PixelGuard: Thúc đẩy quyền riêng tư dữ liệu chăm sóc sức khỏe thông qua hệ thống khử nhận dạng được điều khiển bởi AI cho nghiên cứu hình ảnh y tế

**📖 Bài viết gốc:** https://aws.amazon.com/vi/blogs/publicsector/pixelguard-advancing-healthcare-data-privacy-through-ai-driven-de-identification-system-for-medical-imaging-research/  
**👤 Tác giả:** Shashank Tanksali, Steve Fu, Abhijit Gaonkar, và Javeed Shaik  
**📅 Ngày xuất bản:** 09 tháng 7 năm 2025  
**🌐 Nguồn:** AWS Blog  
**👨‍💻 Người dịch:** Nguyễn Quang Bình  
**📅 Ngày dịch:** 08/07/2025  
**⏱️ Thời gian đọc:** 15 phút

──────────────────────────────────────────────────

## 📋 Tóm tắt

Hình ảnh y tế đóng vai trò quan trọng trong nghiên cứu y học bằng cách cung cấp những hiểu biết có giá trị giúp thúc đẩy sự hiểu biết của chúng ta về sức khỏe con người, quản lý bệnh tật và hiệu quả điều trị. Các nhà nghiên cứu sử dụng hình ảnh y tế để nghiên cứu cấu trúc và chức năng của các cơ qua...

**🎯 Đối tượng đọc:** Developers, Solution Architects, DevOps Engineers  
**📊 Độ khó:** Intermediate  
**🏷️ Tags:** AWS, Cloud Computing, Architecture

──────────────────────────────────────────────────

## 📚 Mục lục

• Giới thiệu  
• Kiến trúc giải pháp  
• Triển khai  
• Kết luận  
• Glossary - Thuật ngữ  
• Tài liệu tham khảo

──────────────────────────────────────────────────



Hình ảnh y tế đóng vai trò quan trọng trong nghiên cứu y học bằng cách cung cấp những hiểu biết có giá trị giúp thúc đẩy sự hiểu biết của chúng ta về sức khỏe con người, quản lý bệnh tật và hiệu quả điều trị. Các nhà nghiên cứu sử dụng hình ảnh y tế để nghiên cứu cấu trúc và chức năng của các cơ quan, mô và tế bào ở trạng thái khỏe mạnh và bệnh tật. Những hình ảnh này cũng được sử dụng để đào tạo và giáo dục các chuyên gia chăm sóc sức khỏe, để tạo ra các tài nguyên giáo dục và hội thảo, và để đào tạo nhân viên y tế trong việc hiểu và diễn giải dữ liệu hình ảnh.

Mặc dù hình ảnh y tế đóng vai trò quan trọng trong nghiên cứu và giáo dục y học, việc bảo vệ quyền riêng tư trong hình ảnh y tế có tầm quan trọng hàng đầu để đảm bảo tính bảo mật của bệnh nhân, sự tin tưởng và tuân thủ các tiêu chuẩn đạo đức và pháp lý, như HIPAA và HITRUST. Hình ảnh y tế chứa thông tin rất nhạy cảm về sức khỏe, tình trạng và tiên lượng của bệnh nhân. Duy trì quyền riêng tư của Thông tin Nhận dạng Cá nhân (PII) và Thông tin Sức khỏe Cá nhân (PHI) đảm bảo rằng thông tin này không bao giờ được tiết lộ mà không có sự đồng ý rõ ràng của bệnh nhân đồng thời xây dựng lòng tin của bệnh nhân trong chăm sóc sức khỏe.

