



WHITE PAPER

Confidentiality and legal artificial intelligence

Anonymisation and pseudonymisation of legal matters
A reference framework for the legal profession

Analysis of French and Californian legal texts
Design principles and evaluation method

How can the exposure of information entrusted to an AI system be reduced while preserving what is necessary for legal reasoning and the restoration of information in the result?

An AIDocs Studio publication · Max
October 2026 · First revised edition

Summary and method

Summary

The confidentiality of a legal matter depends on the information that circulates, its recipients and the uses made of it. Replacing names is a useful measure, but its value must be assessed alongside the remaining content and the work required. A transformation that protects an identity while altering a capacity, a date or an obligation may make the result legally unsuitable.

This white paper proposes a framework for examining information protection, contextual fidelity and the restoration of information in outputs together. It draws on French rules on professional secrecy, the recommendations of the Conseil national des barreaux (CNB), the GDPR, California Senate Bill 574 and institutional references on de-identification. [1–10]

Its central conclusion is that the relevant unit of assessment is the processing workflow. Protection must be examined at intake, during document preparation and successive processing stages, and when information is restored in the output. An isolated check on a text does not, by itself, describe the properties of the whole workflow.

Method and status of the proposals

The approach compares a targeted body of legal texts, professional recommendations and institutional research, selected for their direct relevance to the confidentiality of legal matters. It is supplemented by a decision model and an evaluation protocol; it is not a systematic literature review. The discussion distinguishes the requirements of the texts, their interpretation for AI use and the methodological proposals of this publication. Numbered references provide access to the sources; the methodological proposals are those of this publication.

The examples are fictional. The frameworks and indicators are working tools to be adapted; they are not presented as standards adopted by the institutions cited. The study focuses primarily on lawyers' matters. Other professionals may use the method while taking account of the secrecy duties and obligations specific to their status.

Proposed contributions

A distinction between identity protection, content confidentiality and legal fidelity.

An examination of the complete workflow, from request to deliverables.

A method for choosing safeguards and examining their results for a defined use case.

The French foundations of protection

Professional secrecy covers the matter file

Article 66-5 of the Law of 31 December 1971 protects, in particular, legal advice, correspondence and documents in the matter file. Article 2 of the French National Rules of Professional Conduct for Lawyers (RIN) specifies its scope. Professional secrecy therefore covers information extending beyond data that identifies a natural person. [4][5]

This distinction is decisive for AI. Litigation strategy, a proposed settlement or know-how may remain confidential after names have been removed. The question to examine is what content is actually communicated and under what processing conditions.

The CNB recommendation

The guide *La déontologie et l'intelligence artificielle*, first edition published in April 2026, recommends prior pseudonymisation or anonymisation, limiting context to what is necessary and vigilance regarding indirect identification. It also recalls the constraints specific to particular uses and processing environments. [2]

The guide translates professional obligations into AI practices. Its recommendations must be distinguished from the statutory provisions and the RIN to which it refers. Implementation requires an organised approach: authorised uses, preparation of requests and validation of results.

The role of the GDPR

For personal data, pseudonymisation is among the protective measures envisaged by the GDPR, particularly for data protection by design and the security of processing. Obligations concerning purpose, minimisation, arrangements with service providers and, where applicable, transfers must be examined in the context of the processing. [9]

Implications for professional decisions

Three complementary assessments are needed: whether the data is personal, whether the information is confidential and whether its use is necessary. Data irrelevant to the task may be excluded; necessary data may require transformation; content whose value depends on secrecy may require a more restricted environment.

The choice therefore does not follow from a single label attached to a document. It follows from the task, the information needed and the chosen trust boundary.

