

# Diversité, inclusion et justice algorithmique : les nouveaux défis de l'entreprise responsable à l'ère de l'IA

## Cadre introductif

*Diversité et inclusion : des leviers de performance avérés*

La diversité et l'inclusion ne relèvent pas seulement de l'éthique sociale : ce sont également des **vecteurs puissants de performance organisationnelle**.

De nombreuses études empiriques confirment un lien positif entre la diversité des équipes et les résultats économiques des entreprises. Par exemple, un rapport de McKinsey (Hunt et al., 2015) a démontré que les entreprises figurant parmi les plus diversifiées ont une probabilité significativement plus élevée de surperformer financièrement : +15 % pour la diversité de genre, et jusqu'à +35 % pour la diversité ethnique.

Cette corrélation s'est même renforcée dans les analyses les plus récentes, atteignant +39 % pour les entreprises se situant dans le premier quartile de diversité (Hunt et al., 2023).

En parallèle, **l'inclusion vécue au quotidien** agit comme un véritable multiplicateur de performance. Une étude conduite par Deloitte (Bourke & Dillon, 2018) révèle qu'un renforcement du sentiment d'inclusion chez les salariés améliore significativement plusieurs dimensions du travail collectif : la performance des équipes progresse de 17 %, la qualité des décisions s'accroît de 20 %, et la collaboration entre collègues enregistre une hausse de 29 %. Ces résultats confirment que l'inclusion ne se limite pas à une valeur sociale, mais constitue un levier stratégique pour les organisations.

En somme, les équipes diverses **et** inclusives sont plus innovantes, plus engagées, et prennent de meilleures décisions collectives — créant ainsi un cercle vertueux au sein de l'organisation (Shore et al., 2011).

*Une révolution numérique dans les pratiques RH*

Parallèlement à ces dynamiques humaines, la dernière décennie a vu émerger une **vague d'outils numériques et d'intelligence artificielle (IA)** dans la gestion des ressources humaines.

Du recrutement à l'évaluation des performances, en passant par la mobilité interne ou le développement des compétences, l'IA s'invite aujourd'hui dans de nombreuses décisions stratégiques RH. Il est par exemple devenu courant d'utiliser des algorithmes pour trier automatiquement les CV, de recourir à des systèmes prédictifs pour évaluer le potentiel des candidats, ou encore d'employer des outils d'analyse vidéo fondés sur l'IA dans le cadre d'entretiens virtuels. Ces pratiques, qui se généralisent rapidement, transforment profondément les modes de sélection et d'évaluation en entreprise.

Selon l'enquête mondiale *Global Talent Trends* de Mercer (2022), 55 % des directeurs RH aux États-Unis ont déjà recours à des algorithmes prédictifs dans leurs processus d'embauche. Certaines entreprises vont encore plus loin en mobilisant l'IA pour détecter les collaborateurs à

haut potentiel, anticiper les risques de départ ou encore objectiver les évaluations de performance, dans une logique de rationalisation accrue des décisions RH (Cecchi-Dimeglio, 2021).

Ces innovations sont souvent présentées comme des **leviers d'efficacité** et de rationalisation. Parfois même, elles sont perçues comme plus « neutres » que les jugements humains.

Mais derrière l'enthousiasme technologique, **de nouveaux défis apparaissent**. L'intégration de l'IA dans les pratiques RH soulève des **risques éthiques inédits**, notamment en matière d'équité, de biais et d'inclusion.

## **Les risques d'exclusion liés à l'IA**

### *Des biais systémiques difficilement visibles*

Lorsque les systèmes d'IA sont mal maîtrisés ou mal calibrés, ils peuvent introduire des **biais systémiques** et générer des formes d'exclusion insidieuses dans les pratiques RH.

Le cas désormais emblématique d'Amazon illustre ce phénomène : en 2018, l'entreprise a dû abandonner un algorithme interne de recrutement qui **défavorisait systématiquement les candidatures féminines**. Les CV contenant le mot « femme » ou mentionnant des activités associées au genre féminin étaient pénalisés, en raison d'un entraînement fondé sur dix années d'historiques dominés par des profils masculins (Dastin, 2018).

