

L'impossible droit à l'erreur, l'impossible droit à l'oubli ?

Nos traces numériques – données de navigation, paiements, trajets –, devenues indélébiles, sont désormais agrégées pour dessiner des profils de risque ou des scores de solvabilité. Or, cette mémoire algorithmique doit être mieux régulée pour éviter qu'elle ne constitue un « stigmate définitif » attaché au parcours d'un individu et assurer droit à l'oubli et à une seconde chance.



ARTHUR CHARPENTIER
Professeur de Mathématiques,
Université du Québec à Montréal
(UQAM)

« L'information publique est comme du dentifrice ; une fois sortie du tube, impossible de la faire rentrer », affirmait Tony Doyle en 2010. Quinze ans plus tard, dans un monde désormais entièrement numérique, chaque action laisse une trace, qu'il s'agisse d'un paiement différé, d'une recherche en ligne ou d'un trajet géolocalisé. Agrégées, ces traces dessinent des profils de risque, des scores de solvabilité ou des réputations durables. Mais une société juste ne peut se réduire à la mémoire intégrale : le droit à l'erreur et à l'oubli sont des conditions de la seconde chance. La sagesse japonaise enseigne qu'« on tombe sept fois, on se relève huit » (七転び八起き) : l'humain avance par essais et erreurs, à condition que ses fautes ne soient pas figées en stigmate définitif.

Les enfants et les machines apprennent de leurs erreurs

L'ironie de notre âge numérique tient à ce renversement : nous avons bâti des intelligences artificielles capables de progresser par itération, d'ajuster leurs modèles à partir de chaque échec, alors même que la société tend à figer les humains dans la moindre faute. On le voit d'autant plus dans l'enseignement supérieur, où la « norme » semble être un parcours linéaire, parfait, sans faute. Les algorithmes d'apprentissage automatique (qu'ils évaluent un risque de crédit, une prime d'assurance ou la probabilité d'une rechute médicale) se perfectionnent grâce à la conservation exhaustive des traces passées. Leur performance dépend d'une mémoire sans trou : rien n'est jamais effacé, tout sert à recalibrer la prédiction suivante. Pour la personne évaluée, cette

mémoire totale transforme chaque faux pas en stigmate durable. Le paiement en retard, l'oubli d'une déclaration, la consultation médicale jugée « à risque » sont agrégés, scorés, monétisés ; ils surgissent plus tard dans un refus de prêt ou une surprime. Ainsi, la logique actuarielle traditionnelle (qui mutualise l'aléa) laisse la place à une micro-tarification où le passé individuel décide du futur.

Ce verrouillage heurte deux intuitions fondatrices d'une société juste : le droit à l'erreur, qui reconnaît la perfectibilité humaine, et le droit à l'oubli, qui permet de se défaire d'un passé devenu sans pertinence. Il heurte aussi la pédagogie élémentaire : dès l'école primaire, l'apprentissage repose sur l'essai-erreur, la correction, la possibilité de recommencer sans être réduit à la faute initiale. Les sciences de l'éducation montrent qu'un enfant progresse lorsqu'il peut expérimenter, se tromper, recevoir un retour, puis réessayer. Refuser aux adultes (par le truchement d'algorithmes inflexibles) la latitude accordée aux enfants revient à nier la continuité du processus d'apprentissage tout au long de la vie.

Se pose alors une question éthique et juridique : comment concevoir des dispositifs de décision automatisée capables, comme l'être humain, d'intégrer l'erreur sans la fossiliser ? Autrement dit, comment faire place à la réversibilité (cette capacité de transformer la trace négative en simple étape) alors que la machine, pour apprendre, exige la persistance des données ? C'est dans cet écart entre mémoire algorithmique et droit à la seconde chance que s'inscrit toute la problématique de l'« impossible » droit à l'erreur et à l'oubli.

