

L'intelligence artificielle dans le droit pénal suisse

Potentiel d'efficacité et limites de l'État de droit

Kaspar Brönnimann · Web Edition 2026 · Switzerland · kaspar@kaspair.ch · kaspair.ch

ABSTRACT

Cette étude examine dans quelle mesure l'intelligence artificielle peut améliorer l'efficacité de la justice pénale suisse sans affaiblir les garanties fondamentales de l'État de droit. Elle distingue les fonctions d'enquête, de sanction et de jugement et analyse notamment l'explicabilité, les biais, le biais d'automatisation, le principe de culpabilité individuelle, les garanties du procès équitable et l'attribution de la responsabilité. Elle défend un modèle hybride dans lequel l'IA demeure un outil d'aide à la décision tandis que la responsabilité juridique ultime reste assumée par une personne identifiable.

Note éditoriale : cette étude est fondée sur un travail préparé dans le cadre du cours « AI & Law » proposé par l'Université de Lund sur Coursera. La version remise pour le cours a été abrégée afin de respecter les contraintes de l'exercice. Cette édition web indépendante rétablit un appareil de sources et comporte quelques mises à jour techniques et bibliographiques mineures. Elle ne constitue pas une publication officielle de l'Université de Lund et n'a pas été validée par celle-ci. Ce texte ne constitue pas un avis juridique.

1 Introduction

L'intelligence artificielle (IA) est de plus en plus utilisée dans les processus décisionnels publics, notamment lorsque de grandes quantités de données doivent être traitées ou que des tâches prédictives complexes sont en jeu. Le droit pénal, traditionnellement fondé sur l'attribution individuelle de la faute, l'indépendance judiciaire et de strictes garanties de l'État de droit, est lui aussi concerné par cette évolution.

Pour évaluer le potentiel d'efficacité de l'IA dans la justice pénale, il convient de distinguer les fonctions exercées par les différents acteurs publics. L'usage de l'IA par les autorités d'enquête peut servir une clarification rapide et efficace des faits, tandis que son usage par des autorités qui tranchent ou sanctionnent est soumis à des exigences juridiques plus strictes.

Une particularité de la procédure pénale suisse tient au fait que les décisions de sanction ne sont pas exclusivement réservées aux tribunaux. Dans la procédure de l'ordonnance pénale, le ministère public peut lui-même constater la culpabilité et prononcer une peine, exerçant ainsi

matériellement une fonction juridictionnelle. Dans ces situations, les considérations d'efficacité perdent nettement de leur poids.

Il faut distinguer ces cas des contraventions traitées dans le cadre de la procédure d'amende d'ordre, fondée sur un catalogue standardisé. L'usage de systèmes automatisés ou assistés par IA y est moins problématique, pour autant que la possibilité de former opposition et de transférer ainsi l'affaire à la procédure pénale ordinaire demeure garantie.

L'activité des tribunaux, en revanche, vise à déterminer individuellement la culpabilité et la peine. Dans ce contexte, la rapidité n'a pas le même poids que la transparence, l'obligation de motiver, l'indépendance judiciaire et l'accès à une protection juridique effective.

La présente étude examine, selon une approche juridique doctrinale, si et dans quelle mesure l'usage de l'IA dans le droit pénal suisse peut produire des gains d'efficacité juridiquement admissibles sans compromettre les principes fondamentaux du droit pénal ni les garanties constitutionnelles et conventionnelles.^{7 6 3}

2 Faits et technique : l'IA dans le contexte pénal

L'usage de l'IA dans la justice pénale couvre des applications techniques diverses, aux implications juridiques différentes. En Suisse, l'IA ne prend pas de manière autonome des décisions pénales, mais elle peut soutenir la décision humaine et influencer de facto certaines orientations d'enquête ou de priorisation.^{7 6}

2.1 Formes techniques pertinentes pour le droit pénal

Les systèmes d'IA pertinents pour la justice pénale comprennent notamment des outils d'analyse et de prédiction, des systèmes de classification, de recherche et de priorisation. L'apprentissage automatique déduit des modèles ou fonctions décisionnelles à partir de données, au lieu d'exécuter uniquement des règles explicitement programmées ; un modèle déployé en production ne continue pas nécessairement à s'adapter de manière autonome.