Ce type de dérive révèle un problème de fond : **les données historiques sont rarement neutres**. Comme le soulignent Barocas et Selbst (2016), « un algorithme n'est jamais meilleur que les données dont il est nourri ». Si les jeux de données reflètent les préjugés d'une époque ou d'un contexte organisationnel biaisé, alors l'IA ne fait que les **reproduire, voire les amplifier**, de manière automatisée.

### *De nouvelles formes de discrimination indirecte*

Au-delà des biais explicites, de **nombreux facteurs d'injustice émergent indirectement**.

L'origine ethnique ou sociale, par exemple, peut être inférée à travers des **variables proxy** : le code postal, l'établissement scolaire fréquenté ou même certaines tournures de langage. Ces éléments peuvent conduire un algorithme à écarter des profils issus de milieux populaires, sans que cela ne soit volontaire ni visible (Binns, 2018).

Le **genre** peut lui aussi être inféré de manière implicite : une rupture de carrière, souvent liée à une maternité, ou la mention d'activités genrées peuvent faire baisser la note attribuée à une candidature.

De plus, l'**âge** devient un critère à risque dans les systèmes de présélection. Un cas récent a révélé qu'un algorithme rejetait automatiquement les femmes de plus de 55 ans et les hommes de plus de 60 ans, entraînant une procédure judiciaire pour discrimination automatisée.

### *Une inégalité face au capital numérique*

Au-delà des critères classiques (genre, origine, âge), une nouvelle forme d'inégalité apparaît : **l'inégalité numérique**.

Les candidats et les employés n'ont pas tous le même niveau d'aisance avec les outils digitaux, le même accès à un ordinateur performant, ou encore une présence en ligne optimisée.

Dans un monde de plus en plus digitalisé, certains profils sont **pénalisés simplement parce qu'ils ne savent pas «se vendre» numériquement** ou n'ont pas un profil LinkedIn bien construit. Ces disparités renforcent l'exclusion de populations déjà vulnérables, comme l'ont montré plusieurs enquêtes sur les biais liés à l'IA dans le recrutement (Lee, 2021).

### *Une opacité qui fragilise la confiance*

Un autre risque majeur réside dans **l'opacité des algorithmes**. Contrairement à un jugement humain, que l'on peut interroger, contextualiser ou critiquer, les modèles de machine learning produisent des décisions **difficiles à expliquer**— ce que certains appellent des « boîtes noires ».

Cette opacité fragilise les droits des individus.

Ni le candidat recalé, ni parfois même le recruteur, ne comprennent les critères ayant conduit au rejet ou à une évaluation défavorable.

Cela alimente un **sentiment d'arbitraire**, qui peut nuire à la confiance dans l'organisation. Plusieurs études montrent que les décisions prises par des systèmes automatisés sont perçues comme **moins justes et moins humaines**, ce qui affecte le climat social, la motivation et la loyauté des collaborateurs (Kim, 2022 ; Lee, 2021).

Dès lors, l'IA, si elle n'est pas encadrée, peut **saboter les efforts de diversité et d'inclusion**, pourtant souvent affichés comme des priorités stratégiques par les entreprises.

## **Vers une inclusion numérique responsable**

### *Rétablir un équilibre éthique face aux biais*

Pour éviter que l'IA ne renforce des inégalités ou ne crée de nouveaux biais, il devient impératif d'instaurer une **gouvernance éthique** des outils numériques utilisés en ressources humaines.

Cela implique, en premier lieu, d'adopter des **principes de transparence, d'explicabilité et d'équité** dans la conception comme dans l'usage des algorithmes RH.

Cette exigence de lisibilité est au cœur des travaux sur la justice algorithmique (Binns, 2018 ; Barocas & Selbst, 2016) et fait partie des piliers de la **Corporate Digital Responsibility (CDR)**, une extension éthique et stratégique de la RSE au champ numérique.