Le droit à l'erreur confronté à la froideur du score automatisé

Le droit français a consacré, avec la loi pour un État au service d'une société de confiance, dite « loi ESSOC », de 2018, la possibilité pour un administré de rectifier une déclaration inexacte sans encourir immédiatement de sanction. Par nature, cette garantie suppose une médiation humaine : c'est en exposant sa bonne foi que l'on obtient la bienveillance de l'administration. Or les systèmes de notation utilisés par les banques ou les assureurs figent la moindre entorse sous forme de variable négative. Une omission déclarative, même corrigée, continue d'alimenter un coefficient défavorable, car l'historique, conçu pour être exhaustif, n'intègre pas la nuance morale de la réparation. Ainsi, un micro-entrepreneur peut se voir refuser un prêt malgré la correction rapide de son oubli : la mécanique de score refuse tout dialogue et neutralise le droit à l'erreur.

Parmi les systèmes classés « haut risque » figure la notation de crédit. Or, comme l'a montré Cathy O'Neil dans *Weapons of Math Destruction* (2016), ces modèles « boîte noire » peuvent imputer à tort des défauts ou des retards inexistant : un homonyme ou une adresse partagée suffit parfois à faire baisser le score d'un assuré alors même qu'il n'a jamais manqué un paiement. En l'absence de procédure de recours transparente, l'erreur se transforme en stigmate financier durable pour la personne concernée. Pour préserver l'esprit de la loi, il faudrait imposer la capacité d'amnistie progressive des données : lorsqu'une faute a été réparée, les variables qui la traduisent devraient décroître jusqu'à

disparaître. Cette exigence pourrait s'adosser au règlement général sur la protection des données (RGPD), dont l'article 22 interdit les décisions exclusivement automatisées produisant des effets juridiques significatifs. Sans intervention humaine, la notion même d'erreur périssable s'évanouit au profit d'un passé figé dans la donnée.

Le droit à l'oubli, rempart fragile contre la mémoire illimitée

Institué par l'article 17 du RGPD, le droit à l'effacement paraît, de prime abord, solide. Il impose la suppression des informations personnelles devenues obsolètes ou disproportionnées. Dans le secteur de l'assurance emprunteur, la convention Aeras en offre une illustration : un cancer guéri depuis plusieurs années ne doit plus être déclaré. Pourtant, à l'échelle d'Internet, la suppression reste théorique : les copies et les miroirs rendent l'effacement incertain.

Mais il s'agit également d'un impératif éthique : comme le rappelle le philosophe Luciano Floridi, l'identité numérique est un capital sémantique dont on doit pouvoir maîtriser l'actualisation. L'oubli n'est pas la négation de la mémoire, il en est la « contrepartie pharmacologique », au sens de Bernard Stiegler : un remède à l'hypermnésie technique qui, faute de limites, transforme toute trace en stigmate. Accorder aux anciens malades, aux conducteurs repentis ou aux entrepreneurs faillibles la possibilité de « reprendre date » revient à reconnaître ce que Jacques Derrida appelait le pardon : non pas l'amnésie, mais la faculté de constituer un futur non asservi au passé. Les risques se multiplient avec l'émergence des intelligences

artificielles génératives. En recyclant des archives parfois non vérifiées, ces modèles peuvent ranimer un article diffamatoire ou amplifier une rumeur en la présentant comme un fait. Un justiciable relaxé voit ressurgir une procédure classée ; une militante d’autrefois retrouve son nom associé à un délit amnistié. La faculté d’oubli est alors sapée par la reproduction sans fin des traces. Pour rééquilibrer, il devient nécessaire d’intégrer, au cœur même des architectures d’IA, des mécanismes de dépréciation temporelle et d’alerte : toute référence à des données personnelles sensibles devrait déclencher une vérification humaine avant publication.

