

Französische Datenschutzaufsicht über Datenannotation im KI-Training

Am 22. Juli 2025 veröffentlichte die CNIL (Commission Nationale de l'Informatique et des Libertés) ein [Factsheet](#) zur Datenannotationsphase bei der Entwicklung eines hochwertigen KI-Modells.

Die CNIL stellt dabei heraus, dass Datenannotation eine zentrale Grundlage in der Entwicklung eines qualitativen KI-Modells (sowohl in überwachten als auch nicht-overwachten Verfahren) ist: Jedem Datum wird eine Beschreibung (sog. "**Label**") zugefügt, das als "**Ground Truth**" gilt und anhand dessen das Modell die Daten verarbeitet, klassifiziert und unterscheidet.

- **Annotationstypen:** manuell, semi-automatisch, automatisch; auch Wiederverwendung bestehender Annotationen – z. B. aus medizinischen Diagnosen
- **Praxisbeispiele**
 - Stimmerkennung: Stimmen mit Sprecheridentität versehen
 - Sturzerkennung: Personenbilder mit „stehend“/„liegend“ labeln
 - Kennzeichenerkennung: Pixelbereiche für Nummernschilder markieren
 - Pathologie-Vorhersage: Blutwerte mit ärztlicher Diagnose annotieren
- **Datenschutzanforderungen an die Annotation:** **Minimierung** ("minimization"), **Richtigkeit** ("accuracy"), **Treu und Glauben** ("loyalty") als zentrale GDPR-Prinzipien
 - **Minimierungsprinzip:** nur für Modellzweck „adäquate, relevante, notwendige“ Labels nutzen; bei vordefinierten Datensätzen störende Annotationen aussortieren und begründen; Annotationen können zur Fehler- und Verzerrungskorrektur nützliche Kontextelemente (bspw. Datum, Uhrzeit, Wetter, etc.) enthalten - Vorsicht allerdings bei personenbezogenen Daten (wie Name, etc.) oder sensiblen Daten (wie Religion, Hautfarbe, etc.)
 - **Richtigkeitsprinzip:** Korrektheit (fall erforderlich auch Aktualität) der verarbeiteten Daten; Ergreifung geeigneter Maßnahmen für Sicherstellung objektiver Annotationen; Vermeidung unpräziser Labels (vor allem bestehend aus wenigen Worten), um diskriminierende Fehlklassifikationen zu verhindern

Vorschlag der CNIL eines kontinuierlichen Überprüfungsverfahrens zur Qualitätsüberprüfung der Annotationen

Die CNIL empfiehlt folgende Schritte zur Aufstellung eines **Annotations-Protokolls:**

- auf den Einsatzzweck zugeschnittene, faire und unmissverständliche Labels, die sich auf nützliche Informationen beschränken
- möglichst objektive und eindeutige Annotation -> keine abwertenden oder wertenden Begriffe
- Vermeidung von Voreingenommenheit oder möglicher Diskriminierung
- Bei Ergebnis eines Geschäftsprozesses: möglicherweise Sortier- oder Requalifizierungsphase der Labels

Die CNIL empfiehlt folgende Schritte zur Aufstellung eines **Annotations-Verfahrens**:

- Dokumentation des Verfahrens
- Klare Aufgabenzuweisung (Zugang nur für befugte Personen)
- Den annotierenden Personen Feedback-Möglichkeit einräumen
- Validierungsphase vorsehen
- Stichprobenanalyse der annotierten Daten zur Fehleraufdeckung
- Protokollierung von Änderungen
- Anwendung eines zuverlässigen, robusten und beherrschten Annotations-Werkzeugs (Beachte vor allem die Sicherheitsaspekte bei halbautomatischer Annotationsfunktion!)

Die CNIL empfiehlt folgende Schritte zur Aufstellung eines **kontinuierlichen Überprüfungsverfahrens**:

- Einführung kurz nach Annotation sowie Fortsetzung während der Annotation durch regelmäßige/kontinuierliche Überwachung
- Dokumentation - vor allem basierend auf: Diskussionsrunden zwischen Annotationsteam, Systementwicklungsteam und (sofern möglich) Systembenutzenden; Stichprobenanalysen; internes oder externes Audit; Relevanzanalyse der annotierten Daten; Feedback-Verfahren für die Nutzenden; Qualitätskontrollverfahren

Die CNIL empfiehlt außerdem den **Ethik-Referenten bzw. Ethikkommission** einzubeziehen. Der multidisziplinäre und unabhängige Charakter dieses Gremiums ermöglicht die:

- Wahl der besten Option für die Datenannotation
- Erstellung eines Annotationsprotokoll
- Überprüfung des Annotationsprotokolls
- Kontrolle von Qualität und Eignung der Annotationen

Die CNIL weist auf die Pflicht zur **Informierung der Betroffenen** über:

- Zweck der Annotation
- Verantwortlichen (besonderer Hinweis bei Verantwortlichkeiten/Datentransfer außerhalb der EU)
- Vertragliche Kriterien der sozialen Verantwortung zwischen Annotierenden und dem Verantwortlichen Datenverarbeiter (bspw. Garantien für Arbeitsbedingungen, Entlohnung, etc.)
- Sicherheitsmaßnahmen, v.a. während der Annotation

Die CNIL weist zudem auf die Betroffenenrechte hin - in der Regel stellen Annotationen personenbezogene Daten dar, sodass die Betroffenenrechte der DSGVO gelten:

- Auskunftsrecht
- Recht auf Berichtigung, Löschung, Widerspruch und Einschränkung
- Recht auf Übertragbarkeit gilt nur, sofern die Annotation von dem Betroffenen zur Verfügung gestellt wurde und die Verarbeitung auf einer Rechtsgrundlage oder einem Vertrag beruht (Bsp: Kein Übertragungsrecht bei Tonaufnahmen durch Sprachassistenten, in deren Verarbeitung zum Zwecke der Verbesserung der Betroffenen eingewilligt hat.)

In Abgrenzung Annotation/Profiling/autom. Entscheidung differenziert die CNIL dahingehend, dass die Annotation aus einer Klassifizierung, die als „Grundwahrheit“ für das Modell dient, besteht und in der Regel keine rechtlichen Auswirkungen für den Betroffenen hat. Das Ziel besteht grundsätzlich nicht darin, individuelle Merkmale zu bewerten oder zu beurteilen, und stellt daher kein Profiling i.S.v. Art. 4 DSGVO bzw. keine automatisierte Entscheidung i.S.v. Art. 22 DS-GVO dar.

Im Hinblick auf die **Annotation sensibler Daten** führt die CNIL aus, dass

- auch wenn das ursprüngliche Datum kein sensibles Datum ist, kann eine Annotation aber ein sensibles Datum sein
- Art. 9 DSGVO zu beachten ist und Ausnahmen zu dokumentieren sind
- objektive und sachliche Messkriterien anzuwenden sind (z. B. Hautfarbe nach RGB-System statt „ethnische Herkunft“) -> Empfehlung: Verwendung technischer Kommentierungswerkzeuge
- eine verstärkte Prüfung hinsichtlich einer erhöhten Datensicherheit (Verschlüsselung, Zugriffsbeschränkung) und bestehender Risiken erfolgen muss