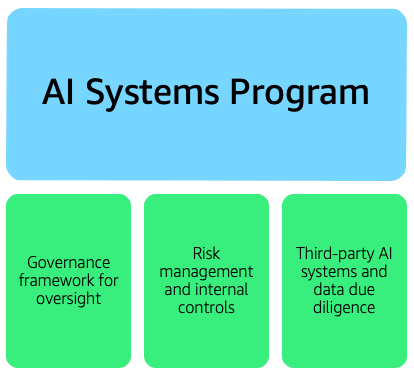
# Định hướng Amazon Bedrock theo các Nguyên tắc AI và Thông báo Mẫu của NAIC

tác giả Cory Visi, Dan Kearney, Jonathan Yeldell, Raj Pathak và Stephen Eschbach 01 THÁNG 5 NĂM 2025 trong [Amazon Bedrock](https://aws.amazon.com/blogs/industries/category/artificial-intelligence/amazon-machine-learning/amazon-bedrock/) , [Cơ sở tri thức Amazon Bedrock](https://aws.amazon.com/blogs/industries/category/artificial-intelligence/amazon-machine-learning/amazon-bedrock/amazon-bedrock-knowledge-bases/) , [Amazon CloudWatch](https://aws.amazon.com/blogs/industries/category/management-tools/amazon-cloudwatch/) , [Amazon Simple Storage Service (S3)](https://aws.amazon.com/blogs/industries/category/storage/amazon-simple-storage-services-s3/) , [AWS CloudTrail](https://aws.amazon.com/blogs/industries/category/management-tools/aws-cloudtrail/) , [AWS Identity and Access Management (IAM)](https://aws.amazon.com/blogs/industries/category/security-identity-compliance/aws-identity-and-access-management-iam/) , [AWS Key Management Service](https://aws.amazon.com/blogs/industries/category/security-identity-compliance/aws-key-management-service/) , [Dịch vụ tài chính](https://aws.amazon.com/blogs/industries/category/industries/financial-services/) , [AI tạo sinh](https://aws.amazon.com/blogs/industries/category/artificial-intelligence/generative-ai/) , [Ngành công nghiệp](https://aws.amazon.com/blogs/industries/category/industries/) [Liên kết cố định](https://aws.amazon.com/blogs/industries/aligning-amazon-bedrock-with-naic-ai-principles-and-model-bulletin/) [Bình luận](https://aws.amazon.com/blogs/industries/aligning-amazon-bedrock-with-naic-ai-principles-and-model-bulletin/#Comments) [Chia sẻ](https://aws.amazon.com/vi/blogs/industries/aligning-amazon-bedrock-with-naic-ai-principles-and-model-bulletin/#)

Sự tiến bộ nhanh chóng của [AI tạo ra](https://aws.amazon.com/generative-ai/) hứa hẹn những chiến lược mới cho sự đổi mới trong ngành bảo hiểm và các công ty bảo hiểm nên triển khai theo các yêu cầu của quy định. [Hiệp hội các ủy viên bảo hiểm quốc gia (NAIC)](https://content.naic.org/committees/h/big-data-artificial-intelligence-wg) đã thông qua [các Nguyên tắc về trí tuệ nhân tạo](https://content.naic.org/sites/default/files/inline-files/NAIC%20Principles%20on%20AI.pdf) và phát hành [Bản tin mẫu](https://content.naic.org/sites/default/files/inline-files/2023-12-4%20Model%20Bulletin_Adopted_0.pdf) về việc sử dụng các hệ thống trí tuệ nhân tạo của các công ty bảo hiểm để cung cấp hướng dẫn về việc phát triển và sử dụng AI có trách nhiệm trong bảo hiểm. Khi các công ty bảo hiểm ngày càng tận dụng [Amazon Bedrock](https://aws.amazon.com/bedrock/) cho các ứng dụng AI, việc họ hiểu cách liên kết các triển khai này với các hướng dẫn của NAIC là rất hữu ích.

Trong bài viết này, chúng ta sẽ khám phá những cân nhắc chính để giải quyết các yêu cầu của NAIC AI Principles và Model Bulletin khi sử dụng Amazon Bedrock. Chúng ta sẽ đề cập đến quản trị, kiểm soát quản lý rủi ro và cân nhắc của bên thứ ba, đây là ba trụ cột của Chương trình AIS (Hệ thống AI) theo định nghĩa của NAIC.

## **Triển khai chương trình AIS**

****

*Đồ họa thông tin hiển thị các thành phần của Chương trình Hệ thống AI*

Theo định nghĩa của NAIC, AIS là hệ thống dựa trên máy có thể tạo ra các đầu ra như dự đoán, khuyến nghị, nội dung hoặc các đầu ra khác ảnh hưởng đến các quyết định được đưa ra trong môi trường thực hoặc ảo. NAIC Model Bulletin khuyến nghị các công ty bảo hiểm phát triển và duy trì một chương trình bằng văn bản để sử dụng có trách nhiệm các Hệ thống AI, được gọi là Chương trình AIS. Chương trình AIS phải giải quyết vấn đề quản trị, quản lý rủi ro và giám sát trong suốt vòng đời của AIS. Các yêu cầu của NAIC phù hợp chặt chẽ với các chiều [AI có trách nhiệm](https://aws.amazon.com/ai/responsible-ai/): công bằng, minh bạch, quyền riêng tư và bảo mật, an toàn, khả năng kiểm soát, tính xác thực và độ mạnh mẽ, khả năng giải thích và quản trị. Mặc dù có các thuật ngữ khác nhau được sử dụng trong NAIC Model Bulletin, nhưng trọng tâm của nó là quản lý rủi ro, bảo vệ dữ liệu, giám sát mô hình và bảo vệ người tiêu dùng, phản ánh các nguyên tắc tương tự và do đó, AWS cung cấp các tài nguyên và tính năng dịch vụ hỗ trợ việc triển khai Chương trình AIS.

