Security and Data Management within AI features

Sécurité et gestion des données au sein des fonctionnalités IA

PUBLIC

v1.0 - April 2025

Table des matières

Summary

This document details the security and data management of Pennylane's AI features.

Data is hosted in the European Union by certified providers with a contractual commitment that data is never used for model training.

The infrastructure includes strict security measures including access control, data isolation between users, and data encryption. Penetration tests are performed twice a year.

The architecture consists of four levels: Pennylane frontend for display, backend for security and authentication, ML Engine for AI logic, and hosting providers for request processing.

The features are developed in compliance with our internal security policies (ISO 27001 certified), GDPR and AI Act, with a focus on transparency and documentation.

Résumé

Le présent document détaille la sécurité et la gestion des données des fonctionnalités IA de Pennylane.

Les données sont hébergées en Union européenne par des fournisseurs certifiés avec l'engagement contractuel que les données ne sont jamais utilisées pour l'entraînement des modèles.

L'infrastructure comprend des mesures de sécurité strictes incluant le contrôle des accès, l'isolation des données entre utilisateurs, et le chiffrement des données. Des tests de pénétration sont effectués deux fois par an.

L'architecture se compose en quatre niveaux : le frontend Pennylane pour l'affichage, le backend pour la sécurité et l'authentification, le ML Engine pour la logique IA, et les hébergeurs pour le traitement des requêtes.

Les fonctionnalités sont développées en conformité avec nos politiques de sécurité internes (certifiées ISO 27001), le RGPD et l'IA Act, avec un focus sur la transparence et la documentation.

Hosting providers

Pennylane relies on a select panel of hosting providers chosen for their compliance with the most stringent security standards on the market, such as ISO 27001 and SOC2 certification. This rigorous selection enables us to ensure an optimum level of security and resilience, validated by independent audits. You can consult the list of our subcontractors at this address.

These hosting providers offer access to a variety of models, so that we can select the most efficient for each context, while offering satisfactory policies in terms of hosting and data processing in the European Union.

Pennylane has contractually obtained a commitment from each of these hosting providers that the data used for AI functionalities will never be used for model training (LLM).

Pennylane s'appuie sur un panel restreint d'hébergeurs choisis pour leur conformité aux standards de sécurité les plus exigeants du marché, à l'instar des certifications ISO 27001 et SOC2. Cette sélection rigoureuse nous permet d'assurer un niveau optimal de sécurité et de résilience, validé par des audits indépendants. Vous pouvez consulter la liste de nos sous-traitants à cette adresse.

Ces hébergeurs permettent l'accès à une variété de modèles, afin de sélectionner le plus performant en fonction de chaque contexte, tout en offrant des politiques satisfaisantes en termes d'hébergement et de traitement des données en Union européenne.

Pennylane a obtenu par contrat auprès de chacun de ces hébergeurs l'engagement que les données utilisées dans le cadre des fonctionnalités IA ne sont jamais utilisées pour l'entraînement des modèles (LLM).

Infrastructure and data security

Storage and hosting

Data is hosted and processed in the European Union by hosting providers that comply with applicable regulations (in particular RGPD). Data stored for technical improvement and performance purposes is subject to access control rules and retention periods determined according to need, in line with our log management policy.

Les données sont hébergées et traitées en Union européenne par des hébergeurs qui respectent la réglementation applicable (notamment le RGPD). Les données stockées à des fins d'amélioration technique et de performance sont soumises à des règles de contrôles d'accès et des durées de rétention déterminées en fonction du besoin, conformément à notre politique de gestion des logs.

Security

Storage security measures are detailed in our Security Whitepaper available on the Pennylane security page.

These measures include, in particular, for IA functionalities:

- Strict access control. Pennylane implements mechanisms to ensure that each user can only access his or her own data. These mechanisms include:
    - Systematic authentication of each request;
    - Complete technical isolation of data between users;
    - Technical limitation of the AI on the data it is working on, so that its answers are not influenced by initial data outside the user's context;
    - Systematic control of user input to protect against injection attacks and manipulation of the AI's prompt, and to ensure that the AI only answers questions that do not raise ethical or legal issues;
    - Access traceability.
- The implementation of measures to limit the risk of jailbreaking (a means of evading the protection mechanisms implemented on the AI model in order to take it out of its security context or deviate from defined ethical requirements)
- Encryption of data in transit and at rest.