Hơn nữa, các luật như Đạo luật Tính di động và Trách nhiệm Bảo hiểm Y tế (HIPAA), Khung Bảo mật Chung của Liên minh Tin cậy Thông tin Y tế tại Hoa Kỳ, và Quy định Bảo vệ Dữ liệu Chung (GDPR) tại Liên minh Châu Âu (EU) đặt ra các hướng dẫn nghiêm ngặt về cách dữ liệu y tế như hình ảnh nên được xử lý, lưu trữ, ẩn danh hóa và chia sẻ. Vi phạm các quy định này có thể dẫn đến các hình phạt nghiêm trọng và có thể làm tổn hại danh tiếng của các nhà cung cấp chăm sóc sức khỏe và các tổ chức của họ. Hơn nữa, các chuyên gia chăm sóc sức khỏe có nghĩa vụ đạo đức và đạo đức để tôn trọng quyền riêng tư của bệnh nhân.

**Hình ảnh Kỹ thuật số và Truyền thông trong Y học (DICOM)**

DICOM là định dạng tiêu chuẩn được sử dụng để lưu trữ, truyền tải và chia sẻ hình ảnh y tế và dữ liệu liên quan. Các tệp DICOM chứa thông tin được cấu trúc thành dữ liệu hình ảnh và siêu dữ liệu khác. Mỗi trường DICOM được sử dụng để mô tả và phân loại các khía cạnh khác nhau của hình ảnh y tế như thông tin bệnh nhân, chi tiết nghiên cứu và các tham số hình ảnh. Ví dụ, các trường DICOM bao gồm tên bệnh nhân, ngày sinh, giới tính, ngày nghiên cứu, bác sĩ giới thiệu, dữ liệu pixel, thông tin chuỗi, nhà sản xuất thiết bị, tên mô hình, phiên bản phần mềm, nghiên cứu và mô tả chẩn đoán, trong số những thứ khác.

Các tệp DICOM chứa Thông tin Nhận dạng Cá nhân (PII), do đó chúng không thể được sử dụng trong nghiên cứu y tế hoặc đào tạo mà không có thông tin nhạy cảm được biên tập. Tuy nhiên, điều quan trọng là đảm bảo rằng quá trình biên tập thông tin nhạy cảm không làm tổn hại chất lượng của thông tin không thể quy cho cá nhân. Cũng quan trọng là giảm thiểu kích thước của hình ảnh y tế đã khử nhận dạng để giảm chi phí lưu trữ và xử lý trong khi cung cấp tính linh hoạt lưu các tệp ẩn danh kết quả dưới dạng DICOM hoặc JPEG. AWS HealthImaging, kho lưu trữ DICOM dựa trên đám mây có khả năng mở rộng và hiệu suất cao, cung cấp truy xuất hình ảnh dưới giây từ bất kỳ đâu. Tổng chi phí sở hữu (TCO) của lưu trữ hình ảnh và truyền dữ liệu có thể được giảm đáng kể bằng cách sử dụng mã hóa hình ảnh High Throughput JPEG 2000 (HTJ2K) tiêu chuẩn ngành.

**Hướng dẫn Giải pháp**

PixelGuard—được xây dựng trên Amazon Web Services (AWS) và được tạo bởi Giáo sư Trợ lý Đại học Northwestern và Người sáng lập Xtasis, LLC, Tiến sĩ Adrienne Kline—là một giải pháp phần mềm tiên tiến khử nhận dạng hình ảnh y tế trong khi bảo tồn tính liên quan và hiệu quả lâm sàng. Nó sử dụng hơn 75 mô hình được điều khiển bởi AI tiên tiến có khả năng phát hiện và biên tập văn bản đa ngôn ngữ, đa hướng trên tất cả các định dạng chính (DICOM, JPEG, PNG, NIfTI, v.v.), cùng với ẩn danh hóa siêu dữ liệu có thể cấu hình. Với giao diện người dùng trực quan, SSO doanh nghiệp và triển khai trong tenant (không có dữ liệu egress), PixelGuard cung cấp khử nhận dạng hình ảnh an toàn, tuân thủ và thông lượng cao. PixelGuard có sẵn trên AWS Marketplace. ScaleCapacity—một AWS Partner—đóng vai trò quan trọng trong việc phát triển giao diện người dùng, cơ sở hạ tầng đám mây và triển khai lên AWS Marketplace.