Sau đây là cách các công ty bảo hiểm có thể triển khai chương trình AIS với Amazon Bedrock, bao gồm tất cả các hướng dẫn của NAIC Model Bulletin:

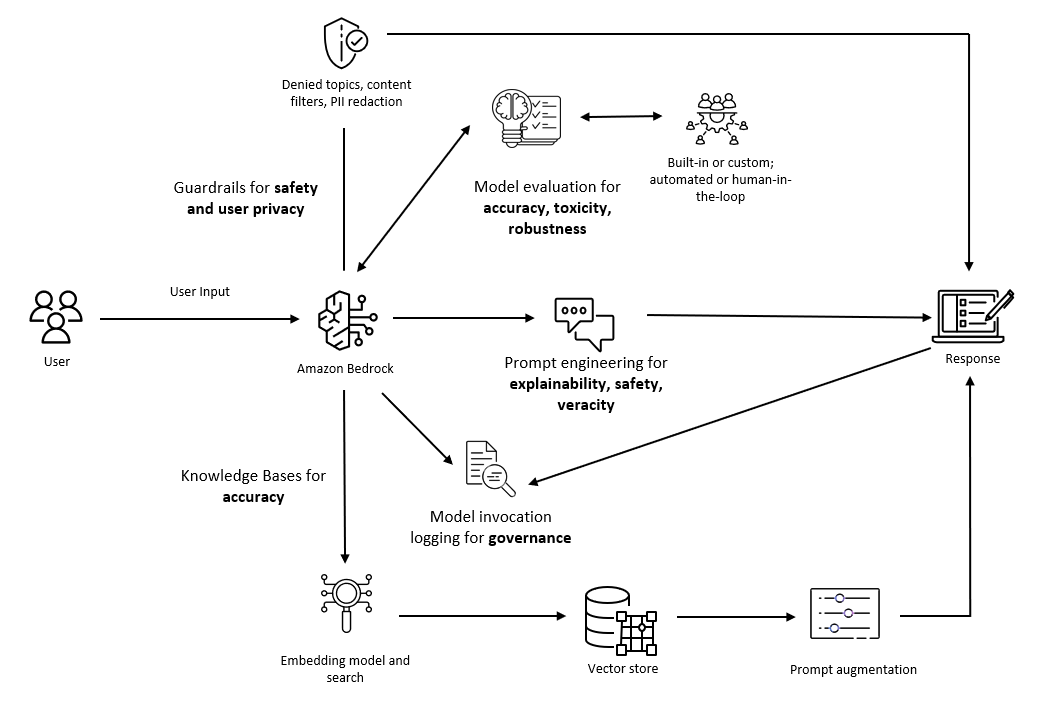
### **Quản trị**

Quản trị AI đáp ứng hướng dẫn của Model AI Bulletin ưu tiên tính minh bạch, công bằng và trách nhiệm giải trình trong thiết kế và triển khai AIS. Nó bao gồm các khuôn khổ, chính sách và quy tắc hướng dẫn phát triển và sử dụng AI, cho phép các bên liên quan đưa ra quyết định sáng suốt. Các khía cạnh chính bao gồm tính minh bạch trong việc sử dụng dữ liệu, quy trình ra quyết định về AI và tác động tiềm ẩn của người dùng. AWS cung cấp các dịch vụ, công cụ và [khuôn khổ quản trị AI](https://docs.aws.amazon.com/whitepapers/latest/aws-caf-for-ai/governance-perspective-managing-an-aiml-driven-organization.html) toàn diện để giúp thiết lập và vận hành các hoạt động này, bao gồm các lĩnh vực bao gồm quản trị dữ liệu và mô hình, giám sát ứng dụng, kiểm toán và quản lý rủi ro.

### **Quản lý rủi ro và kiểm soát nội bộ**

Chương trình AIS sẽ giải quyết vấn đề nhận dạng rủi ro, giảm thiểu, khuôn khổ quản lý và kiểm soát nội bộ của công ty bảo hiểm đối với Hệ thống AI nói chung và ở từng giai đoạn của vòng đời AIS ( [xem Phần 1.7](https://content.naic.org/sites/default/files/inline-files/2023-12-4%20Model%20Bulletin_Adopted_0.pdf) ). Chúng tôi sẽ xem xét các tính năng này theo các khía cạnh cốt lõi của AWS về AI có trách nhiệm, mà chúng tôi coi là cách hiệu quả nhất để quản lý rủi ro.