Pennylane organizes penetration tests twice a year on its entire application perimeter. AI functionalities are integrated within the Pennylane application and are therefore tested in the same way twice a year by external consultants.

As with all its functionality projects, Pennylane carried out an extensive Security & Privacy by Design analysis on our AI projects before they were implemented.

Les mesures de sécurité sont détaillées dans notre Plan Assurance Sécurité (Security Whitepaper) disponible sur notre page dédiée à la sécurité.

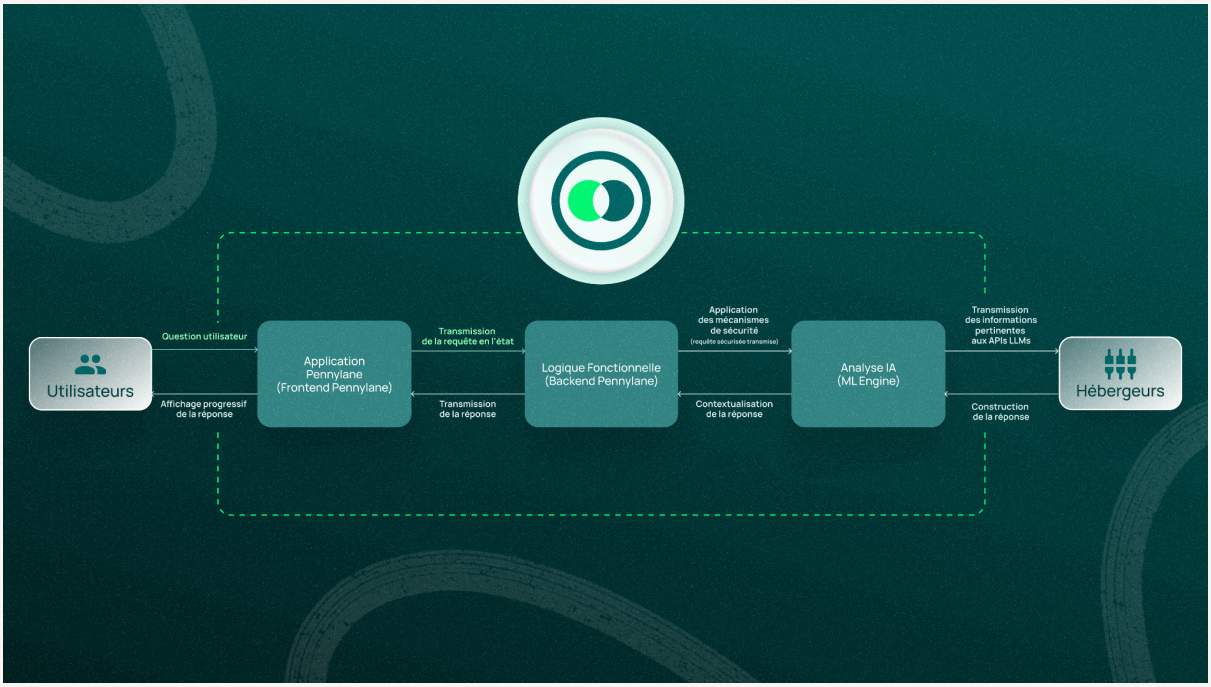
Ces mesures incluent notamment dans le cadre des fonctionnalités IA :

- Un contrôle strict des accès. Pennylane met en œuvre des mécanismes afin de garantir que chaque utilisateur ne puisse accéder qu'à ses propres données. Ces mécanismes incluent :
    - Une authentification systématique de chaque requête ;
    - Une isolation technique complète des données entre les utilisateurs ;
    - Une limitation technique de l'IA sur les données sur lesquelles elle travaille afin que ses réponses ne soient pas influencées par des données initiales en dehors du contexte de l'utilisateur ;
    - Un contrôle systématique des entrées saisies par les utilisateurs pour nous protéger d'attaques d'injection, de manipulation du prompt de l'IA et nous assurer que l'IA ne réponde qu'à des questions ne posant pas de problématique éthique ou juridique ;
    - Une traçabilité des accès.
- La mise en place de mesures visant à limiter le risque de jailbreaking (moyen d'échapper aux mécanismes de protection implémentés sur le modèle d'IA afin de le faire sortir de son contexte de sécurité ou de dévier des exigences d'éthiques définies) ;
- Le chiffrement en transit et au repos des données.