The contribution of California Senate Bill 574

SB 574, Chapter 858 of the Statutes of 2026, was signed into law on 30 September 2026. It is expected to take effect on 1 January 2027 under the ordinary constitutional timetable. [1][3][6]

The new section 6068.1 of the Business and Professions Code prohibits delegating the practice of law to generative AI. It regulates the input of confidential, identifying or non-public information by restricting access to the lawyer and authorised persons subject to confidentiality obligations. It also requires reasonable verification of outputs and correction of errors. AI use must be disclosed to the court for the covered documents submitted to it; disclosure must be considered for content intended for the public. [1]

Section 128.7 of the Code of Civil Procedure, as amended, requires the lawyer responsible for a filing to personally verify citations, including those produced by AI. This analysis concerns lawyers; the provisions concerning arbitrators and judicial institutions are not discussed in detail. The legislation does not prescribe general anonymisation. [1]

Two regimes with shared concerns

The French and Californian texts are not interchangeable. The former combine professional secrecy, data protection and practice recommendations. The latter notably establish an access restriction requirement for the uses they cover. An activity subject to several jurisdictions must be assessed under each of them.

Comparing them nevertheless reveals common design questions: what information is communicated, to whom does it become accessible and how does the professional retain control over the result? This comparative reading focuses on the problems to be solved without treating the rules as uniform.

What pseudonymisation contributes to this analysis

Pseudonymisation can reduce the exposure of identities in the processing concerned. Access conditions and the confidentiality of the remaining content constitute a separate level of assessment. It is useful to document these dimensions separately: a measure applied to identifiers does not, in itself, describe access permissions.

A rigorous evaluation therefore examines the properties of processing rather than matching terminology. Terms such as “private”, “professional” or “secure” must be explained by reference to the scope of the data, the operations performed and the applicable commitments.

Defining safeguards without conflating them

The following terms describe different properties. Distinguishing them makes it possible to choose a measure according to the intended result and to define what must be verified. [7][8][9]

Measure	Intended property	Implication for the work
Anonymisation	Making identification irreversibly impracticable, taking account of the means reasonably likely to be used.	Suitable for a use that does not require a return to the identities.
Pseudonymisation	Separating data from identity through protected additional information.	Allows controlled re-attribution; restoration in the deliverable still needs to be checked.
Minimisation	Limiting data to what the task requires.	Reduces the context to be processed and the information exposed.
Access restriction	Limiting access to authorised recipients.	Governs the use of originals and transformed content.

Reversibility is a functional choice

A publication may require an anonymous result; a legal instrument may require the parties to be identified again. Reversibility is then a functional option, not an inherently superior quality. Separating the re-attribution information and restricting access to it are part of pseudonymisation within the meaning of the GDPR. [9, Art. 4(5)]

Context determines the risk of identification

In assessing anonymisation, the CNIL distinguishes, in particular, the risks of singling out, linkability and inference. The possibility of recognising a person also depends on supplementary information available. Assessment must therefore consider combinations of clues and the recipients. [7]

For a law firm that retains the means of re-attribution, replacing identities alone does not take the processing outside the scope of personal data protection. [8][9]

A risk model suited to legal matters

The proposed model distinguishes four dimensions. It helps prepare the analysis of a use case; it does not assume knowledge of a provider's internal processes.

Direct identifiers

Names, contact details and individual references may immediately connect a text to a person or an organisation. Their presence must be examined in the request and associated content, including file names and comments relevant to the processing.

Recognition clues

An unusual role, a location, an amount and a period may make a transaction recognisable. Each element considered separately may seem general; together they may be distinctive. Assessment must consider the combination of information retained and the recipient's knowledge.

Information that is confidential by its content

A reservation price, a possible concession or a procedural weakness has value independently of the client's name. Its protection also depends on access boundaries and whether it needs to be transmitted. A method focused solely on identities would miss this dimension of the matter.

Circulation and duration

Information may reappear when a document is added, work is resumed or an intermediate summary is prepared. It may also appear in a log, a copy or a deliverable. What needs to be examined is every point at which the information changes recipient, form or purpose.

Formulating a testable risk hypothesis

A useful formulation specifies a recipient, content and a consequence: "Combining these three elements would allow this participant to recognise the transaction." It is then possible to examine whether the elements are necessary, whether they can be transformed and which access conditions are appropriate.

Institutional literature on de-identification highlights the trade-off between utility and protection, as well as the possibility of re-identification in some transformed datasets. This observation supports a contextual evaluation rather than a conclusion based solely on the presence of masking. [10]

Examining the complete information workflow

The following framework describes functional stages. It can be used for an integrated service or a combination of several tools. Each stage requires its content and recipients to be checked.