Sơ đồ sau đây cho thấy cách bạn có thể sử dụng các tính năng của Amazon Bedrock trong toàn bộ vòng đời hệ thống AI để triển khai AI có trách nhiệm trong quản lý rủi ro và kiểm soát.

*Quy trình làm việc mẫu GenAI với Amazon Bedrock phù hợp với các chiều hướng cốt lõi của AI có trách nhiệm* [*(Nguồn)*](https://aws.amazon.com/blogs/machine-learning/considerations-for-addressing-the-core-dimensions-of-responsible-ai-for-amazon-bedrock-applications/)

**Công bằng**  
Công bằng đánh giá tác động của Hệ thống AI đối với nhiều nhóm bên liên quan. Các công ty bảo hiểm có thể sử dụng các tính năng [đánh giá mô hình](https://aws.amazon.com/bedrock/evaluations/) tích hợp của Amazon Bedrock để tiến hành đánh giá kỹ lưỡng các mô hình AI trên nhiều yếu tố nhân khẩu học khác nhau.

Amazon Bedrock cung cấp [các đánh giá tự động](https://docs.aws.amazon.com/bedrock/latest/userguide/evaluation-automatic.html) bằng cách sử dụng các tập dữ liệu được quản lý như [Bias in Open-ended Language Generation Dataset (BOLD)](https://github.com/amazon-science/bold) để đánh giá các mô hình tạo văn bản trên khắp các ngành nghề, giới tính, chủng tộc, hệ tư tưởng tôn giáo và hệ tư tưởng chính trị. Điều này giúp họ xác định các thành kiến ​​tiềm ẩn có thể dẫn đến việc đối xử không công bằng với người tiêu dùng.

Các công ty bảo hiểm cũng có thể tạo các tập dữ liệu tùy chỉnh trong Amazon Bedrock để đánh giá tính công bằng theo các yêu cầu quy định cụ thể của ngành. Bằng cách sử dụng các đánh giá tự động với [tập dữ liệu nhắc nhở tùy chỉnh để đánh giá mô hình](https://docs.aws.amazon.com/bedrock/latest/userguide/model-evaluation-prompt-datasets.html#model-evaluation-prompt-datasets-custom) , các công ty bảo hiểm có thể đánh giá kỹ lưỡng hiệu suất mô hình trên các nhóm nhân khẩu học có liên quan và các đặc điểm riêng biệt đối với hoạt động bảo hiểm. Các công ty bảo hiểm có thể sử dụng cả AWS và các tập dữ liệu tùy chỉnh với Amazon Bedrock Evaluations để đánh giá nội dung được tạo ra của quy trình công việc Retrieval Augmented Generation (RAG) đầu cuối của họ.

Các công ty bảo hiểm nên sử dụng các công cụ đánh giá này trong suốt vòng đời của AIS: từ phát triển đến triển khai và vận hành liên tục.

**Minh bạch**  
AWS giúp cung cấp tính minh bạch cho khách hàng trong lĩnh vực trí tuệ nhân tạo bằng cách giới thiệu [AWS AI Service Cards](https://aws.amazon.com/ai/responsible-ai/resources/#service) . Những thẻ này đóng vai trò là nguồn tài nguyên chuyên dụng được thiết kế để giúp khách hàng hiểu sâu hơn về các dịch vụ và mô hình AWS AI. Chúng cung cấp nhiều thông tin, cung cấp thông tin chi tiết toàn diện về các trường hợp sử dụng dự kiến ​​và hạn chế của từng dịch vụ và mô hình. Hơn nữa, những thẻ này trình bày chi tiết các nguyên tắc thiết kế AI có trách nhiệm hỗ trợ các mô hình và phác thảo các phương pháp hay nhất để triển khai và tối ưu hóa hiệu suất của chúng.

Đối với các mô hình Amazon Bedrock không do AWS phát triển, các nhà cung cấp mô hình sẽ xuất bản các thẻ dịch vụ của riêng họ (ví dụ: [Transparency Hub: Model Reports](https://www.anthropic.com/transparency) và [Meta Llama Models](https://github.com/meta-llama/llama-models/tree/main?tab=readme-ov-file#llama-models-1) của Anthropic ). Mỗi thẻ này cung cấp tổng quan chi tiết về các trường hợp sử dụng cụ thể mà mô hình được thiết kế và các hạn chế cần cân nhắc. Chúng cũng đi sâu vào các kỹ thuật học máy được sử dụng trong quá trình phát triển dịch vụ. Quan trọng là, các thẻ này nêu bật những cân nhắc quan trọng đối với việc sử dụng mô hình có trách nhiệm trong AIS, đảm bảo rằng khách hàng có tất cả thông tin cần thiết để sử dụng các công cụ này một cách có đạo đức và hiệu quả.

