DROIT & POLITIQUES
HUMANITAIRES

Droit international, détention et intelligence artificielle : vers une humanité externalisée ?

décembre 23, 2025, Détention / Droit et conflits / Générer le respect du DIH / Nouvelles technologies

20 mins read

Terry Hackett
Chef de la Division des personnes privées de liberté, CICR

Alexis Comninos
Conseiller juridique thématique, Comité international de la Croix-Rouge (CICR)

Alors que l'intelligence artificielle (IA) commence à influencer la prise de décision en matière de détention de personnes en période de conflit armé, ainsi que la façon dont les lieux de détention sont administrés, des questions qui relevaient autrefois de la science-fiction deviennent aujourd'hui de réelles préoccupations au regard du droit et de l'éthique. L'exploitation massive de données et la volonté d'optimiser l'efficacité risquent d'éliminer tout discernement humain de l'un des domaines les plus sensibles de la guerre : la privation de liberté. Utilisée dans ce contexte, l'IA dépouillerait les détenus de leur dernière part d'humanité, les réduirait à des données et saperait les protections humanitaires fondamentales que les Conventions de Genève sont censées accorder.

Dans le présent article, Terry Hackett, chef de la Division des personnes privées de liberté au CICR, et Alexis Comninos, conseiller juridique thématique, examinent les liens entre l'IA et le droit international humanitaire (DIH) dans les opérations de détention et expliquent pourquoi seuls des humains peuvent traiter les détenus avec humanité. S'appuyant sur les récentes

recommandations du CICR au Secrétaire général des Nations Unies, ils font valoir que si le DIH ne s'oppose pas à l'innovation, il fixe des limites morales et juridiques qui garantissent que les avancées technologiques ne mettent pas en péril la dignité humaine.

La vie des personnes placées en détention en lien avec un conflit armé dépend entièrement des décisions de leurs ravisseurs. Coupées de leur famille et de leur communauté, elles sont particulièrement exposées aux risques de détention arbitraire, de mauvais traitements, de disparition et de négligence. Or quand la technologie s'invite dans ce contexte précaire, les enjeux deviennent encore plus élevés. L'intégration de l'intelligence artificielle (IA) dans le cadre de la privation de liberté risque de réduire les individus à des données et de remettre leur sort entre les mains de systèmes opaques, insensibles aux souffrances, exempts de doutes et dénués de toute compassion.

Conscient de ce danger, le Comité international de la Croix-Rouge (CICR) a tiré la sonnette d'alarme dans son *rapport sur les défis posés au DIH 2024*, où il signale que les biais, le manque de transparence et la perte de contrôle par les humains pourraient entraver le respect du droit international humanitaire (DIH). Le risque s'étend à l'ensemble des aspects de la détention : de l'identification des personnes à interner à la gestion des installations, en passant par l'examen de la situation des détenus et la libération de ceux qui doivent l'être.

Le respect du DIH par l'État doit servir de garde-fou et veiller à ce que la technologie contribue à traiter les personnes avec plus d'humanité, et non l'inverse. Les Conventions de Genève *[1]* imposent aux États parties de respecter et faire respecter le DIH, notamment de traiter les personnes privées de liberté avec humanité, et cette obligation s'applique également aux systèmes pilotés par l'IA. Les arguments qui suivent s'appuient sur les *recommandations préliminaires du CICR* au Secrétaire général des Nations Unies, au regard de trois dimensions essentielles permettant de maintenir l'humanité au cœur des progrès technologiques en temps de guerre : une approche centrée sur l'humain, des garanties sûres et une mise à l'essai, et un examen juridique permanent.

Maintenir une composante humaine

Grâce aux fonctionnalités d'analyse prédictive, l'IA peut aider les humains à prendre des décisions, à améliorer les services destinés aux détenus et à gérer les lieux de détention plus efficacement. Toutefois, c'est bien la façon dont les systèmes sont conçus, mis en œuvre et utilisés par les personnes qui déterminera si les détenus sont toujours traités avec humanité ou si ce traitement se dégrade.