Stage	Central question	What to examine
Intake	What context enters the service?	Request, relevant exchanges, documents and initial access.
Preparation	Who accesses the content to make it usable?	Extraction, text recognition and services involved.
Transformation	What information is changed or excluded?	Data scope and consistency of preparation.
Processing	What context circulates between stages?	Research, analyses, intermediate results and dependencies.
Output delivery	How is a usable result recovered?	Response, files, parties' roles and restoration checks.
Retention	What remains after the work?	Documents, logs, persistent access and chosen retention periods.

Preparation is part of the scope

Legal processing is not necessarily the first point at which a service accesses content. Document conversion or text recognition may take place beforehand. Mapping includes these operations and their recipients, even if the content is subsequently pseudonymised.

Output delivery is a new confidentiality stage

A deliverable in which identities have been restored once again contains explicit identifiers; its content may already have been confidential before this restoration. Its recipients, storage location and access rights must be examined. Returning to the original information is a useful function whose conditions must be planned from the outset.

This mapping provides a common basis for legal teams, data protection officers and service designers. It connects each stage to data, recipients and observable evidence.

Preserving the fidelity of legal reasoning

What must remain accurate

The quality of a transformation is not simply a matter of how many elements have been replaced. Reasoning may depend on a chronology, a legal capacity or a relationship between two parties. Preparation must preserve these elements when changing them would alter the analysis.

Preserving a one-month period, a date of receipt or a rate may be decisive. Automatically generalising a date to a year or an amount to an order of magnitude may be appropriate for a statistical publication, but unsuitable for calculating a deadline or a claim.

What must remain consistent

The same person may appear as an agent in one document, a director in another and a signatory in a third. The distinction between the actor's identity and the capacity in which they act must be maintained. Connecting the roles is a legal property to preserve.

Consistency also concerns successive documents. An amendment may change only one obligation or only certain signatories. Transformation must allow these states to be compared without merging distinct actors or artificially separating the same participant.

What must be capable of restoration

Reversibility must be checked on the transformed elements and the media actually used. Faithfully recovering a prepared text is a different check from restoring information in a generated document: the latter may contain new wording, tables or comments.

Two dimensions to assess together

Protection examines what can be learned about individuals and the matter. Fidelity examines what still allows correct reasoning and the production of a usable result. A suitable solution must be assessed along both dimensions for the defined use case.

The evaluation protocol must therefore record both the information requiring protection that remains and transformations that degrade meaning. Restoration completes the checks; it replaces neither of them.

Choosing safeguards according to the need

The proposed decision model starts with the work required. It allows several responses to be compared, including reducing context before resorting to any transformation.

1 Can the question be abstracted?

To understand a general legal mechanism, stating the applicable law and a typical situation may be sufficient. An abstract request avoids communicating matter-specific elements. It is suitable as long as the specificity of the answer does not depend on omitted facts.

2 What data is necessary?

For the information retained, specify its function: identifying a role, establishing a fact, calculating a deadline or interpreting a clause. An element with no function in the task may be removed. This step reduces the scope to be protected and clarifies the need for context.

3 Does the need require a return to identities?

Where the result must be returned to the matter with its parties identified, reversible pseudonymisation is an option to consider. For external distribution without a return to identities, the objective of anonymisation requires a separate assessment of the identifiability of the complete output. [7][8]

4 Does the content retain its own sensitivity?

Strategic information may remain sensitive after identifiers have been transformed. Keeping it depends on the task's needs and the safeguards of the environment. The choice may involve further reducing context or processing it under more restricted access.

5 Can the result be validated?

Define who will check the restored content, the references and the intended use. If the available context does not allow a sufficiently substantiated answer, supplement the request or reconsider the task before using the result.

The choice may combine several measures. An abstract question requires little context; a publication may require anonymisation; work requiring a return to the parties may require pseudonymisation. Processing without transformation requires an examination of access and confidentiality. Where suitable safeguards are not in place, reduce the task or choose another environment.

A protocol for evaluating results

A useful evaluation is reproducible and tied to a use case. The proposed protocol can be applied to a market offering or an internal system, using representative fictional matters.

Building a reference set

Include several related documents, participants with distinct roles, variations in names or descriptions and recognisable contextual elements. Include the media used in practice: request text, documents and relevant comments. Define in advance the elements to protect and those whose meaning must be preserved.