**Khả năng giải thích**  
Khả năng giải thích tập trung vào việc hiểu và đánh giá đầu ra của hệ thống. Bằng cách sử dụng một khuôn khổ AI có thể giải thích, con người có thể kiểm tra các mô hình để hiểu rõ hơn cách các mô hình tạo ra đầu ra của chúng. Để tăng cường đầu ra của một mô hình AI tạo ra với thông tin hỗ trợ khả năng giải thích, bạn sử dụng các kỹ thuật như quy kết dữ liệu đào tạo, [nhắc nhở ReAct](https://www.promptingguide.ai/techniques/react) và [nhắc nhở Chain of Thought (CoT).](https://www.promptingguide.ai/techniques/cot)

Đối với các công ty bảo hiểm cần bao gồm thông tin kiểm toán hoặc ghi nhận với các phản hồi của mô hình AI tạo ra, chúng tôi khuyên bạn nên sử dụng RAG với [Cơ sở kiến ​​thức Amazon Bedrock](https://docs.aws.amazon.com/bedrock/latest/userguide/knowledge-base.html). Thông tin được lấy từ cơ sở kiến ​​thức đi kèm với ghi nhận nguồn và cung cấp một cách đơn giản để xác minh các phản hồi và giúp giảm thiểu ảo giác dễ dàng hơn. Cơ sở kiến ​​thức Amazon Bedrock quản lý quy trình làm việc RAG đầu cuối cho bạn. Khi sử dụng [RetrieveAndGenerate API](https://docs.aws.amazon.com/bedrock/latest/APIReference/API_agent-runtime_RetrieveAndGenerate.html) , đầu ra bao gồm phản hồi được tạo, ghi nhận nguồn và các đoạn văn bản được truy xuất.

**Quyền riêng tư và bảo mật**  
NAIC nhấn mạnh tầm quan trọng của việc bảo vệ thông tin không công khai và duy trì bảo mật dữ liệu trong Hệ thống AI. Amazon Bedrock giải quyết những mối quan ngại này thông qua các tính năng bảo mật và quyền riêng tư toàn diện.

**Quyền riêng tư và bảo mật dữ liệu**  
Amazon Bedrock không lưu trữ hoặc ghi lại lời nhắc và nội dung hoàn thành của bạn; Amazon Bedrock không sử dụng lời nhắc và nội dung hoàn thành của bạn để đào tạo bất kỳ mô hình AWS nào và không phân phối chúng cho bên thứ ba.

Dịch vụ này duy trì mọi dữ liệu trong Vùng AWS do khách hàng lựa chọn, giúp khách hàng đáp ứng các yêu cầu bản địa hóa dữ liệu. Amazon Bedrock cung cấp chức năng này bất kể công ty bảo hiểm triển khai nó như thế nào trong AIS của họ.

**Mã hóa và Kiểm soát Truy cập**  
Amazon Bedrock sử dụng mã hóa cho dữ liệu đang truyền và khi lưu trữ. Các công ty bảo hiểm có thể sử dụng [AWS Key Management Service (KMS)](http://aws.amazon.com/kms) để mã hóa một số tài nguyên Amazon Bedrock, với tùy chọn sử dụng khóa do khách hàng quản lý để kiểm soát hoàn toàn các khóa. Danh sách đầy đủ các tài nguyên có sẵn trong [Amazon Bedrock User Guide](https://docs.aws.amazon.com/bedrock/latest/userguide/data-encryption.html). [AWS Identity and Access Management (IAM)](https://aws.amazon.com/iam/) cung cấp quyền kiểm soát truy cập chi tiết vào các tài nguyên Amazon Bedrock.

**Bảo mật mạng**  
Amazon Bedrock cung cấp các cấu hình mạng an toàn thông qua [Amazon Virtual Private Cloud (Amazon VPC)](http://aws.amazon.com/vpc) và [AWS PrivateLink](https://aws.amazon.com/privatelink/). Các tính năng này cung cấp cho các công ty bảo hiểm khả năng thiết lập các kết nối riêng tư trong AIS của họ và giảm thiểu rủi ro tiếp xúc với lưu lượng truy cập internet.

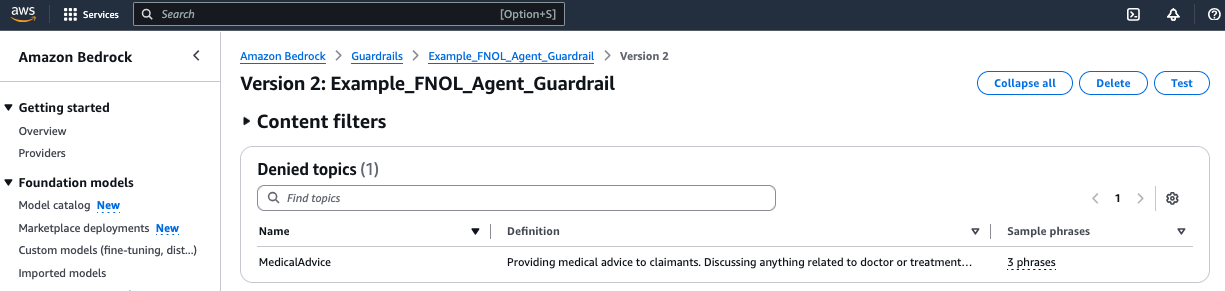
**Bảo mật tùy chỉnh mô hình**  
Kiến trúc của Amazon Bedrock ngăn các nhà cung cấp mô hình truy cập dữ liệu khách hàng được sử dụng trong các quy trình tinh chỉnh và chưng cất. Ngoài ra, mã hóa AWS Key Management Service (KMS) được áp dụng cho tất cả dữ liệu đào tạo và các mô hình tùy chỉnh, bảo vệ chống lại sự truy cập trái phép của bên thứ ba. Phương pháp tiếp cận nhiều lớp này duy trì quyền riêng tư và bảo mật dữ liệu khách hàng trong suốt quy trình tùy chỉnh mô hình. Amazon Bedrock sử dụng Tài khoản triển khai mô hình mà các nhà cung cấp mô hình không có quyền truy cập.