**Lớp Ingestion**

Lớp ingestion cung cấp giao diện người dùng và API để gửi hình ảnh y tế để khử nhận dạng. API cũng có thể được sử dụng bởi nhà sản xuất thiết bị y tế bên thứ ba để cung cấp khả năng biên tập. Trước khi ingestion, web UI và API có thể được bảo mật với ủy quyền dựa trên nhà cung cấp danh tính doanh nghiệp. Cùng với việc ingesting hình ảnh cần khử nhận dạng, lớp này cũng phải nắm bắt cấu hình khử nhận dạng cụ thể, định nghĩa các trường nào cần được biên tập từ hình ảnh. Đối với các tình huống biên tập phổ biến như biên tập để tuân thủ quy định HIPAA, các bộ trường được xác định trước được định nghĩa và ghi lại.

**Lớp Pre De-identification**

Lớp lưu trữ được sử dụng để lưu trữ hình ảnh trước khi khử nhận dạng. Hơn nữa, lớp lưu trữ cũng có thể lưu trữ một tệp nén chứa một số hình ảnh cần khử nhận dạng. Hơn nữa, lớp này cũng lưu trữ siêu dữ liệu liên quan đến công việc khử nhận dạng như ID công việc, thời gian gửi, các trường cụ thể được biên tập, định dạng của tệp được cung cấp, v.v.

**Lớp Xử lý**

Lớp lưu trữ xác định định dạng của hình ảnh cần khử nhận dạng và dựa trên cấu hình khử nhận dạng cụ thể, siêu dữ liệu hình ảnh được định nghĩa bởi các thẻ DICOM và khử nhận dạng cấp Pixel được thực hiện. Hơn nữa, hình ảnh có thể được nén để tối ưu hóa lưu trữ và một tệp crosswalk tham chiếu một ID duy nhất được tạo. Tệp crosswalk làm cho việc tra cứu ngược của tệp gốc trở nên khả thi nếu cần thiết. Cần thận trọng để bảo mật tệp crosswalk và lưu trữ nó riêng biệt với tệp đã khử nhận dạng.

**Đường ống De-identification**

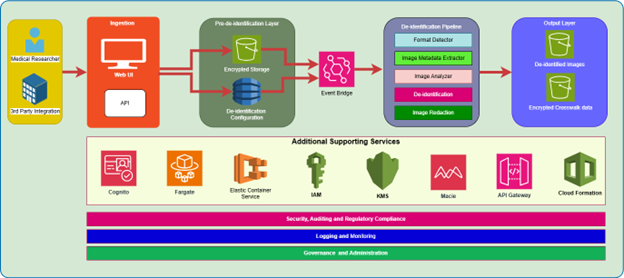
Lớp lưu trữ đã khử nhận dạng lưu trữ hình ảnh y tế đã khử nhận dạng được xử lý, có thể được sử dụng bởi các nhà nghiên cứu y tế như một phần của nghiên cứu. Hơn nữa, hình ảnh có thể được tạo ra trong các định dạng khác nhau phù hợp với nhu cầu của nghiên cứu.

**Lớp Đầu ra De-identified**

Các hình ảnh đã khử nhận dạng có thể được tạo ra cùng với một tệp crosswalk có thể duy trì ánh xạ giữa một định danh trong hình ảnh đã khử nhận dạng và một định danh liên quan trong hình ảnh gốc. Tệp crosswalk được duy trì riêng biệt và được mã hóa để chỉ cho phép các cá nhân được ủy quyền truy vết hình ảnh đã khử nhận dạng trở lại nguồn gốc của nó nếu cần.

**Kiểm toán, Giám sát và Khả năng Quan sát**

Lớp kiểm toán, giám sát và khả năng quan sát lưu trữ nhật ký truy cập để đảm bảo rằng các bản ghi về ai đã truy cập gì, khi nào và như thế nào có thể được lưu trữ cho mục đích lưu giữ hồ sơ. Hơn nữa, nhật ký lỗi chi tiết nếu có thể được lưu trữ để cho phép khắc phục sự cố nếu một số hình ảnh y tế không thể được khử nhận dạng.