Pennylane organise des tests de pénétration deux fois par an sur l'ensemble de son périmètre applicatif. Les fonctionnalités IA sont intégrées au sein de l'application Pennylane et sont donc testées au même titre deux fois par an par des consultants externes.

Pennylane a conduit, comme pour l'ensemble de ses projets de fonctionnalités, une analyse étendue de Security & Privacy by Design sur nos projets d'IA avant leur mise en place.

Integration and encapsulation



Pennylane application (Frontend)

Concerns only the logic for displaying responses from AI Functionalities (conversational or otherwise). The frontend is connected only to the backend, via a REST API and a WebSocket connection.

Concerne uniquement la logique d'affichage des réponses des fonctionnalités IA (conversationnelles ou non). Le frontend est connecté uniquement au backend, via une API REST et une connexion WebSocket.

Functional logic (Backend)

The backend is the link between the frontend and the ML Engine.

All requests (questions, discussions) are transmitted to the backend in order to guarantee the same security requirements, authentication, verification of user permissions and data compartmentalization that apply to the entire Pennylane application.

The APIs made available for the ML Engine are also the responsibility of the backend for the same reasons, and the same security mechanisms described above are therefore in place.

Requests received in this way are forwarded to the ML Engine once the security checks have been carried out. The ML Engine analyzes them and sends the complete responses back to the frontend, or if necessary, partial response streams via WebSocket for progressive display.

Le backend fait le lien entre le frontend et le ML Engine.

L'ensemble des requêtes (questions, discussions) est transmis au backend afin de garantir les mêmes exigences de sécurité, d'authentification, de vérification des permissions de l'utilisateur et de cloisonnement des données qui s'appliquent sur l'ensemble de l'application Pennylane.

Les API mises à disposition pour le ML Engine sont également de la responsabilité du backend pour les mêmes raisons, les mêmes mécanismes de sécurité décrits précédemment sont en place.

Les requêtes ainsi reçues sont ensuite transmises au ML Engine une fois les vérifications de sécurité effectuées. Ce dernier les analyse et renvoie les réponses complètes au frontend, ou le cas échéant des flux de réponses partielles via WebSocket pour un affichage progressif.

AI analysis (ML engine)

The ML Engine, deployed at Pennylane, contains all the agentic AI logic: declaration of AI agents, definition of tools available to LLMs, prompt libraries, content anonymization where appropriate, etc.

The ML Engine acts as the link between the Pennylane backend and the hosting providers, and is responsible for monitoring AI functionalities to ensure the right level of quality and performance.

Le ML Engine, déployé chez Pennylane, contient toute la logique d'IA agentique : déclaration des agents IA, définition des outils à disposition des LLMs, librairies de prompts, anonymisation de contenu le cas échéant, etc.

Le ML Engine fait le lien entre le backend de Pennylane et les hébergeurs, et est responsable du monitoring des fonctionnalités IA pour assurer le bon niveau de qualité et de performance.

Hosting providers

Hosting providers are restricted to the role of "LLM API":

- Application of proprietary filters on requests and responses (violence, hate, etc.);
- Processing of requests formalized by Pennylane to optimize the relevance of results provided by hosters' LLMs.

No data persists in the hosted LLMs' AI tools/services: data relevance is managed solely by Pennylane's ML Engine, for limited retention periods proportional to the need to improve Pennylane's AI functionalities.

Les hébergeurs sont cantonnés à un rôle d'"API LLM" :

- Application de filtres propriétaires sur les requêtes et les réponses (violence, haine, etc.) ;
- Traitement de requêtes formalisées par Pennylane pour optimiser la pertinence des résultats fournis par les LLMs des hébergeurs.

Aucune donnée ne persiste dans les outils / services IA des LLMs hébergés : la pertinence des données est gérée uniquement par le ML Engine de Pennylane, pour des durées de rétention limitées et proportionnelles au besoin d'amélioration des fonctionnalités IA de Pennylane.

Governance and control

To ensure the safe use of our AI functionalities, Pennylane has put in place several measures to secure the use of these functionalities.