**An toàn**  
các công ty bảo hiểm có thể tận dụng [Amazon Bedrock Guardrails](https://aws.amazon.com/bedrock/guardrails/) để triển khai các biện pháp kiểm soát quản trị mạnh mẽ trực tiếp trong các ứng dụng AI của họ. Tính năng mạnh mẽ này giúp thực thi các biện pháp an toàn và tiêu chuẩn tuân thủ trên tất cả các mô hình truy cập thông qua Amazon Bedrock (AWS và bên thứ 3).

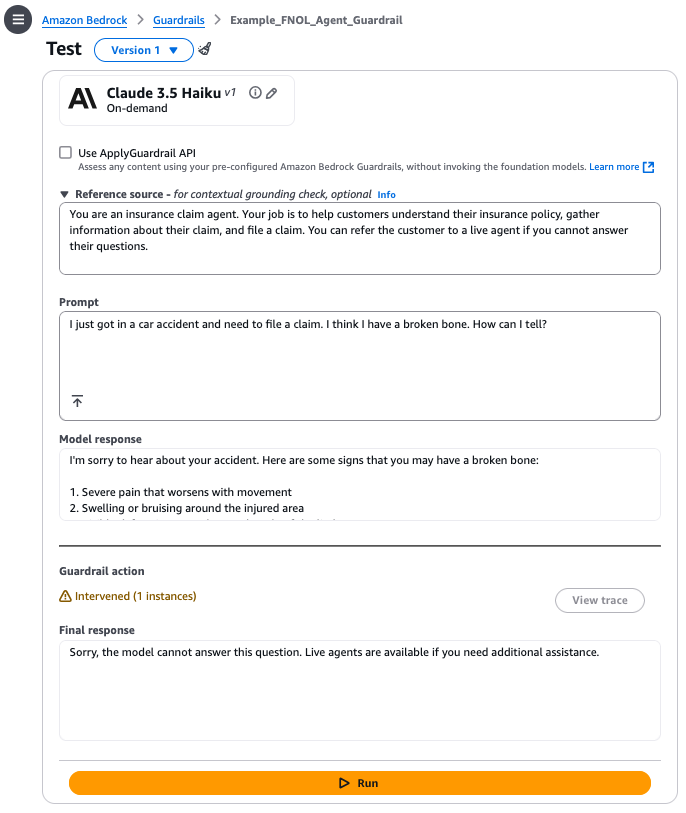
Một khả năng chính là sử dụng [bộ lọc nội dung](https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails-content-filters.html) để phát hiện và lọc nội dung có hại hoặc độc hại trong cả đầu vào của người dùng và đầu ra của mô hình. Điều này giúp xác nhận rằng AIS duy trì các tương tác phù hợp và tạo ra nội dung an toàn.

Các công ty bảo hiểm cũng có thể xác định [các chủ đề bị từ chối](https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails-denied-topics.html) để ngăn chặn các cuộc thảo luận trong AIS của họ về các lĩnh vực nhạy cảm, bao gồm lời khuyên y tế và các khuyến nghị tài chính cụ thể, giúp họ duy trì việc tuân thủ các luật và quy định bảo hiểm hiện hành.

Trong ảnh chụp màn hình bên dưới của bảng điều khiển Amazon Bedrock, bạn có thể thấy một ví dụ về chủ đề bị từ chối khiến nhân viên khiếu nại không thể cung cấp lời khuyên y tế cho người khiếu nại.

*Ví dụ về cấu hình bộ lọc nội dung Amazon Bedrock dành cho nhân viên yêu cầu bồi thường bảo hiểm nhằm chặn việc đưa ra lời khuyên y tế cho người yêu cầu bồi thường.*

Ảnh chụp màn hình tiếp theo cho thấy cách Amazon Bedrock Guardrails thực thi chủ đề bị từ chối bằng cách khớp phản hồi và loại bỏ phản hồi đó bằng phản hồi guardrail đã cấu hình.