De l'avis du CICR, le DIH exige que les décisions juridiques soient prises par des humains, car ce sont aux humains – et aux États ou aux groupes armés qu'ils représentent – qu'il incombe de respecter les obligations fixées par le droit. Ce principe fondamental reste valable lorsque des humains utilisent l'IA, y compris pour générer des recommandations censées orienter leur prise de décision.*[2]*

D'aucuns prétendent que l'IA est garante d'une meilleure impartialité et d'une plus grande cohérence dans la décision par rapport au jugement humain. Or l'idée que l'IA serait impartiale par nature est fallacieuse : les systèmes peuvent être conçus ou formés sur des données biaisées, qu'ils peuvent ensuite reproduire et même amplifier. L'impartialité humaine n'est certes jamais garantie, mais aucun algorithme ne saurait remplacer l'empathie, la discrétion ou le discernement humains – un cri de douleur, un appel à l'aide ou une demande de contact avec des proches ne doit jamais être rejeté ou considéré comme bruit de fond par un algorithme. L'impartialité n'existe pas en soi : elle dépend de systèmes délibérément conçus pour la respecter. L'accès aux soins et le contact avec des proches sont des droits qui doivent être garantis, et l'obligation de respecter ces droits incombe aux humains et non aux machines.

La détention concerne avant tout des êtres humains. Le maintien de l'ordre et de la discipline nécessite des interactions attentives et régulières entre gardes et détenus, afin de permettre à ces derniers d'avoir toujours une bonne perception de la situation et de maintenir la confiance. Externaliser la gestion d'un lieu de détention à un algorithme risque d'exclure la part d'humanité dans l'équation.

Les humains doivent donc rester présents à chaque étape : réception d'informations, analyse, production de résultats et les décisions finales – et veiller à ce que l'exigence de traiter chaque personne avec humanité demeure au cœur de toute initiative. À cette fin, les gardes doivent recevoir une formation portant non seulement sur leurs obligations juridiques et les techniques de gestion des détenus, mais aussi sur la manière d'intégrer l'IA de façon responsable dans leur travail.

Pour assurer une approche véritablement centrée sur l'humain, les planificateurs militaires, les autorités détentrices et les développeurs de systèmes d'IA doivent avant tout avoir une compréhension commune de l'environnement de détention et des facteurs de vulnérabilité qu'il présente. Les défis liés à la garantie d'un traitement humain, aux conditions de détention, à l'accès aux services essentiels et à la protection des données personnelles doivent être traités dès la phase de conception de tout système d'IA.

Mesures préparatoires permettant de respecter le DIH

Dans *un article récemment* publié, Isabelle Gallino et Sylvain Vité présentent les mesures préparatoires requises pour que les États respectent le DIH en détention en situation de conflit armé international. Ils expliquent comment adapter les infrastructures, les institutions et les instructions bien avant le début des hostilités. Ces principes doivent être appliqués de la même manière dans le contexte de l'introduction de l'IA dans le milieu carcéral. C'est en temps de paix que les États peuvent œuvrer au déploiement de technologies fondées sur l'IA conformes au DIH, en élaborant des structures de détention/d'internement et des systèmes d'IA qui respectent cette branche du droit, en établissant des cadres juridiques et politiques clairs et en répartissant les responsabilités entre les institutions publiques, les acteurs privés et les personnes.*[3]*

Planifier le respect du DIH implique également de tester et d'adapter rigoureusement toutes les technologies fondées sur l'IA qui sont à l'étude. Cela est particulièrement important lorsque les États réutilisent des systèmes qui ont été élaborés dans d'autres contextes. Puisque les systèmes de justice pénale du

monde entier expérimentent des outils fondés sur l'IA, les États peuvent être tentés de les importer dans les situations de détention en lien avec des conflits armés pour gagner en efficacité. Pourtant, Ashley Deeks nous met en garde dans *son article* : une telle approche risque de nous faire tomber dans le « piège de la portabilité », c'est-à-dire l'incapacité à comprendre comment des solutions algorithmiques conçues pour un environnement social (ou juridique) peuvent fausser les résultats dans un autre.*[4]*

Aucun système d'IA n'est intégré à un contexte de détention de but en blanc. Tout système interagit avec les lois, les infrastructures et les dynamiques sociales existantes à tous les stades – de la conception des installations aux relations entre gardes et prisonniers. Un algorithme formé sur les données d'une population dans un contexte culturel, social ou juridique différent peut produire des résultats biaisés ou trompeurs lorsqu'il est appliqué à un autre contexte. Un deuxième piège de portabilité juridique peut se présenter lorsqu'un système fondé sur l'IA et développé dans un cadre dans lequel les droits de l'homme sont applicables ou pour des prisons de haute sécurité est réutilisé dans le contexte de la détention en période de conflit armé, où d'autres règles et des principes s'appliquent, y compris le principe d'assimilationapplicable aux prisonniers de guerre.*[5]*

Les formations sur des systèmes fondés sur l'IA jouent un rôle important à la fois pour mettre à l'essai ces technologies et pour limiter certains des risques évoqués ci-dessus. Ces formations peuvent être dispensées dans le cadre d'exercices militaires qui reproduisent de façon réaliste les conditions d'un conflit armé de haute intensité. Or, compte tenu des limites inhérentes à la mise à l'essai lorsqu'elle est faite dans un environnement aussi contrôlé qu'un exercice militaire *[6]*, celle-ci ne saurait constituer à elle seule une garantie efficace. En réalité, aucun exercice militaire ne peut couvrir l'ensemble des scénarios possibles ni reproduire pleinement les conditions hostiles d'un conflit armé.