Examining protection and fidelity

For each case, compare the result with this reference. Record information requiring protection that remains, unnecessary transformations and changes in meaning. Record the sensitive elements correctly handled relative to the expected elements, then unjustified substitutions relative to all substitutions made. Distinguish these results from errors in roles, dates or amounts. Publish counts and the severity of errors: a percentage alone conceals critical cases.

Dimension	Observation to record
Protection	Expected elements correctly handled and sensitive information still present.
Fidelity	Roles, facts, relationships, dates and amounts preserved where necessary.
Restoration	Elements restored, mapping errors and unresolved content.
Coverage	Formats, stages and situations actually examined.

Also evaluating sequences

Use a result in a subsequent stage, add a document and check the final restoration. Observe behaviour when a document is incomplete or a transformation cannot be completed. The expected response to these situations must be defined within the scope of the use case.

Making the conclusion interpretable

Record the version, settings, date, cases, repetitions and corrections. Have disputed annotations reviewed by a second evaluator. Before testing, define the errors that would prevent use, then retest the relevant cases after correction. Fictional matters make tests shareable; they do not, on their own, demonstrate coverage of real matters.

Three practical cases and their lessons

These fictional cases illustrate how to apply the method. They do not describe the measured performance of a tool.

A contract and two amendments

The task is to determine which obligations still apply. The matter includes parties referred to in different ways and one person acting in two capacities. Preparation must allow the different designations to be linked while preserving capacities and chronology.

The check concerns the relationship between the texts: which clauses are amended, for which parties and on what date? Identities may be substituted during processing and then restored in the result. Decision criterion: confusion about a signatory or a date prevents use of the result, even if every name has been replaced.

A negotiation note without a client name

A note describes an unusual transaction and the minimum amount acceptable to the client. The absence of a name removes neither the possibility of recognising the transaction nor the strategic value of the amount. Recognition clues and content confidentiality must be examined separately.

If a general analysis is sufficient, the request may be reformulated with a more abstract context. If the exact amount is necessary, its transmission must be assessed in the chosen environment. Decision criterion: use the abstract version if it is sufficient; otherwise examine the processing safeguards for the amount and strategy before transmission.

A request followed by the production of a letter

An initial analysis is then used to draft a letter. Transformed information passes through several stages, and identities are restored in the deliverable. The check concerns the whole sequence: initial content, intermediate result, final document and recipients.

Preparation and restoration must avoid attributing a fact to the wrong party or reusing context that has become incomplete. Decision criterion: approve the letter only after checking attributions, restored information and its recipient. A correct intermediate result does not automatically validate the deliverable.

A common lesson

Appropriate protection depends as much on the relationships between information as on its form. Evaluation must start from the intended legal result and trace back through the stages that make it possible.

Organising responsibility and traceability

Defining a trust boundary

The firm identifies the categories of work entrusted to AI, the information involved and the authorised persons. Mapping recipients includes preparation services and the access needed to operate the service. Confidentiality commitments and the processing framework are examined with the relevant responsible parties. [9]

Distinguishing validations

Validation of preparation concerns the chosen context and safeguards. Validation of the result concerns facts, references and legal relevance. The transmission decision concerns the recipient and intended use. These validations may be performed by the same person, but their objects remain distinct.

Keeping a useful record

The record documents the task, the chosen scope, the checks and the final decision. It must allow the work to be understood without multiplying unnecessary copies of documents or confidential information. Access to records and their retention periods fall within the same assessment as those of the documents.

Keeping the firm's rules operational

A short procedure can specify common uses, situations requiring particular scrutiny and what to do when content is incomplete. Fictional examples help teams recognise issues of indirect identification and strategic confidentiality.

A review becomes useful when a use case, a format, a participant or a processing condition changes. Documentation must make it possible to determine which conclusions remain applicable and which checks must be repeated.

Preserving freedom of choice

Describing needs and criteria without depending on a supplier's terminology makes offerings easier to compare. The firm can request answers about the workflow, data, access, fidelity and restoration, then assess the evidence relevant to its uses.

This discipline also benefits the service designer: it specifies the expected properties and the situations that a demonstration must make observable.