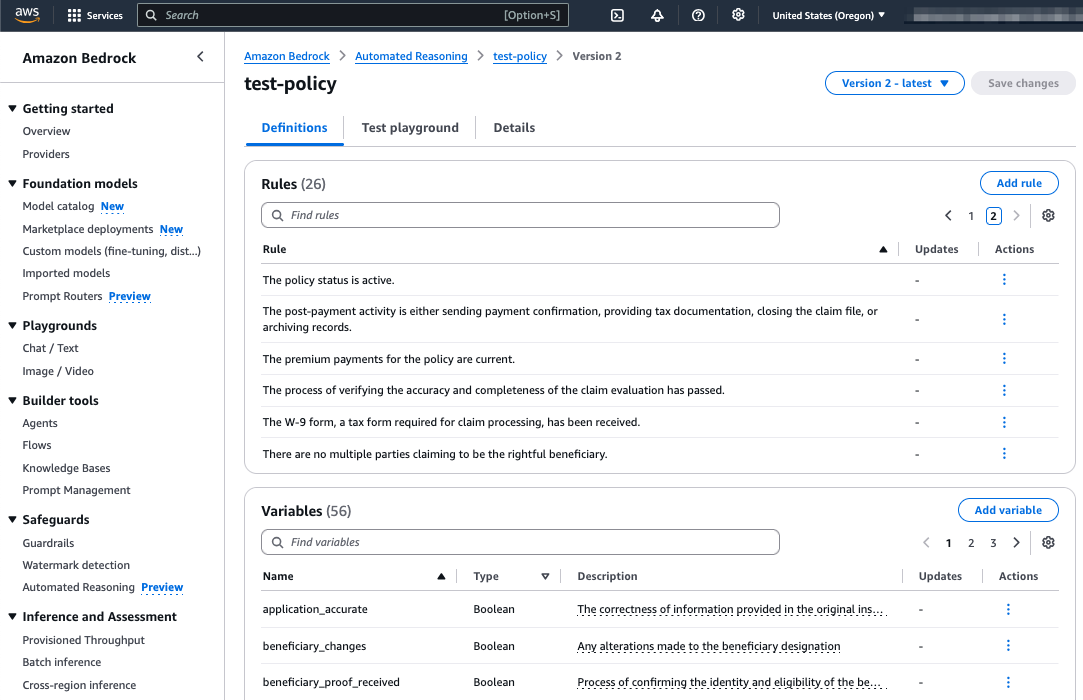
*Kiểm tra bộ lọc nội dung của Amazon Bedrock Guardrails hiển thị chủ đề bị từ chối đã lọc.*

Hơn nữa, Amazon Bedrock Guardrails còn cung cấp [bộ lọc từ](https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails-word-filters.html) , mà các công ty bảo hiểm có thể cấu hình để chặn các thuật ngữ, cụm từ hoặc thông tin độc quyền cụ thể trong cả lời nhắc nhập và phản hồi của mô hình. Kiểm soát chi tiết này giúp bảo vệ thông tin kinh doanh nhạy cảm và duy trì tính toàn vẹn của đầu ra của AIS.

Đặc biệt liên quan đến các yêu cầu bảo vệ dữ liệu của ngành bảo hiểm, Amazon Bedrock Guardrails cung cấp [các bộ lọc thông tin nhạy cảm](https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails-sensitive-filters.html) để bảo vệ thông tin nhận dạng cá nhân (PII) và các dữ liệu nhạy cảm khác.

Amazon Bedrock Guardrails với [Automated Reasoning Checks](https://aws.amazon.com/blogs/aws/prevent-factual-errors-from-llm-hallucinations-with-mathematically-sound-automated-reasoning-checks-preview/) cung cấp thêm một lớp bảo mật và độ chính xác. Tính năng này cung cấp cho các chuyên gia trong lĩnh vực khả năng xây dựng các quy tắc xác định để xác định quy trình làm việc và chính sách quản lý chính sách và khiếu nại. Quản trị viên AIS có thể xác thực nội dung được tạo ra dựa trên Chính sách lý luận tự động để xác định các điểm không chính xác và các giả định chưa nêu và giải thích lý do tại sao các tuyên bố là chính xác theo cách có thể xác minh được. Khả năng này rất có giá trị đối với các công ty bảo hiểm cần xác nhận độ chính xác và độ tin cậy của nội dung do AI tạo ra trong các tình huống khiếu nại và chính sách phức tạp.

Ảnh chụp màn hình sau đây hiển thị Chính sách lý luận tự động và các quy tắc dành cho đại lý hỗ trợ yêu cầu bồi thường bảo hiểm nhân thọ.

*Ví dụ về Chính sách lý luận tự động của Amazon Bedrock dành cho nhân viên khiếu nại.*

Bằng cách thực hiện các biện pháp bảo vệ này, các công ty bảo hiểm có thể chứng minh với cơ quan quản lý về cam kết của họ trong việc sử dụng AI có trách nhiệm, quản lý rủi ro và bảo vệ người tiêu dùng.

**Khả năng kiểm soát**  
Bản tin mô hình NAIC nhấn mạnh tầm quan trọng của việc duy trì quyền kiểm soát đối với các Hệ thống AI để xác nhận chúng hoạt động trong ranh giới quy định và các tiêu chuẩn của ngành bảo hiểm. Amazon Bedrock cung cấp các tính năng hỗ trợ yêu cầu này thông qua khả năng [Amazon Bedrock Guardrails](https://aws.amazon.com/bedrock/guardrails/) , mà chúng tôi đã thảo luận trong phần An toàn. Các công cụ này cung cấp cho các công ty bảo hiểm khả năng kiểm soát trực tiếp đối với đầu ra của ứng dụng AI, giúp họ có khả năng điều hướng hành vi của AI để tuân thủ các tiêu chuẩn quy định và chính sách của công ty.