En plus des exercices militaires, les États peuvent effectuer des « tests parallèles » sur ces technologies lors de vraies opérations menées en temps réel, en veillant à ce que le système fondé sur l'IA n'ait aucune capacité de prise de décision ni même d'assistance à la décision. Ce type de test permet aux experts d'examiner et d'évaluer a posteriori les résultats du système, sans risque d'engager la responsabilité individuelle ou étatique et, surtout, sans risque de porter atteinte aux droits des personnes sur lesquelles il est testé. Une fois cette mise à l'essai analysée, la technologie et son utilisation peuvent être ajustées, adoptées ou rejetées. on the assessment of such testing, the technology and its use may be adjusted, adopted, or discarded.

Poser les conditions nécessaires au respect des dispositions du DIH relatives à la non-discrimination *[7]* nécessite d'adopter une approche qui prenne en compte les biais du système dès le départ – en les atténuant, en les réduisant et en les corrigeant à chaque étape de la durée de vie d'un système fondé sur l'IA, de sa conception et sa formation à son déploiement et son évaluation. Ce n'est qu'en mettant en place des mesures préparatoires permettant de respecter le DIH dès le temps de paix que les États peuvent espérer respecter pleinement leurs obligations en cas de guerre.

Examens juridiques, transparence et droit de recours

L'article 1 commun aux quatre Conventions de Genève impose à toutes les Parties de « respecter et faire respecter » les Conventions. Il est impossible d'être certain que l'application d'une nouvelle technologie telle que l'IA à la détention dans le cadre d'un conflit armé respecte les obligations des Conventions sans procéder à un examen juridique.*[8]*

Pour être efficaces, ces examens doivent non seulement être menés au stade de la création et de la conception ainsi qu'en amont du déploiement d'une nouvelle technologie d'IA, mais également chaque fois qu'un outil existant doit être utilisé dans un but, un contexte opérationnel ou un cadre juridique différents (conflits armés internationaux vs conflits armés non internationaux ; cadres internationaux vs cadres régionaux des droits de l'homme). Plusieurs règles de DIH relatives à la détention s'appliquent de manière totalement différente dans les conflits armés internationaux et non internationaux. Par ailleurs, les normes en matière de traitement doivent être adaptées à la personne concernée, notamment à son genre, son âge et son handicap éventuel. Les algorithmes d'évaluation des menaces développés avec des données relatives à une population particulière ne peuvent pas être considérés comme transposables à une autre population sans être minutieusement adaptés. En d'autres termes, les examens juridiques ne peuvent pas être considérés comme des processus en une seule étape ; ils doivent être permanents ou reproduits à plusieurs reprises.

S'agissant des systèmes d'aide à la décision fondés sur l'IA appliqués à la détention dans les conflits armés, les obligations juridiques au regard desquelles le respect du DIH doit être examiné comprennent notamment le traitement humain, le caractère non arbitraire de la détention, la non-discrimination, la mise en œuvre effective des garanties judiciaires ainsi que le régime de détention applicable (internement en vertu de la Troisième ou de la Quatrième Conventions de Genève, justice pénale, etc.).

Prenons l'exemple de l'utilisation de systèmes d'aide à la décision fondés sur l'IA pour des décisions relatives à l'internement des personnes protégées en vertu de la Quatrième Convention de Genève : il faudrait non seulement veiller à ce que l'utilisation de cette technologie soit compatible avec les exigences d'absolue nécessité pour la « sécurité de la Puissance au pouvoir » (s'agissant de l'internement sur le territoire d'un belligérant) et de nécessité « pour d'impérieuses raisons de sécurité » (s'agissant de l'internement sur un territoire occupé*[9]*) mais aussi que les garanties procédurales *[10]* puissent être appliquées efficacement, que les décisions soient non discriminatoires et qu'elles soient prises *au cas par cas*.