Hình 1. Kiến trúc giải pháp logic PixelGuard

Việc khử nhận dạng hình ảnh y tế đòi hỏi việc scrubbing siêu dữ liệu và dữ liệu pixel. Trong một số tình huống, siêu dữ liệu có thể cần được ẩn danh hóa thay vì bị xóa hoàn toàn. Khi siêu dữ liệu được ẩn danh hóa thay vì bị xóa, quá trình ẩn danh hóa không được làm loãng giá trị nghiên cứu của hình ảnh theo bất kỳ cách nào. Ví dụ, một hình ảnh y tế của một người 20 tuổi không thể được ẩn danh hóa để chỉ ra rằng hình ảnh thuộc về một người 80 tuổi, vì điều đó có thể làm loãng giá trị nghiên cứu.

Để scrub dữ liệu pixel, nhận dạng ký tự quang học (OCR) để phát hiện văn bản được đốt cháy có thể được sử dụng. Điều này phải được sử dụng cùng với các mô hình học máy (ML) để định vị và làm mờ các khu vực có thông tin nhận dạng. Các mô hình ML cũng có thể được sử dụng để tạo ra điểm tin cậy liên quan đến việc làm mờ thông tin. Đối với các hình ảnh có điểm tin cậy dưới một ngưỡng được xác định trước nhất định, một hàng đợi đánh giá của con người cũng có thể được định nghĩa để điều tra và có thể phê duyệt.

**Công cụ và Khung**

**Ingestion**: Lớp ingestion có thể được xử lý với Web UI và API có thể được tiếp xúc thông qua API Gateway và có thể thực hiện xác thực yêu cầu, xác thực và ủy quyền cùng với thực thi quy tắc bảo mật. API Gateway tích hợp với Amazon Cognito để thực thi xác thực và ủy quyền chi tiết.

**Lưu trữ Pre-de-identification**: Lớp này có thể lưu trữ các hồ sơ khử nhận dạng hình ảnh được xác định trước như những hồ sơ cho HIPAA, và có thể lưu trữ nó vào một bucket Amazon Simple Storage Service (Amazon S3) hoặc vào một kết hợp của bucket S3 và Amazon DynamoDB.