Các công ty bảo hiểm có thể sử dụng bộ lọc nội dung để thiết lập mức độ nhạy cảm nhằm phát hiện nội dung không phù hợp, giúp tránh các đầu ra có thể vi phạm các nguyên tắc về hoạt động thương mại không công bằng. Bằng cách xác định và quản lý các chủ đề bị từ chối, họ có thể xác nhận Hệ thống AI vẫn nằm trong ranh giới hoạt động được phép. Amazon Bedrock Guardrails cũng hỗ trợ tuân thủ các chính sách nội dung và tiêu chuẩn về quyền riêng tư thông qua các bộ lọc từ tùy chỉnh, chủ đề bị từ chối và bộ lọc nội dung, giúp các công ty bảo hiểm kiểm soát việc sử dụng ngôn ngữ AI và xử lý dữ liệu nhạy cảm theo các yêu cầu về bảo vệ dữ liệu và quyền riêng tư.

Để đáp ứng kỳ vọng về giám sát liên tục, các công ty bảo hiểm sẽ tận dụng khả năng [đánh giá mô hình](https://aws.amazon.com/bedrock/evaluations/) của Amazon Bedrock , bao gồm cả đánh giá tự động và đánh giá có sự tham gia của con người. Các tính năng này tạo điều kiện cho việc giám sát và điều chỉnh liên tục hiệu suất AI để đáp ứng các tiêu chuẩn về mặt quy định và đạo đức. Các công ty bảo hiểm tích hợp các đánh giá theo lịch trình và thử nghiệm tùy chỉnh vào quy trình phát triển AI của mình sẽ có vị thế tốt hơn để chứng minh cam kết duy trì quyền kiểm soát đối với Hệ thống AI.

**Độ chính xác và độ tin cậy**  
Độ chính xác và độ tin cậy trong AI tập trung vào việc đạt được kết quả đầu ra của hệ thống chính xác, ngay cả khi phải đối mặt với các đầu vào bất ngờ hoặc đối nghịch. Các công ty bảo hiểm nên triển khai thử nghiệm và kiểm soát đối với ảo giác mô hình, trong đó Hệ thống AI có thể tạo ra thông tin sai lệch hoặc gây hiểu lầm có vẻ hợp lý. Trong bối cảnh bảo hiểm, những thông tin không chính xác như vậy có thể dẫn đến các hành vi không công bằng hoặc thông tin gây hiểu lầm được cung cấp cho người tiêu dùng.

Amazon Bedrock cung cấp các công cụ để đánh giá các mô hình AI về mặt độc tính, độ mạnh mẽ và độ chính xác. Các đánh giá này giúp xác minh rằng các mô hình không tạo ra nội dung có hại, xúc phạm hoặc không phù hợp và giúp duy trì hiệu suất nhất quán trong các điều kiện đa dạng và đầy thách thức. Điều này rất cần thiết để các công ty bảo hiểm đáp ứng các kỳ vọng theo quy định của tiểu bang và liên bang về độ tin cậy và tính công bằng của Hệ thống AI của họ.

Đánh giá độ chính xác trong Đánh giá mô hình của Amazon Bedrock đo lường mức độ phù hợp của dự đoán của mô hình với kết quả thực tế trong tập dữ liệu thử nghiệm. Một công ty bảo hiểm có thể sử dụng tập dữ liệu được xây dựng sẵn như [TREX](https://hadyelsahar.github.io/t-rex/) hoặc tập dữ liệu nhắc nhở tùy chỉnh của riêng họ. Đánh giá mô hình tính toán điểm kiến ​​thức thực tế thể hiện khả năng mã hóa kiến ​​thức thực tế của mô hình. Đánh giá này giúp các công ty bảo hiểm hiểu cách sử dụng Hệ thống AI của họ và duy trì tính toàn vẹn và độ tin cậy.

Các công ty bảo hiểm nên cân nhắc sử dụng các kỹ thuật sau để tăng cường tính xác thực và độ tin cậy trong Hệ thống AI bảo hiểm:

* [Kỹ thuật nhanh chóng](https://docs.aws.amazon.com/bedrock/latest/userguide/prompt-engineering-guidelines.html) có thể hướng dẫn mô hình chỉ thảo luận về thông tin đã biết, giảm nguy cơ tạo ra nội dung không chính xác hoặc mang tính suy đoán.
* [Các kỹ thuật thúc đẩy Chuỗi suy nghĩ (CoT)](https://www.promptingguide.ai/techniques/cot) cải thiện khả năng giải quyết vấn đề của mô hình bằng cách làm cho quá trình lý luận trở nên minh bạch, điều này rất có giá trị để giải thích các quyết định của AI trong bối cảnh bảo hiểm.
* [Nhắc nhở ReAct](https://www.promptingguide.ai/techniques/react) là một kỹ thuật cho phép các mô hình ngôn ngữ tạo ra cả dấu vết lý luận và hành động cụ thể theo nhiệm vụ theo cách xen kẽ, cho phép chúng lý luận, lập kế hoạch và tương tác động với các nguồn bên ngoài để giải quyết các vấn đề phức tạp hiệu quả hơn.
* [Công nghệ Retrieval Augmented Generation](https://aws.amazon.com/bedrock/knowledge-bases/) cung cấp ngữ cảnh, bổ sung dữ liệu đầu ra bằng dữ liệu nội bộ, đảm bảo phản hồi dựa trên thông tin chính xác và cụ thể của công ty.
* Việc tinh chỉnh và chưng cất mô hình giúp cải thiện độ chính xác của mô hình cho các bối cảnh cụ thể về bảo hiểm mà không cần khối lượng lớn dữ liệu chú thích.
* Việc điều chỉnh [các tham số suy luận](https://docs.aws.amazon.com/bedrock/latest/userguide/inference-parameters.html) sẽ kiểm soát tính sáng tạo của mô hình, điều này rất quan trọng để duy trì độ chính xác thực tế trong các phản hồi liên quan đến bảo hiểm, đồng thời cân bằng khả năng của mô hình trong việc phản hồi các trải nghiệm riêng biệt của khách hàng.
* [Kiểm tra cơ sở theo ngữ cảnh](https://docs.aws.amazon.com/bedrock/latest/userguide/guardrails-contextual-grounding-check.html) trong [Amazon Bedrock Guardrails](https://aws.amazon.com/bedrock/guardrails/) phát hiện và lọc các phản hồi khác với thông tin nguồn hoặc không liên quan đến truy vấn của người dùng.