L'examen de la licéité des systèmes d'IA au regard des garanties procédurales visées aux articles 43 et 78(2) de la Quatrième Convention de Genève nécessiterait notamment de s'assurer que l'utilisation des systèmes d'aide à la décision fondés sur l'IA ne contrevienne pas au droit de l'interné de contester la décision d'internement, y compris, par exemple, l'accès aux informations sur les motifs de son internement.*[11]* L'opacité par nature de la prise de décision assistée par l'IA (souvent appelée l'« effet boîte noire ») limite la capacité de l'utilisateur – en l'espèce, de l'autorité détentrice – à saisir pleinement le processus ayant conduit à une recommandation spécifique. S'agissant des internements dans le cadre d'un conflit armé, il faut également tenir compte de la décision des autorités de ne pas communiquer certaines informations à la personne internée, en raison de la nécessité militaire et pour des motifs liés au secret militaire.

En un sens, cet ensemble de facteurs peut créer un *double* « effet boîte noire » pour la personne internée. Comme l'indique *Jelena Pejic*, l'accès de chaque individu à des informations sur les raisons pour lesquelles il a été privé de liberté peut être « considéré comme un des éléments constitutifs de l'obligation de traitement humain, car on sait que l'incertitude d'une personne quant aux motifs de sa détention représente une source de stress psychologique aigu."*[12]*. Il est donc peu probable que l'utilisation de systèmes fondés sur l'IA qui ne sont pas en mesure de fournir aux internés des informations significatives sur les raisons de leur détention soit conforme au DIH.

Des systèmes utilisant l'IA peuvent être utilisés dans des lieux de détention en période de conflit armé pour gérer les flux et les déplacements des personnes privées de liberté. Cela peut par exemple permettre de réduire le nombre de gardes nécessaires et de redéployer les ressources humaines sur d'autres tâches (y compris des activités liées au combat). Cependant, appliqué à l'internement des prisonniers de guerre dans le cadre de la Troisième Convention de Genève, un système fondé sur l'IA, quel qu'il soit, devrait permettre à son utilisateur d'appliquer les principes de l'internement c'est-à-dire, une privation de liberté non punitive, dans laquelle les prisonniers ne sont pas consignés et jouissent d'une grande liberté de mouvement à l'intérieur du périmètre d'un camp. Il va sans dire que si ce système fondé sur l'IA était en revanche appliqué en détention dans le cadre d'un conflit non international, il ne serait pas évalué et examiné au regard des mêmes normes, et que ces deux systèmes ne seraient en aucun cas interchangeables sans adaptations importantes.

Conclusion

Le droit international humanitaire n'est pas l'ennemi de l'innovation ou du progrès. Au contraire, il fournit les garde-fous et les garanties nécessaires aux développeurs et aux utilisateurs pour atténuer les risques et utiliser l'IA de façon responsable dans le domaine militaire. Alors que les États souhaitent recourir à l'IA pour exploiter des données, réaliser des gains d'efficacité et maximiser les ressources dans les opérations militaires, ils ne pourront le faire au détriment des droits et de la dignité des personnes privées de liberté ou sans s'acquitter des obligations juridiques qui leur incombent.

Les quelque 130 conflits armés qui ravagent le monde actuellement et les mauvaises interprétations du DIH qui sont parfois utilisées pour justifier son non-respect, ainsi que les très lourdes conséquences humanitaires qui en découlent, laissent penser que l'IA ne sera pas toujours utilisée avec les meilleures intentions en détention. Et même dans le cas contraire, il n'est pas certain qu'elle respecte le droit. Comme tout autre outil, c'est la façon dont les humains conçoivent les systèmes, les développent, les mettent à l'essai et les utilisent qui déterminera s'ils deviennent une force au service du bien ou un vecteur de nouvelles violations et de souffrances.

Cet article a initialement été publié en anglais le 13 novembre 2025.

Notes

[1] Articles 1 et 3 communs aux quatre Conventions de Genève de 1949.

[2] « ICRC Position Paper : Artificial intelligence and machine-learning in armed conflict : a human-centred approach », *Revue internationale de la Croix-Rouge*, n° 913, mars 2022 : « Il est essentiel de préserver le contrôle et le jugement humains dans les systèmes fondés sur l'IA et l'apprentissage automatique pour les tâches et les décisions qui peuvent avoir de graves conséquences sur la vie des personnes, en particulier lorsque ces tâches et ces décisions présentent des risques qui mettent en péril la vie humaine et lorsqu'elles sont régies par des normes spécifiques du droit international humanitaire. Les systèmes fondés sur l'IA et d'apprentissage automatique doivent être au service des acteurs et des décideurs humains et non les remplacer. », disponible sur : *https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913*.