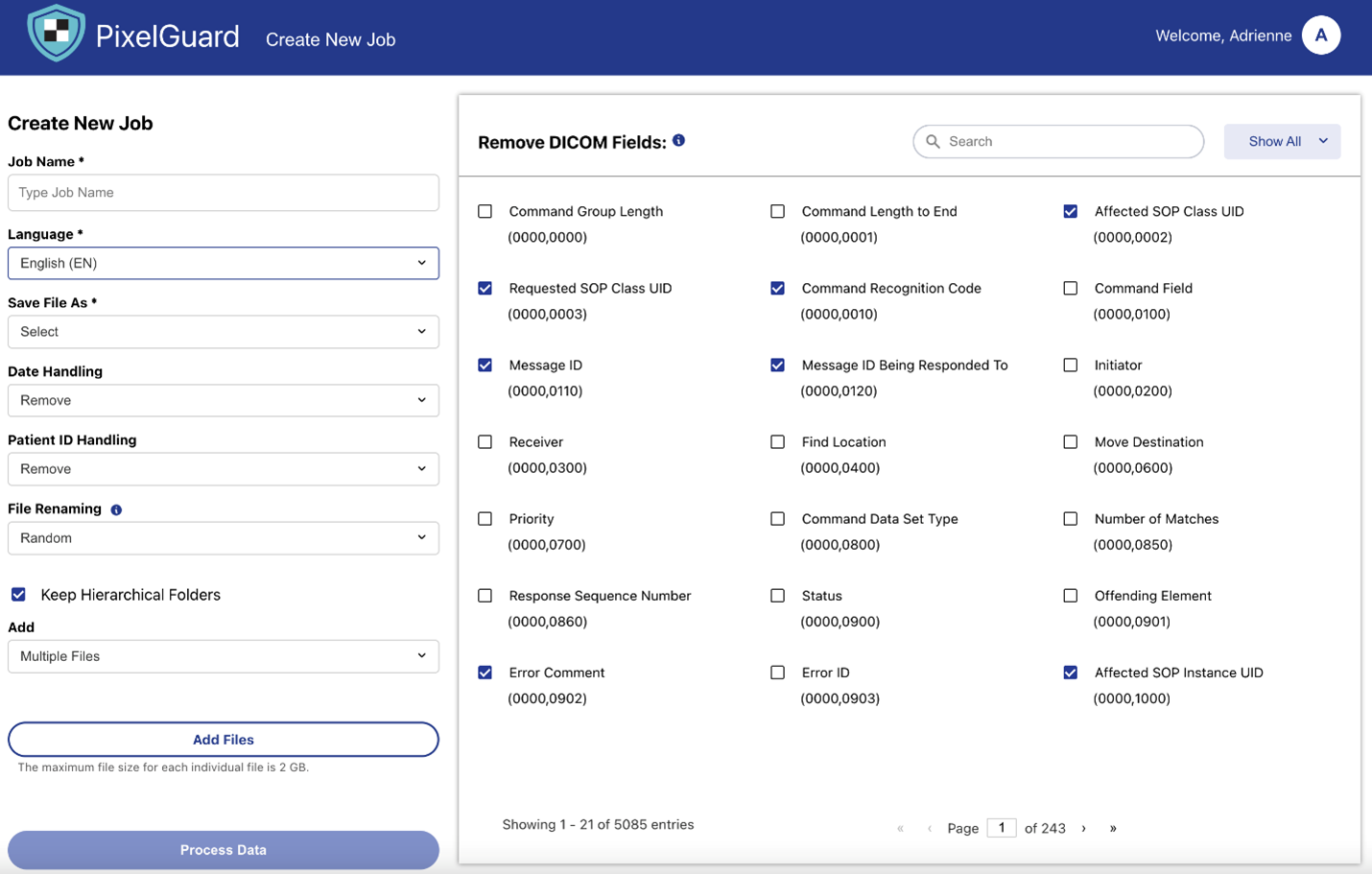
**De-identification**: De-identification là một quy trình nhiều bước có thể sử dụng các thư viện parser DICOM như pydicom và các công cụ scrubbing như dicom-anonymizer. Amazon Textract và Amazon Comprehend Medical cũng có thể được sử dụng để xác định văn bản nhúng trong hình ảnh.

**Đầu ra**: Hình ảnh y tế sau khử nhận dạng có thể được nén, chuyển đổi sang định dạng khác và lưu trữ với các định danh có thể được ánh xạ đến nguồn gốc nếu cần. Việc lưu trữ có thể được xử lý với bucket S3 hoặc hệ thống tệp khi cần với hoặc không có nén.