A practical worksheet to adapt to each use

This worksheet summarises useful decisions and observations. It may be incorporated into a firm procedure, an evaluation file or a discussion with a provider.

The task

Expected result: _____

Relevant law and context: _____

Person responsible for validation: _____

The information

Necessary documents and content: _____

Data to remove or transform: _____

Confidential information to handle separately: _____

The workflow and recipients

Document preparation and participants: _____

Processing and authorised access: _____

Restoration, storage and deliverable recipients: _____

Retention of documents and records: _____

Properties to preserve

Essential roles, facts and relationships: _____

Decisive dates and amounts: _____

Elements to restore in the result: _____

Observations and decision

Representative cases examined: _____

Remaining sensitive information: _____

Changes in meaning or restoration errors: _____

Decision, conditions of use and date: _____

A use case is documented when the professional can explain what they expect from processing, what information they provide, which safeguards they choose and the basis for their decision to use the result.

Conclusions and perspectives

The choice of safeguards starts with the legal need and the information required. Abstraction, minimisation, anonymisation, pseudonymisation and access restriction serve different purposes and may be combined. Their suitability is judged on the actual workflow, content confidentiality and result fidelity, without presuming that one measure suits every matter.

Evaluating a workflow rather than an isolated text

The request, documents, preparation services, successive processing stages and restoration form a whole. The same information may change form and recipient several times. This continuity must be reflected in the mapping and testing.

Measuring protection alongside legal fidelity

Preserving capacities, facts and relationships is a quality criterion as necessary as handling identifiers. An evaluation must make protection errors and changes in meaning visible, then examine restoration in the media actually used.

Documenting transferable decisions

A clear description of the task, scope and criteria makes it possible for professionals to discuss choices. The proposed frameworks are intended to be adapted, tested against new cases and enriched through experience; they are not a certification standard. They may provide a common language for law firms, data protection officers and service designers.

Areas for further work

Evaluations would benefit from covering more multi-document matters, resumed contexts and output formats. A useful comparison would publish the cases, criteria and types of errors while protecting matter information. It would distinguish the properties of a component from those of the complete workflow.

About this publication

This white paper is published by AIDocs Studio, the publisher of Max. It offers a framework for reflection and methodological tools applicable to any legal AI service. The analyses are those of the publisher; the institutions cited are the reference sources. Documentary research current as of 16 October 2026.

Legal and professional references

[1] California • Senate Bill 574

Chapter 858 of the Statutes of 2026. Enacted text, in particular sections 1 and 3. Reproduced by LegiScan.

Source: legiscan.com

[2] Conseil national des barreaux • 2026

La déontologie et l'intelligence artificielle. Practical guide, first edition, April 2026, in particular pages 15–18.

Source: cnb.avocat.fr

[3] Governor of California • 30 September 2026

Official press release mentioning SB 574 among the bills signed into law.

Source: www.gov.ca.gov

[4] France • Law No. 71-1130 of 31 December 1971

Article 66-5, concerning lawyers' professional secrecy.

Source: www.legifrance.gouv.fr

[5] Conseil national des barreaux • RIN

Title I, Article 2: principles and scope of professional secrecy.

Source: cnb.avocat.fr

References on information protection

[6] Constitution of California

Article IV, section 8. Ordinary timetable for entry into force. The expected date for SB 574 is derived from this timetable. Institutional reproduction by the California courts.

Source: courts.ca.gov

[7] CNIL • 19 May 2020

L'anonymisation de données personnelles. Definition and risks of re-identification.

Source: www.cnil.fr

[8] CNIL • 13 January 2022

Recherche scientifique hors santé : enjeux et avantages de l'anonymisation et de la pseudonymisation. Reference for distinguishing the concepts.

Source: www.cnil.fr

[9] European Union • Regulation (EU) 2016/679

GDPR, in particular Article 4(5); Articles 5, 25, 28 and 32; Chapter V for relevant transfers.

Source: eur-lex.europa.eu

[10] Garfinkel S L • NIST • 2015

De-Identification of Personal Information. NISTIR 8053. Research overview on de-identification, its utility and re-identification risks. DOI: 10.6028/NIST.IR.8053.

Source: csrc.nist.gov