[3] Croix-Rouge française, Croix-Rouge australienne et CICR, *Private Businesses and armed conflict, and Introduction to relevant rules of international humanitarian law*, 4829/002, 2024, section 2.2.

[4] Andrew D. Selbst *et al.*, *« Fairness and Abstraction in Sociotechnical Systems »*, *FAT* '19 : Conference on Fairness, Accountability, and Trans parency* (FAT* '19), janvier 2019 29-31, p. 61.

[5] Le principe d'assimilation est le fondement sur lequel reposent plusieurs règles de la Troisième Convention de Genève. Il exprime l'idée selon laquelle les prisonniers de guerre doivent bénéficier d'un traitement et de conditions de vie semblables à ceux dont jouissent les membres des troupes de la Puissance détentrice. Voir Commentaire de la CGIII, CICR, 2020, par. 30-38.

[6] Robin Geiss, Henning Lahmann, « The use of AI in military contexts : opportunities and regulatory challenges », *The Military Law and the Law of War Review*, vol. 59 n° 2, 2021, pp. 182-183.

[7] Article 3 commun ; CGIII, art.16. ; CGIV, art. 27.

[8] Le Protocole additionnel I (PAI) aux Conventions de Genève établit une obligation distincte de procéder à l'examen de tout nouveau moyen ou méthode de guerre (art. 36, PA I). Les États peuvent donc utiliser des processus et des structures similaires pour examiner la légalité des technologies fondées sur l'IA, y compris lorsqu'elles sont utilisées en détention.

[9] CG IV, art. 42 et 78.

[10] Visées à l'art. 43 et à l'art. 78(2).

[11] Comme cela est précisé à l'article 75(3) du Protocole additionnel I ; voir aussi Jelena Pejic, « Principes en matière de procédure et mesures de protection pour l'internement/la détention administrative dans le cadre d'un conflit armé et d'autres situations de violence », *Revue internationale de la Croix-Rouge*, vol. 87, n° 858, juin 2005, p. 341 : « Les informations fournies doivent être suffisamment détaillées pour que la personne privée de liberté puisse contester la légalité de son internement, ou de sa détention administrative, et exiger que la décision soit reconsidérée. »

[12] *Ibid.*, p. 341.

Voir aussi :

- Ruben Stewart and Georgia Hinds, *Les algorithmes de la guerre : le recours à l'IA dans la prise de décision dans les conflits armés*, 22 février 2024
- Erica Harper, *L'IA va-t-elle profondément transformer la manière dont les conflits armés sont déclenchés, menés et résolus ?*, 12 décembre 2025.

Tags: CG III, Conventions de Genève, détention, droit international humanitaire, IA, intelligence artificielle, prisonniers de guerre, respect du DIH, traitement humain

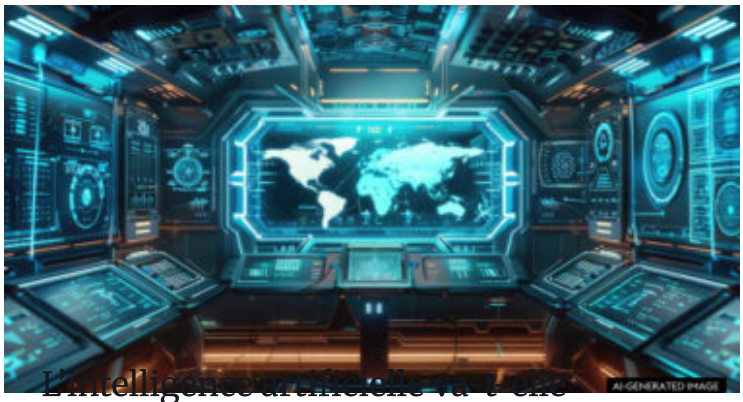
You should also read these articles

Même les « cyberguerres » ont des limites. Et si elles n'en avaient pas ?

14 mins read

Détention / Droit et conflits / Générer le respect du DIH / Nouvelles technologies Tilman Rodenhäuser & Kubo Mačák

Les cyberopérations sont devenues une réalité des conflits armés contemporains, et il est probable qu'elles ...



L'intelligence artificielle va-t-elle profondément transformer la manière dont les conflits armés sont déclenchés, menés et résolus ?

14 mins read

Détention / Droit et conflits / Générer le respect du DIH / Nouvelles technologies Erica Harper

Dans le cadre du débat concernant les effets de l'intelligence artificielle (IA) sur les stratégies ...