à des résultats injustes ou discriminatoires. Les données d'entraînement reflètent-elles des biais historiques ? Ce modèle est-il plus précis pour un groupe que pour un autre ?
Efficacité	Le système n'est ni suffisamment précis, ni suffisamment fiable, ni suffisamment robuste pour remplir sa mission. À quelle fréquence le système commet-il des erreurs (faux positifs/faux négatifs) ? Est-ce qu'on peut facilement le tromper ou le casser ?
Transparence	Il est difficile de comprendre comment fonctionne le système ou comment ses décisions sont prises. S'agit-il d'une « boîte noire » ? Les utilisateurs savent-ils qu'une IA est utilisée ? Peut-on expliquer un résultat ?
Sécurité et confidentialité	Le système présente des failles de sécurité susceptibles d'entraîner des fuites de données ou une utilisation abusive. Ce système pourrait-il être piraté ? Collecte-t-il plus de données que nécessaire ?
Interaction homme-machine	La manière dont les utilisateurs se servent du système comporte des risques. Les utilisateurs font-ils aveuglément confiance à l'IA (biais d'automatisation) ? L'interface utilisateur prête-t-elle à confusion ?

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Ces catégories de causes déterminent également la stratégie d'évaluation de l'IA à venir. Les causes identifiées ici permettront de déterminer les indicateurs à suivre, les tests techniques à réaliser et les mesures de gouvernance de l'IA à mettre en place dans le cadre des mesures d'atténuation et de la surveillance continue. En d'autres termes : comprendre *les causes* des préjudices est essentiel pour déterminer *comment les détecter*.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Annexe B : Utilisation des bases de données d'incidents liés à l'IA

Avant de vous lancer dans une réflexion à partir de zéro, vérifiez ce qui a déjà posé problème avec des systèmes similaires. Ces bases de données correspondent à des dossiers d'affaires constitués avant le début de l'enquête :

- **Base de données sur les incidents liés à l'IA**: archives consultables répertoriant les préjudices réels causés par l'IA
- **Observatoire des incidents liés à l'IA de l'OCDE**: suivi des incidents axé sur les politiques publiques
- **Répertoire des risques liés à l'IA**, base de données classée par catégories des risques connus liés à l'IA

Recherchez des systèmes similaires à celui qui fait l'objet de l'évaluation. Les tendances mises en évidence permettront d'affiner l'évaluation.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.

Annexe C : Ressources

Cadres et normes

- Guide pratique du cadre de gestion des risques liés à l'IA du NIST
- Règlement européen sur l'IA ; classification des systèmes à haut risque et mesures de contrôle requises
- Règlement européen sur l'IA ; classification des systèmes à haut risque et mesures de contrôle requises

Journalisme d'investigation et études de cas

- Rapports LightHouse : Enquêtes menées par un système de suspicion automatisé (Suède, France, Pays-Bas)
- Human Rights Watch : « Négligence automatisée » : le programme de transferts monétaires Takaful en Jordanie
- Fondation Eticas : Audit contradictoire du système espagnol VioGén

---

Cet article fait partie de la série intitulée « Feuille de route pour une IA responsable pour les OSC », élaborée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup, avec le soutien de Google.org. Toutes les ressources sont publiées sous la licence Creative Commons Attribution 4.0 International.

A propos de l'auteur

Ayşegül Güzel est une architecte en gouvernance de l'IA responsable qui aide les organisations axées sur une mission à transformer l'anxiété liée à l'IA en systèmes dignes de confiance. Sa carrière combine le leadership social exécutif — notamment en tant que fondatrice de Zumbara, le plus grand réseau mondial de banques de temps — et la pratique technique de l'IA, en tant qu'auditrice certifiée en IA et ancienne data scientist. Elle guide les organisations à travers des transformations complètes de la gouvernance de l'IA et réalise des audits techniques en IA. Elle enseigne à ELISAVA et intervient à l'international sur les approches centrées sur l'humain dans la technologie. Pour en savoir plus, consultez le site https://aysegulguzel.info ou abonnez-vous à sa lettre d'information AI of Your Choice à l'adresse https://aysegulguzel.substack.com.

Réaliser la première évaluation des risques liés à l'IA de votre OSC, initialement élaboré par Ayşegül Güzel pour la série « Feuille de route pour une IA responsable pour les OSC », développée dans le cadre de l'initiative L'IA au service du changement social au sein du Digital Activism Program de TechSoup avec le soutien de Google.org. Tous les contenus sont mis à disposition sous la licence Creative Commons Attribution 4.0 International, sauf indication contraire. L'auteur a utilisé l'IA pour créer ce contenu. Toutefois, son intégralité a été conçue, révisée et vérifiée par l'auteur et l'équipe de TechSoup.