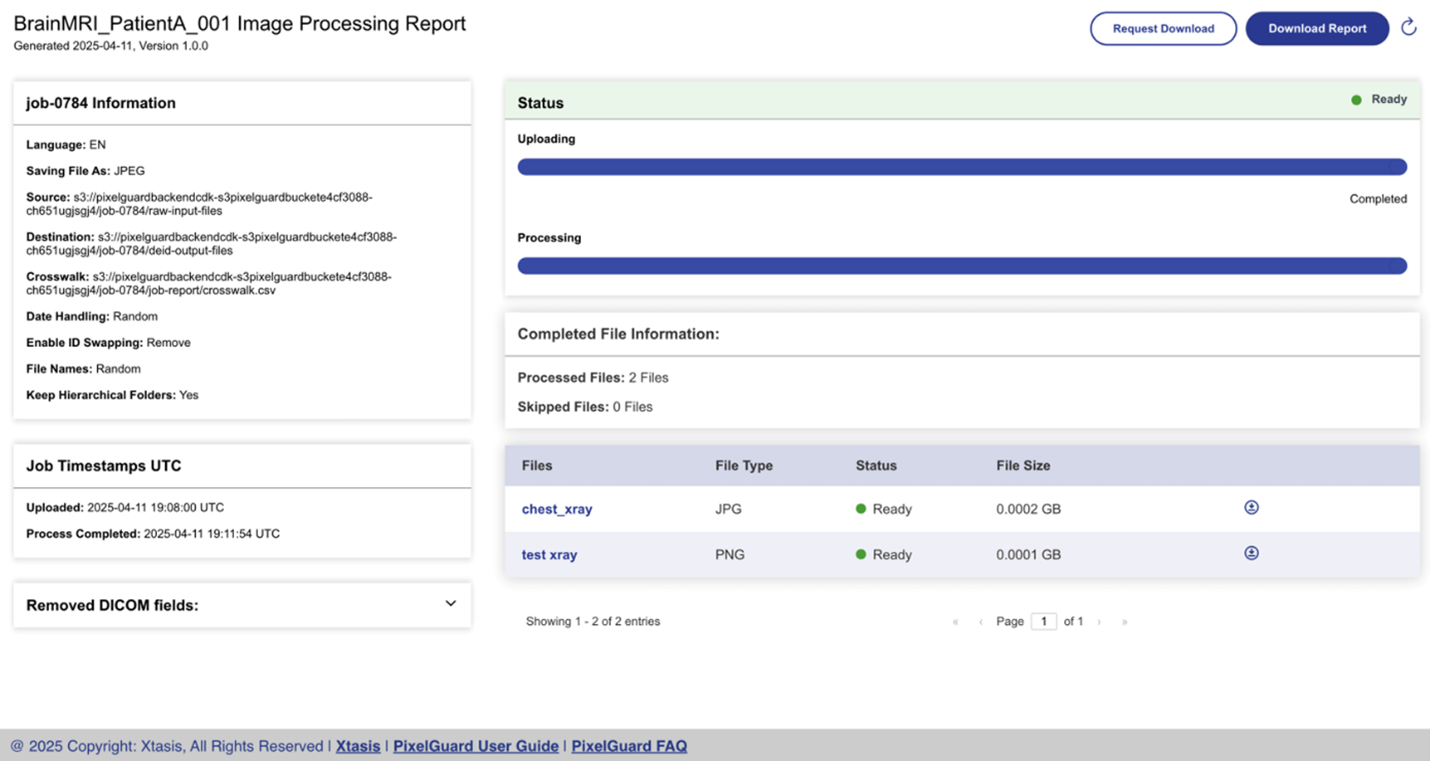
**Các dịch vụ hỗ trợ khác**: Các dịch vụ hỗ trợ khác đóng vai trò quan trọng trong việc cho phép bảo mật, khả năng mở rộng, khả năng quản lý và tự động hóa cơ sở hạ tầng. Các dịch vụ như Amazon API Gateway làm cho việc thực thi các chính sách AuthN/AuthZ phù hợp với nhu cầu tổ chức trở nên khả thi. AWS Key Management Service (AWS KMS) cung cấp một cách an toàn để quản lý các khóa mật mã và bảo vệ dữ liệu của bạn. AWS Fargate có thể được sử dụng để chạy hàng nghìn container mà không cần quản lý máy chủ hoặc cụm của các instance Amazon Elastic Compute Cloud (Amazon EC2).

Hơn nữa, các dịch vụ như Amazon Macie có thể thêm một lớp bảo vệ thứ cấp để xác định dữ liệu nhạy cảm và bảo vệ nó cùng với các dịch vụ khắc phục khác. Hơn nữa, các dịch vụ như AWS CloudFormation cho phép định nghĩa các tài nguyên cơ sở hạ tầng AWS theo cách khai báo, sử dụng các mẫu ở định dạng JSON hoặc YAML và có thể tự động hóa việc tạo, cấu hình và quản lý các tài nguyên này, hoạt động như một công cụ cơ sở hạ tầng dưới dạng mã (IaC).