Une difficulté centrale tient à l'explicabilité limitée. Lorsque le fondement d'un résultat algorithmique ne peut pas être suffisamment reconstruit par le décideur humain, l'obligation de motiver et le droit d'être entendu peuvent être affectés.^{7 8}

2.2 Utilisation par les autorités d'enquête en Suisse

Dans la poursuite pénale, l'IA peut soutenir le travail d'enquête, par exemple lors de l'analyse de grands volumes de données ou de contenus visuels. Ces systèmes ont en principe une fonction préparatoire, mais peuvent néanmoins créer des préorientations de fait lorsque leurs

recommandations influencent la suite de l'enquête.⁶

2.3 Utilisation par les tribunaux et autorités de sanction

L'usage de l'IA par les tribunaux demeure comparativement limité en Suisse. Une décision autonome de culpabilité ou de peine par une IA n'est ni envisagée ni compatible avec la structure de la responsabilité pénale individuelle. Des systèmes algorithmiques peuvent toutefois soutenir des processus administratifs, la recherche juridique ou la structuration de dossiers volumineux.

Dans d'autres systèmes juridiques, des outils d'évaluation du risque ont été utilisés pour soutenir des décisions relatives à la détention, à la probation ou à la peine. Ces exemples illustrent les tensions qui apparaissent lorsque des prédictions algorithmiques se rapprochent du cœur de la décision juridictionnelle.

Sur le plan juridique, de tels usages sont particulièrement sensibles parce qu'ils peuvent affecter directement les droits fondamentaux, l'indépendance judiciaire et le principe de culpabilité.^{6 11}

2.4 Risques techniques et limites structurelles

Les systèmes d'IA utilisés dans la justice pénale peuvent reproduire des biais présents dans les données, produire des effets discriminatoires et réduire la traçabilité des décisions. S'y ajoute le biais d'automatisation, c'est-à-dire le risque que les décideurs humains accordent un poids excessif aux recommandations automatisées.

Ces caractéristiques sont centrales pour l'évaluation juridique, car elles conditionnent la possibilité d'un contrôle effectif, d'une contestation réelle et d'une responsabilité humaine identifiable.^{8 9 10}

3 Analyse juridique : défis posés par l'IA en droit pénal

3.1 Principes fondamentaux du droit pénal suisse

Le droit pénal suisse repose sur le principe de culpabilité individuelle. L'art. 47 CP exige une fixation individualisée de la peine, tenant compte de la situation personnelle de l'auteur et des circonstances concrètes de l'infraction.

Les systèmes d'IA fonctionnent en revanche fréquemment sur la base de corrélations statistiques et de probabilités. Ils peuvent détecter des modèles ou attribuer des scores de risque, mais ils ne peuvent eux-mêmes apprécier la faute morale ni assumer une responsabilité

juridique. Une tension structurelle apparaît lorsque l'inférence statistique se rapproche de l'appréciation de la culpabilité individuelle.²

3.2 Garanties du procès équitable et transparence

Des garanties essentielles découlent notamment de l'art. 29 Cst. et de l'art. 6 CEDH : droit à une procédure équitable, droit d'être entendu et accès effectif aux voies de recours.

Lorsqu'une décision pénalement pertinente est matériellement influencée par un processus algorithmique opaque, la personne concernée peut avoir des difficultés à comprendre ou contester les motifs réels de la décision. L'explicabilité et la documentation deviennent dès lors des exigences procédurales, et non de simples questions techniques.^{1 4 8}

3.3 Égalité devant la loi et discrimination

L'art. 8 Cst. garantit l'égalité devant la loi. Les systèmes d'IA peuvent mettre ce principe en danger lorsque les données d'entraînement ou d'utilisation reflètent des inégalités sociales existantes et produisent ainsi des effets indirectement discriminatoires.

La discrimination algorithmique peut être difficile à détecter lorsqu'elle résulte d'effets statistiques plutôt que de distinctions explicites, ce qui complique le contrôle judiciaire et la protection juridique effective.^{1 9}

3.4 Responsabilité et attribution

Une autre question fondamentale concerne l'attribution de la responsabilité. Dans une décision assistée par IA, les contributions techniques et organisationnelles sont réparties entre développeurs, autorités, opérateurs et décideurs. L'exercice de la puissance publique doit néanmoins rester attribuable à une autorité identifiable et responsable.

Si, dans une atteinte à la liberté assistée par IA, il devient impossible de déterminer qui répond de la décision, une exigence fondamentale de l'État de droit est affaiblie. La gouvernance doit donc rendre la responsabilité visible au lieu de la laisser se dissoudre dans le système.^{6 13}

4 Discussion : efficacité et État de droit

L'IA peut analyser rapidement de grandes quantités de données, favoriser une certaine cohérence et soulager des processus surchargés. Ces avantages potentiels sont particulièrement pertinents dans les fonctions d'enquête et d'administration.