### *Expliquer les décisions algorithmiques : une exigence stratégique*

La transparence algorithmique constitue un principe central. Elle vise à rendre les décisions compréhensibles pour les individus concernés — qu'il s'agisse d'un candidat, d'un collaborateur ou d'un manager. Concrètement, cela implique de communiquer clairement les critères généraux de sélection, d'être en mesure de justifier toute décision automatisée, et de faciliter les recours en cas de contestation ou de doute sur l'équité du processus.

Comme l'ont souligné Buhmann, Paßmann et Fieseler (2021), les entreprises doivent développer une capacité organisationnelle à **rendre des comptes sur leurs algorithmes**, surtout lorsqu'ils affectent la carrière des individus.

Cette transparence permet non seulement de renforcer la confiance, mais aussi de **repérer plus rapidement les biais ou dysfonctionnements** internes.

### *Élargir la RSE à l'ère numérique*

De plus en plus d'auteurs plaident pour une approche holistique, en **prolongeant les principes de la RSE dans le domaine digital**.

La notion de **responsabilité numérique de l'entreprise (CDR)** s'impose comme un cadre pertinent. Elle peut être définie comme « l'ensemble des normes, pratiques et valeurs guidant l'usage éthique des technologies numériques dans une organisation » (Kim, 2022 ; Buhmann et al., 2021).

La CDR couvre l'ensemble du cycle de vie des systèmes numériques, depuis leur conception et l'acquisition des données jusqu'à leur utilisation, leur évaluation et leur adaptation continue. Elle ne vise pas à remplacer la RSE, mais à en constituer une extension nécessaire dans un environnement de plus en plus algorithmisé. Les principes classiques de non-discrimination, de transparence, de redevabilité ou encore de respect des parties prenantes s'y traduisent aujourd'hui par des exigences telles que l'auditabilité des algorithmes, la supervision humaine des décisions automatisées, ou encore le respect strict de la vie privée et des droits numériques des individus (Ministère du Travail, 2018 ; Cecchi-Dimeglio, 2021).

### *Une attente sociétale et une opportunité stratégique*

Au-delà de la conformité, ces pratiques répondent à une **attente croissante des parties prenantes**.

Les jeunes générations de collaborateurs, notamment, se montrent de plus en plus **attentives à l'éthique numérique** de leur employeur (Bourke & Dillon, 2018 ; Deloitte Insights).

À long terme, intégrer ces exigences peut devenir un **levier de réputation, d'attractivité et de performance durable**.

L'entreprise qui saura conjuguer innovation numérique et justice sociale sera perçue comme plus responsable, plus crédible et plus fiable dans un monde digitalisé.

## Pistes concrètes pour les organisations

### 1. Former au leadership inclusif « augmenté »

La première action structurante consiste à **former les managers et dirigeants** à un **leadership inclusif à l'ère de l'IA**.

Le leadership inclusif repose sur la capacité à reconnaître, valoriser et faire coexister des identités, trajectoires et cultures différentes au sein des équipes.

Or, selon une étude de Deloitte, seuls 5 % des leaders dans le monde répondraient actuellement aux critères d'un leadership véritablement inclusif (Bourke & Dillon, 2018).

À l'ère des outils numériques, ce leadership doit évoluer vers une posture augmentée par les données, sans pour autant se laisser dominer par elles. Le manager doit ainsi être en mesure d'interroger de manière critique les recommandations issues des outils algorithmiques, de croiser ces résultats avec des évaluations humaines, et d'identifier les biais potentiels pouvant se dissimuler dans les décisions proposées par l'IA.

Comme le souligne Cecchi-Dimeglio (2021), les leaders RH ont un rôle central à jouer : ce sont les gardiens de la culture d'entreprise, garants de l'équité et de la diversité dans la durée. Leur exemplarité dans l'usage raisonné et éthique de l'IA est déterminante.

### 2. Mettre en place une veille active sur les biais algorithmiques

Une autre piste incontournable consiste à instaurer une veille continue sur les biais algorithmiques. Cela implique la mise en place de dispositifs d'audit réguliers des outils RH automatisés, afin de vérifier qui est sélectionné, évalué ou promu par les algorithmes, d'identifier d'éventuelles disparités entre groupes — en fonction du genre, de l'âge ou de l'origine — et de repérer les effets indirects ou inattendus que ces systèmes pourraient engendrer dans les processus décisionnels.