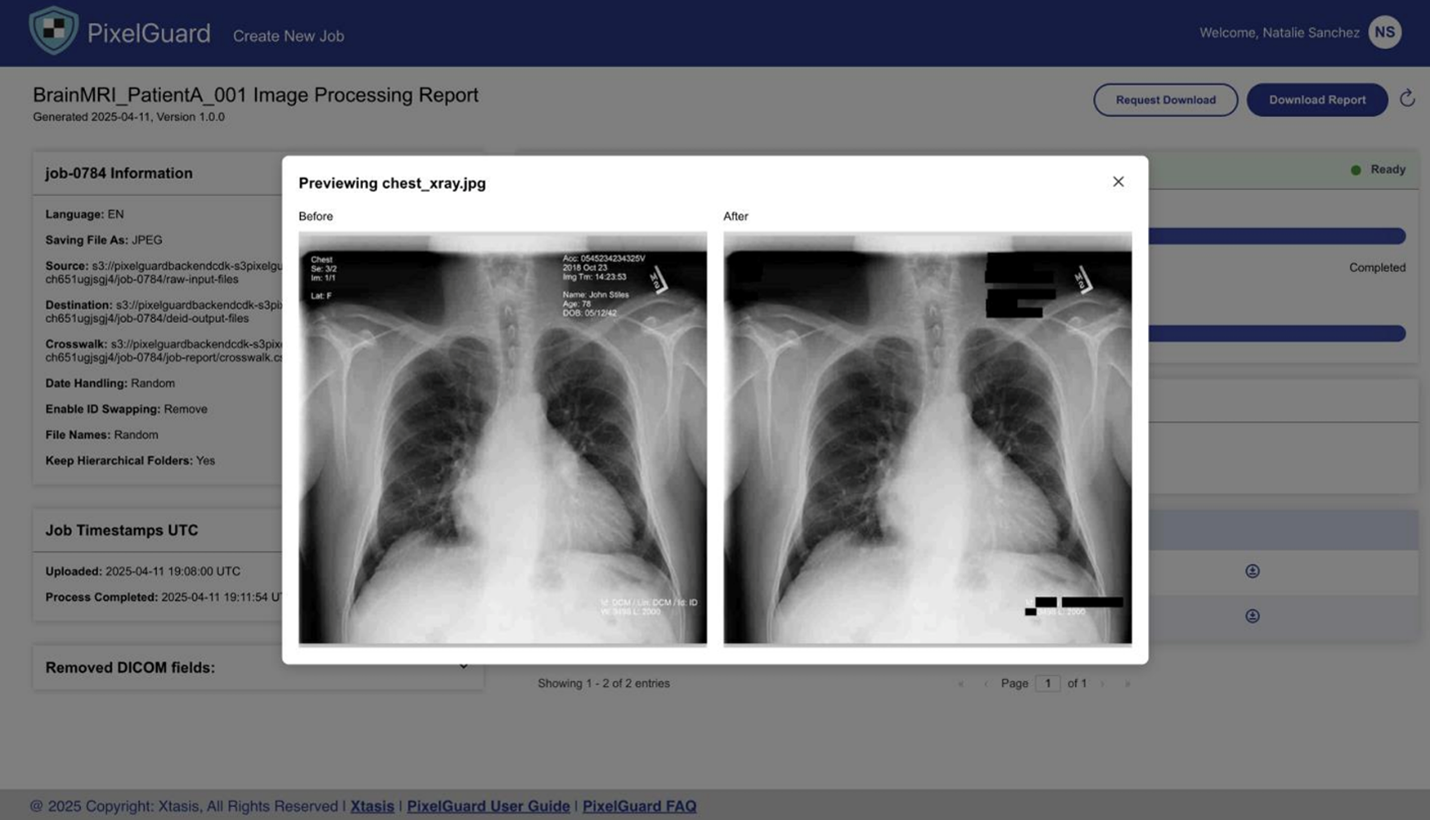
Các hình ảnh sau cung cấp tổng quan về trải nghiệm biên tập hình ảnh sử dụng PixelGuard.