Ils doivent toutefois être mis en balance avec des risques importants pour l'État de droit. Les décisions entraînant de graves atteintes aux droits fondamentaux — détention, constatation de culpabilité ou fixation de la peine — exigent un jugement humain, de la transparence et une responsabilité clairement attribuée. Dans ces contextes, l'IA ne peut au mieux jouer qu'un rôle d'assistance.

Une voie possible réside dans un modèle hybride : l'IA soutient l'analyse et la préparation, tandis que la décision finale et la responsabilité demeurent humaines. Un tel modèle suppose des règles claires en matière de transparence, de documentation, de contrôlabilité et de supervision humaine. Le règlement européen sur l'IA qualifie lui aussi plusieurs usages en matière de poursuite pénale et de justice de systèmes à haut risque et leur impose des exigences de gouvernance et de supervision.^{5 6 10}

5 Conclusion

L'usage de l'intelligence artificielle dans le droit pénal suisse offre un potentiel d'efficacité tout en posant des défis considérables aux principes du droit pénal et de l'État de droit.

Une délégation complète de la décision pénale à l'IA est incompatible avec la culpabilité individuelle, les garanties du procès équitable et une justice dont la responsabilité doit pouvoir être attribuée. L'IA ne peut donc être justifiée que comme outil auxiliaire, intégré dans un cadre juridique clair et soumis à une supervision humaine effective.

La conclusion centrale n'est donc pas que l'IA n'a aucune place dans la justice pénale. Elle est plutôt que l'efficacité ne peut être achetée au prix de la responsabilité. Les droits individuels, la transparence, la contrôlabilité et l'attribution de la responsabilité doivent demeurer intacts lorsque des systèmes numériques interviennent dans des décisions publiques lourdes de conséquences.^{1 3 4 5}

SOURCES ET BASES JURIDIQUES

¹ Constitution fédérale suisse (Cst.), art. 8 et 29 / Constitución Federal Suiza, arts. 8 y 29.

<https://www.fedlex.admin.ch/eli/cc/1999/404/en>

² Code pénal suisse (CP), art. 47 / Código Penal Suizo, art. 47.

https://www.fedlex.admin.ch/eli/cc/54/757_781_799/en

³ Code de procédure pénale suisse (CPP), art. 352-354 / Código de Procedimiento Penal Suizo, arts. 352-354.

<https://www.fedlex.admin.ch/eli/cc/2010/267/en>

- 4 Convention européenne des droits de l'homme, art. 6 / Convenio Europeo de Derechos Humanos, art. 6.
<https://www.echr.coe.int/en/web/echr/european-convention-on-human-rights>
- 5 Règlement (UE) 2024/1689 sur l'intelligence artificielle / Reglamento (UE) 2024/1689 de Inteligencia Artificial.
<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- 6 CEPEJ, Charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires (2018).
<https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>
- 7 Harry Surden, "Machine Learning and Law", 89 Washington Law Review 87 (2014).
<https://digitalcommons.law.uw.edu/wlr/vol89/iss1/5/>
- 8 Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms", Big Data & Society 3(1) (2016).
<https://doi.org/10.1177/2053951715622512>
- 9 Solon Barocas & Andrew D. Selbst, "Big Data's Disparate Impact", 104 California Law Review 671 (2016).
<https://doi.org/10.15779/Z38BG31>
- 10 Raja Parasuraman & Victor Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse", Human Factors 39(2), 230-253 (1997).
<https://web.mit.edu/16.459/www/parasuraman.pdf>
- 11 Tania Sourdin, "Judge v Robot? Artificial Intelligence and Judicial Decision-Making", 41(4) UNSW Law Journal 1114 (2018).
<https://www.unsw.edu.au/content/dam/pdfs/law/unsw-law-journal/2010-2019/Vol-No-41-4-Sourdin.pdf>
- 12 Conseil fédéral suisse, « Artificial Intelligence and International Rules » (2022).
<https://www.admin.ch/gov/en/start/documentation/media-releases.msg-id-88019.html>
- 13 Office fédéral de la justice, aperçu de la réglementation de l'IA en Suisse / Oficina Federal de Justicia, regulación de IA en Suiza.
<https://www.bj.admin.ch/bj/en/home/staat/gesetzgebung/kuenstliche-intelligenz.html>