Certaines entreprises développent aujourd'hui des **comités d'équité algorithmique**, réunissant RH, data scientists, juristes et représentants des parties prenantes (Kim, 2022 ; Buhmann et al., 2021).

Cette gouvernance permet de **détecter les signaux faibles** et d'anticiper d'éventuelles contestations, en favorisant une logique de responsabilité partagée.

Cette pratique rejoint le principe de **redevabilité numérique (algorithmic accountability)**, qui devient un enjeu-clé des régulations à venir — notamment dans le cadre du futur AI Act européen (Kim, 2022).

### 3. Élaborer une charte d'éthique numérique RH

À l'instar des codes de conduite RSE, il devient pertinent d'élaborer une **charte d'éthique numérique RH**, document de référence pour encadrer l'usage des outils digitaux dans la gestion du personnel.

Une première initiative pionnière a vu le jour en France avec la Charte Éthique et Numérique RH (Ministère du Travail, 2018), issue d'une co-construction entre partenaires sociaux et experts en intelligence artificielle. Ce type de charte permet de formaliser les principes de non-discrimination et d'équité applicables aux processus algorithmiques, de poser des engagements clairs en matière de transparence vis-à-vis des candidats, et d'instaurer des garde-fous, notamment en garantissant un contrôle humain systématique sur toute décision sensible.

Au-delà de sa fonction régulatrice, la charte est un **levier d'appropriation collective**. Elle responsabilise les acteurs RH, structure les choix technologiques et constitue un signal fort adressé aux parties prenantes internes comme externes (Cecchi-Dimeglio, 2021).

#### 4. Renforcer les synergies entre RH, IT et gouvernance RSE

Enfin, répondre aux enjeux d'équité numérique exige une **collaboration transversale** entre fonctions : RH, équipes IT/data, et pilotage de la RSE doivent travailler en synergie.

Cette gouvernance partagée peut se traduire concrètement par la création de comités de pilotage réunissant les parties prenantes autour de l'IA éthique, par la mise en œuvre de démarches de co-évaluation des outils RH impliquant des experts techniques, ou encore par le développement de programmes de formation croisée à destination des recruteurs, des data analysts et des juristes, afin de favoriser une compréhension mutuelle des enjeux.

Comme le préconisent Buhmann et al. (2021), cette intégration organisationnelle permet d'ancrer l'éthique dans la conception même des systèmes. Elle traduit une vision stratégique : **la diversité, l'inclusion et la justice algorithmique ne peuvent être pensées séparément** des logiques de transformation digitale.

## Conclusion

### *L'IA au service de la justice sociale, et non l'inverse*

Diversité, inclusion et intelligence artificielle ne doivent pas s'opposer. Elles doivent au contraire être pensées ensemble, comme **des composantes indissociables** de la transformation responsable de l'entreprise.

L'ère de l'IA offre des opportunités immenses en matière de gestion des talents, d'optimisation des processus RH et d'innovation managériale. Mais ces opportunités ne peuvent se concrétiser que si les organisations adoptent une **posture de vigilance éthique** et de **responsabilité numérique**.

L'entreprise responsable de demain sera celle qui saura former un leadership inclusif véritablement augmenté par la technologie, gouverner l'intelligence artificielle en s'appuyant sur les principes d'équité et de transparence, ancrer ses outils RH dans une culture organisationnelle orientée vers la justice sociale, et articuler de manière cohérente ses engagements RSE avec les exigences émergentes de la responsabilité numérique des entreprises (CDR).

Il ne s'agit pas de ralentir la transformation digitale, mais de l'**orienter intelligemment**. L'IA ne sera un levier de performance durable que si elle respecte les valeurs humaines fondamentales, à commencer par l'égalité des chances, la non-discrimination et la dignité des parcours individuels.

En intégrant ces principes dès aujourd'hui, les entreprises peuvent faire de l'intelligence artificielle **non pas une menace**, mais un **formidable levier de progrès collectif et inclusif**.