Hình 2. Cấu hình các trường để biên tập trong hình ảnh y tế



Hình 3. Báo cáo hoàn thành công việc biên tập



Hình 4. Xem trước hình ảnh chưa biên tập và đã biên tập cùng nhau

**Kết luận**

Khử nhận dạng hình ảnh y tế là một nhu cầu quan trọng trong nghiên cứu và đào tạo y tế. Nó đảm bảo rằng hình ảnh đã khử nhận dạng không thể được liên kết trở lại với một cá nhân. Một kiến trúc kỹ thuật vững chắc để khử nhận dạng hình ảnh y tế là quan trọng cho việc sử dụng an toàn và tuân thủ hình ảnh y tế cho nghiên cứu và khám phá y tế. Kiến trúc này có thể đảm bảo rằng PII được loại bỏ hoặc che khuất một cách đáng tin cậy bảo vệ quyền riêng tư của bệnh nhân mà không làm tổn hại tính hữu ích của dữ liệu.

PixelGuard như một giải pháp được xây dựng trên AWS cho phép khử nhận dạng linh hoạt, do đó cho phép tuân thủ luật quyền riêng tư, tăng cường bảo mật và giảm rủi ro. Đồng thời, nó tạo điều kiện cho việc chia sẻ dữ liệu và hợp tác, cuối cùng thúc đẩy nghiên cứu AI hình ảnh y tế nhanh hơn trong sức khỏe cộng đồng và tiến bộ y tế, tất cả trong khi duy trì trách nhiệm đạo đức và pháp lý.

──────────────────────────────────────────────────

## 📖 Glossary - Thuật ngữ

|  |  |  |
| --- | --- | --- |
| English | Tiếng Việt | Định nghĩa |
| Auto Scaling | Tự động mở rộng quy mô | Khả năng tự động tăng/giảm resources dựa trên demand |
| Load Balancer | Bộ cân bằng tải | Phân phối traffic đến multiple servers |
| Microservices | Kiến trúc microservices | Architectural pattern chia application thành small services |

## 🔗 Tài liệu tham khảo

**Tài liệu gốc**

• AWS Blog: https://aws.amazon.com/vi/blogs/publicsector/pixelguard-advancing-healthcare-data-privacy-through-ai-driven-de-identification-system-for-medical-imaging-research/

• Related Articles: https://aws.amazon.com/vi/blogs/publicsector/solving-medical-mysteries-aws-cloud-medical-data-sharing-innovation-undiagnosed-diseases-network/

──────────────────────────────────────────────────

## 💬 Ghi chú của người dịch

**Challenges trong quá trình dịch**

• Technical Terms: Giữ nguyên các thuật ngữ AWS để đảm bảo tính chính xác

• Cultural Context: Điều chỉnh ví dụ cho phù hợp với bối cảnh Việt Nam

• Complex Concepts: Giải thích rõ ràng các khái niệm phức tạp

**Insights gained**

• Technical Learning: Hiểu sâu hơn về kiến trúc AWS và best practices

• Language Skills: Cải thiện kỹ năng dịch thuật kỹ thuật

• Industry Knowledge: Nắm bắt xu hướng công nghệ mới

──────────────────────────────────────────────────

## 🤝 Đóng góp và Feedback

Bài dịch này được thực hiện trong khuôn khổ FCJ Internship Program.

**📧 Liên hệ:** ngqbinh456@gmail.com.  
**💬 Feedback:** Mọi góp ý để cải thiện chất lượng dịch thuật xin gửi về email trên  
**🔄 Updates:** Bài dịch sẽ được cập nhật dựa trên feedback từ cộng đồng

──────────────────────────────────────────────────

© 2024 - Bản dịch thuộc về Nguyễn Quang Bình. Vui lòng credit khi sử dụng.