

# Abstract

AI systems trained on personal data create a real problem for data protection rules: once information becomes part of a model’s internal parameters, can it actually be removed, and can the model’s decisions be properly explained to people? This research addresses that question by comparing two legal systems – the GDPR and Algeria’s Law 18-07 – alongside the technical methods currently available.

The first part examines how both laws treat AI-related rights, focusing mainly on the right to erasure and the right to explanation. Each right is mapped onto a seven-step machine learning process, from data collection to deployment. This comparison reveals that most rights can be applied in practice, but erasure and explanation turn out to be by far the hardest to achieve technically.

The second part reviews machine unlearning techniques – full retraining, SISA, certified removal, gradient ascent, and knowledge distillation – together with the methods used to verify whether unlearning actually worked, such as membership inference attacks, shadow-model tests, and likelihood-ratio approaches. No single technique manages to be cheap, easy to interpret, and resistant to manipulation all at once. The analysis also covers explanation tools like SHAP, LIME, Integrated Gradients, and counterfactual explanations, showing that being mathematically sound doesn’t automatically mean a method meets legal standards.

The main finding is that unlearning methods sit somewhere on a scale between cost and reliability, but a dependable way to confirm that erasure truly happened is still missing. Explanation methods, while technically solid, still lack any agreed standard for what counts as legally sufficient under either the GDPR or Law 18-07.

**Keywords**— Data Protection Law, GDPR, Algeria Law 18-07, Machine Unlearning, Explainable AI, Membership Inference, Right to be Forgotten, Right to Explanation, Privacy by Design

## ملخص

تطرح النماذج المدربة على البيانات الشخصية إشكالية حقيقية بالنسبة لقواعد حماية البيانات: فبمجرد أن تصبح المعلومات جزءاً من الأوزان الداخلية (parameters) للنموذج، هل يمكن حقاً حذفها؟ وهل يمكن تفسير قرارات النموذج بشكل مفهوم للأفراد؟ يتناول هذا البحث هذا السؤال من خلال مقارنة قانونية وتقنية بين إطارين: اللائحة الأوروبية العامة لحماية البيانات (GDPR) والقانون الجزائري رقم 07-18.

يبدأ البحث بمقارنة كيفية تعامل كلا الإطارين مع الحقوق المرتبطة بالذكاء الاصطناعي، وبالأخص الحق في المحو (right to erasure) والحق في التفسير (right to explanation)، خلال إسقاط كل حق على مراحل خط أنابيب التعلم الآلي (machine learning pipeline) السبع، بدءاً من جمع البيانات وصولاً إلى النشر. تُظهر هذه المقارنة أن معظم الحقوق قابلة للتطبيق تقنياً، في حين يبقى الحق في المحو والحق في التفسير الأصعب تحقيقاً من الناحية التقنية.

بعد ذلك، يستعرض البحث تقنيات "النسيان الآلي" (machine unlearning) مثل: إعادة التدريب الكامل (exact retraining)، وطريقة SISA، والحذف الموثق (certified removal)، والصعود التدريجي (gradient ascent)، وتقطير المعرفة (knowledge distillation)، إلى جانب أساليب التحقق من فعالية هذا النسيان، مثل هجمات استدلال العضوية (membership inference)، واختبارات النموذج الظل (shadow-model)، وطرق نسبة الاحتمالية (likelihood-ratio). تكشف هذه المراجعة أنه لا توجد طريقة واحدة تجمع في آن واحد بين انخفاض التكلفة، وسهولة الفهم، ومقاومة التلاعب. كما يستعرض البحث أدوات التفسير مثل SHAP و LIME والتدرجات المتكاملة (Integrated Gradients) والتفسيرات المضادة للواقع (counterfactual explanations)، مبرزاً الفجوة القائمة بين السلامة الرياضية لهذه الأدوات وكفائتها من الناحية القانونية.

يخلص البحث إلى أن طرق النسيان الآلي تتوزع على مقياس يجمع بين التكلفة والضمانات، غير أن مسألة التحقق من نجاح عملية الحذف تبقى دون حل واضح. أما أساليب التفسير، فرغم رسوخها الرياضي، فإنها لا تزال تفتقر إلى معايير متفق عليها لتحديد الكفاية القانونية سواء بموجب اللائحة الأوروبية GDPR أو القانون الجزائري 07-18.

**كلمات مفتاحية:** قانون حماية البيانات، اللائحة الأوروبية العامة لحماية البيانات (GDPR)، القانون الجزائري رقم 07-18، النسيان الآلي، الذكاء الاصطناعي القابل للتفسير، استدلال العضوية، الحق في النسيان، الحق في التفسير، الخصوصية حسب التصميم

# Résumé

Les systèmes d'intelligence artificielle entraînés sur des données à caractère personnel soulèvent un véritable problème au regard des règles de protection des données : dès lors qu'une information est intégrée aux paramètres internes d'un modèle, peut-elle réellement être supprimée, et les décisions produites par ce modèle peuvent-elles être expliquées de manière compréhensible aux personnes concernées ? Cette recherche aborde cette problématique à travers une comparaison entre deux cadres juridiques, le Règlement général sur la protection des données (RGPD) et la loi algérienne 18-07, ainsi qu'une analyse des méthodes techniques actuellement disponibles.

La première partie étudie la manière dont les deux cadres juridiques prennent en compte les droits liés à l'intelligence artificielle, en se concentrant principalement sur le droit à l'effacement et le droit à l'explication. Chaque droit est mis en relation avec les sept étapes d'un processus d'apprentissage automatique, depuis la collecte des données jusqu'au déploiement du modèle. Cette comparaison montre que la plupart des droits peuvent être pris en compte sur le plan technique, mais que l'effacement et l'explication demeurent de loin les plus difficiles à mettre en œuvre.

La deuxième partie porte sur les techniques d'oubli automatique, notamment le réentraînement complet, SISA, la suppression certifiée, la remontée de gradient et la distillation des connaissances. Elle examine également les méthodes permettant de vérifier si l'oubli a effectivement été réalisé, telles que les attaques par inférence d'appartenance, les tests fondés sur des modèles auxiliaires et les approches par rapport de vraisemblance. Aucun de ces procédés ne permet à lui seul de réunir un faible coût, une interprétabilité satisfaisante et une résistance suffisante à la manipulation. L'analyse porte également sur des méthodes d'explication telles que SHAP, LIME, Integrated Gradients et les explications contrefactuelles, en montrant qu'une méthode peut être mathématiquement solide sans pour autant satisfaire aux exigences juridiques.

Les résultats montrent ainsi que les techniques d'oubli automatique se situent sur un compromis entre coût et fiabilité, tandis qu'il n'existe toujours pas de méthode véritablement fiable permettant de confirmer qu'un effacement a été effectué de manière effective. De même, malgré leur solidité technique, les méthodes d'explication ne disposent pas encore de critères communément admis permettant de déterminer ce qui constitue une explication juridiquement suffisante au regard du RGPD ou de la loi algérienne 18-07.

## *Mots-clés :*

Droit de la protection des données, RGPD, loi algérienne 18-07, oubli automatique, intelligence artificielle explicable, inférence d'appartenance, droit à l'oubli, droit à l'explication, protection de la vie privée dès la conception