Prior to development, we have:

- Carried out an analysis of national and European legislation to determine the applicable legal framework;
- Carried out an analysis of issues relating to the security and protection of personal data, based on guidelines and methodologies published by national authorities such as the CNIL or ANSSI;
- Selected our suppliers in compliance with our procurement process, which require joint validation by the Security, Legal and Finance teams.

During the development phase:

- We have put in place mechanisms to ensure the quality and confidentiality of information passing through our functionalities;
- We have anonymized content where necessary;
- We document the risks identified and the mitigation measures put in place.

Pennylane a mis en place plusieurs mesures visant à sécuriser les fonctionnalités IA.

En amont du développement, nous avons:

- Mené une analyse de la législation nationale et européenne pour déterminer le cadre légal applicable,
- Réalisé une analyse des enjeux relatifs à la sécurité et à la protection des données à caractère personnel en nous appuyant sur les lignes directrices et méthodologies publiées par les autorités nationales, comme la CNIL ou l'ANSSI,
- Sélectionné nos fournisseurs dans le respect de nos procédures achats, qui requièrent une validation conjointe des équipes Sécurité, Juridique et Finance.

Durant la phase de développement :

- Nous avons mis en place des mécanismes permettant d'assurer la qualité et la confidentialité des informations qui transitent dans nos fonctionnalités;
- Nous avons procédé à une anonymisation des contenus là où cela est nécessaire;
- Nous documentons les risques relevés et les mesures de mitigations mises en place.

Compliance and regulations

The AI Functionalities comply with applicable regulations, including the AI ACT and the GDPR.

Concerning compliance with the RGPD, the processing of personal data induced by the use of the AI Functionalities is governed by our Data Processing Agreement accessible from our General Terms and Conditions.

Concerning compliance with the IA Act, we develop our functionalities in compliance with the principles of the regulation, in particular the following principles:

- **Transparency**: you will be informed as soon as one of the functionalities you wish to use is based on an AI system within the meaning of European regulations;
- **Documentation**: we implement and maintain the necessary documentation on the AI functionalities we work on to ensure that their level of risk is acceptable, and that they otherwise comply with other applicable regulations, such as the RGPD. We also ensure that only relevant data is accessed when using these functionalities;
- **Post-marketing monitoring** : we have carried out extensive testing before releasing the AI Features into production, and we will monitor them after launch to ensure that they operate in line with our expectations and your needs, and to identify any potential risks;
- **Instructions for use:** we have drawn up clear and accessible documentation on the functionalities and how to use them, in particular to help you understand their capabilities and limitations.

Les fonctionnalités IA sont conformes à la réglementation applicable, notamment l'IA Act et le RGPD.

Concernant la conformité au RGPD, les traitements de données à caractères personnels induits par l'utilisation des fonctionnalités IA sont encadrés par notre accord sur le traitement des données accessible depuis nos conditions générales.

Concernant la conformité à l'IA Act, nous développons nos fonctionnalités dans le respect des principes de la réglementation, en particulier les principes suivants :

- **Transparence** : vous serez informés dès que l'une des fonctionnalités que vous souhaitez utiliser repose sur un système d'IA au sens de la réglementation européenne ;
- **Documentation** : nous mettons en place et tenons à jour la documentation nécessaire sur les fonctionnalités IA sur lesquelles nous travaillons pour nous assurer que leur niveau de risque est acceptable, et qu'elles respectent par ailleurs les autres réglementations applicables, comme le RGPD. Nous nous assurons également que seules les données pertinentes sont accédées dans le cadre de l'utilisation de ces fonctionnalités ;
- **Surveillance après commercialisation** : nous avons réalisé de nombreux tests avant la mise en production des Fonctionnalités IA et nous procéderons à une surveillance après leur lancement pour nous assurer qu'elles fonctionnent conformément à nos attentes, à vos besoins, et identifier les potentiels risques ;
- **Instructions d'usage** : nous avons rédigé une documentation claire et accessible sur les fonctionnalités et la façon de les utiliser, notamment pour vous permettre d'appréhender leurs capacités et leurs limites.

Contact

Guillaume Gohin - Head of Information and Security

guillaume.gohin@pennylane.com

Stanislas de Villoutreys - Head of Legal, Compliance & Risk

stanislas.devilloutreys@pennylane.com

security@pennylane.com

legal@pennylane.com