Le préjudice disproportionné comme boussole de la régulation

En Europe, le RGPD fonde la licéité d’un traitement sur une balance d’intérêts : l’impact sur la personne ne doit jamais excéder la finalité poursuivie. Dans la réalité algorithmique, cette proportionnalité s’efface. Les outils de scoring éducatif, par exemple, croisent taux de décrochage scolaire, codes postaux ou indicateurs de fragilité sociale ; ils aboutissent à écarter un candidat brillant sous prétexte qu’il réside dans une zone défavorisée. Le contraste est encore plus marqué si l’on traverse l’Atlantique. Dans de nombreux États américains, les antécédents judiciaires sont considérés comme des documents publics. Cette transparence se double d’un commerce florissant : des courtiers en données agrègent ces dossiers,

puis les revendent à des employeurs, des bailleurs ou des compagnies d’assurance, de sorte qu’une simple arrestation (même classée sans suite) peut se muer en exclusion durable. Dans cet environnement, le préjudice disproportionné n’est plus l’exception : il devient la règle, parce que l’information sur l’accusation (encore incertaine) est traitée avec le même poids qu’une condamnation définitive. L’algorithme, nourri d’une telle base, reproduit l’asymétrie : il neutralise la présomption d’innocence et pérennise la sanction sociale. Pour redonner chair à la proportionnalité, deux leviers se dessinent. D’abord, l’audit obligatoire et public des algorithmes à fort impact, afin de mesurer leurs biais et d’en dévoiler la logique. Ensuite, l’instauration d’un contrôle humain systématique

Toute référence à des données personnelles sensibles devrait déclencher une vérification humaine avant publication.

lorsqu’une décision négative se fonde sur un score. Sans ces garde-fous, le droit au pardon et à la seconde chance restent lettre morte, confisqués par la

mécanique froide d’une mémoire numérique sans prescription.

L’IA générative et l’illusion de la vérité reconstituée

Les modèles génératifs extraient des milliards de tokens pour produire des récits fluides, cohérents, immédiatement convaincants. Cette virtuosité stylistique masque pourtant une faiblesse structurelle : l’incapacité à distinguer un fait avéré d’une hallucination statistique. Les premiers contentieux montrent l’ampleur du risque. En mai 2025, la justice de l’État de Géorgie a dû examiner l’affaire *Walters v. OpenAI*, née d’une fausse accusation inventée par ChatGPT ; trois mois plus tôt, l’activiste Robby Starbuck avait poursuivi Meta AI pour avoir affirmé qu’il avait participé à l’assaut du Capitole. Ces dossiers illustrent la « déformation à grande échelle » : une information infondée, dès qu’elle circule dans un modèle puissant, acquiert la force persuasive d’un fait. L’AI Act tout juste adopté impose, à l’article 50, que les contenus synthétiques soient signalés comme artificiels et, à l’article 52, que les fournisseurs de grands modèles publient des résumés des données d’entraînement. Encore faut-il traduire ces principes dans la pratique : tout passage généré qui fait référence à une personne identifiable devrait être relu et validé par un humain avant d’être versé dans un dossier. Tant que cet ultime filtre n’est pas systématique, la puissance narrative de l’IA continuera à transformer des approximations en verdicts sociaux.

Le modèle du scoring social : vers une prédestination normative

La Chine a institutionnalisé un système de crédit social. Des pratiques semblables infiltrent désormais les économies occidentales. Dans l’assurance automobile, les boîtiers télématiques enregistrent vitesse, freinages brusques, trajets de nuit et localisation pour établir un « *driver score* ». Les fintechs de crédit brassent jusqu’à 20 000 données par individu pour calculer la probabilité de défaut. D’un point de vue européen, cette logique heurte frontalement le principe de finalité inscrit à l’article 5 du RGPD. Elle viole aussi le principe de minimisation. Conscient du danger, le législateur de l’Union a gravé dans l’AI Act un interdit clair : tout système de « *social scoring* » est classé à « risque inacceptable », et simplement interdit. Réaffirmer le droit à l’erreur et à la singularité suppose donc de proscrire la note globale automatique et de circonscrire les usages sectoriels. Sans ces garde-fous, l’addition de micro-fautes se mue en stigmate permanent. L’article 5 § 1 e) du RGPD rappelle que les données personnelles ne peuvent être conservées indéfiniment : elles doivent l’être seulement pour une durée proportionnée à la finalité poursuivie. Dans le secteur financier, cette exigence est partiellement satisfaite par la supervision prudentielle, mais encore faut-il que les algorithmes utilisés soient transparents. D’où la nécessité d’un règlement sur l’intelligence artificielle, dont l’efficacité dépendra de sa mise en œuvre effective : un défaut de contrôle pourrait, dans l’Union, produire les mêmes effets que l’absence de régulation aux États-Unis ou en Chine.

Le casier judiciaire et la double peine numérique

Le casier judiciaire est historiquement un registre interne à la justice. À l’ère des bases interconnectées, il devient cependant un référentiel de réputation permanente. Ainsi, une condamnation – parfois mineure, parfois ancienne – peut refaire surface lors d’une candidature ou d’une demande de visa, bien après que la peine a été exécutée. La diffusion non contextualisée de ces données contrevient aux principes de finalité et de minimisation. Le résultat est une forme de double peine numérique où la sanction initiale se transforme en obstacle invisible, mais durable, à l’emploi ou au logement. Une réforme équilibrée supposerait l’effacement automatique ou la mise sous scellés des mentions après des délais différenciés selon la gravité de l’infraction. Sans ces garde-fous, la fonction probatoire du casier dégénère en mécanisme d’exclusion systémique.

Mémoire, oubli et identité : quelques repères philosophiques

Hannah Arendt rappelait que la liberté humaine suppose une part de retrait, un espace de l’intime non exposé en permanence au regard d’autrui. Paul Ricoeur, dans *La mémoire, l’histoire, l’oubli*, montrait que l’oubli n’est pas l’ennemi de la mémoire mais sa condition d’équilibre : sans effacement, la mémoire devient encombrement. Maurice Halbwachs soulignait, de son côté, que toute mémoire est sociale et sélective : la société filtre et réordonne les traces pour qu’elles fassent sens. Or la mémoire

numérique, intégrale et indéfinie, ne connaît pas ces filtres. Viktor Mayer-Schönberger parlera d’« obésité informationnelle », où trop de traces tue la signification. Jean-Jacques Rousseau, déjà, faisait de l’erreur une étape nécessaire de l’apprentissage : *Émile, ou De l’éducation* (1762) posait le droit de se tromper comme condition de la formation d’un individu autonome.

L’oubli, loin d’être une perte, est un outil de liberté et de création.

Ces réflexions rappellent que l’oubli, loin d’être une perte, est un outil de liberté et de création.

Pour une société de la réversibilité

Au demeurant, le principe du droit à l’oubli est beaucoup plus solidement inscrit dans le droit européen que dans les autres grandes juridictions, ce qui confère à l’UE une responsabilité particulière de le faire vivre. Mais il faut garder en mémoire que sans régulation, la société calculatoire risque de figer chacun dans la somme de ses erreurs passées. Restaurer un droit effectif à l’erreur et à l’oubli impose d’inscrire l’amnistie temporelle dans les bases de données, d’obliger les

algorithmes à la proportionnalité et de rendre compte publiquement de leurs biais. Plus largement, il s’agit de construire une « assurance de la seconde chance » : un écosystème où le passé informe sans condamner.



Références

Arendt, Hannah. *The Human Condition*. University of Chicago Press, 1958.

Doyle, Tony. « Information in the Public Domain Is Like Toothpaste ». *The Irish Examiner*, 15 juin 2010.

Floridi, Luciano. *The Ethics of Information*. Oxford: Oxford University Press, 2013.

Halbwachs, Maurice. *La mémoire collective*. Paris : Presses Universitaires de France, 1950.

Mayer-Schönberger, Viktor. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press, 2009.

O’Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing, 2016.

Ricoeur, Paul. *La mémoire, l’histoire, l’oubli*. Paris : Seuil, 2004.

Rousseau, Jean-Jacques. *Émile, ou De l’éducation*. 1762.