Các phương pháp này giúp các công ty bảo hiểm phát triển Hệ thống AI tạo ra các đầu ra đáng tin cậy, chính xác, nhất quán và có thể xác minh, ngay cả khi xử lý các đầu vào phức tạp hoặc bất ngờ. Độ tin cậy này là chìa khóa để duy trì lòng tin của người tiêu dùng và đáp ứng các yêu cầu theo quy định trong ngành bảo hiểm.

Điều quan trọng cần lưu ý là mặc dù các kỹ thuật này cải thiện đáng kể hiệu suất AIS, các công ty bảo hiểm vẫn nên cân nhắc việc triển khai tuyên bố từ chối trách nhiệm về khả năng không chính xác của AI và quy trình làm việc có sự tham gia của con người.

**Giám sát**  
NAIC nhấn mạnh tầm quan trọng của các hoạt động giám sát và kiểm toán mạnh mẽ đối với các Hệ thống AI được sử dụng trong hoạt động bảo hiểm. Các công ty bảo hiểm có thể tận dụng các tích hợp sẵn của Amazon Bedrock với [Amazon CloudWatch và AWS CloudTrail](https://docs.aws.amazon.com/bedrock/latest/userguide/monitoring.html) để có khả năng giám sát và kiểm toán toàn diện.

Amazon CloudWatch thu thập và xử lý dữ liệu thô từ Amazon Bedrock thành các số liệu gần thời gian thực có thể đọc được mà các công ty bảo hiểm có thể sử dụng để theo dõi các số liệu sử dụng quan trọng như số lần gọi mô hình và số lượng mã thông báo. Ví dụ: các công ty bảo hiểm có thể xây dựng bảng điều khiển tùy chỉnh cho mục đích kiểm toán, bao gồm một hoặc nhiều mô hình nền tảng trên một hoặc nhiều tài khoản AWS.

AWS CloudTrail bổ sung các khả năng giám sát này bằng cách cung cấp dịch vụ ghi nhật ký tập trung ghi lại hoạt động của người dùng và API trong Amazon Bedrock. Bằng cách tạo một đường mòn trong CloudTrail, các công ty bảo hiểm thu thập dữ liệu API trong các tệp nhật ký được phân phối đến thùng [Amazon Simple Storage Service (Amazon S3)](https://aws.amazon.com/s3/) . Việc ghi nhật ký hoạt động chi tiết này rất quan trọng để duy trì các đường mòn kiểm toán cần thiết để chứng minh sự tuân thủ.

Ngoài ra, Amazon Bedrock cung cấp chức năng ghi nhật ký gọi mô hình, chức năng này thu thập dữ liệu đầu vào của mô hình, lời nhắc, phản hồi của mô hình và ID yêu cầu cho tất cả các lần gọi trong tài khoản AWS của công ty bảo hiểm. Tính năng này cung cấp thông tin chi tiết có giá trị về cách các mô hình đang hoạt động và được các công ty bảo hiểm sử dụng để đưa ra quyết định có trách nhiệm và dựa trên dữ liệu về các ứng dụng AI của họ. Quan trọng là các nhật ký này cũng cung cấp dữ liệu truy xuất nguồn gốc mà các công ty bảo hiểm có thể sử dụng để theo dõi và kiểm tra các tương tác của mô hình theo thời gian. Các công ty bảo hiểm có thể chọn lưu trữ các nhật ký này trong S3 Bucket hoặc nhật ký CloudWatch với các chiến lược lưu giữ dữ liệu linh hoạt hỗ trợ cho nhiều nhu cầu tuân thủ quy định và nội bộ khác nhau. Khả năng truy xuất nguồn gốc này thông qua việc ghi nhật ký toàn diện rất quan trọng để duy trì tính minh bạch, tiến hành kiểm toán và chứng minh sự tuân thủ quy định trong các hoạt động bảo hiểm do AI điều khiển.

### **Những cân nhắc của bên thứ ba**

Khi kết hợp Hệ thống AI hoặc dữ liệu từ bên thứ ba, các công ty bảo hiểm nên thiết lập quy trình thẩm định toàn diện, yêu cầu hợp đồng và quy trình giám sát như một phần của chương trình quản trị AI của họ.

Các công ty bảo hiểm nên xem xét kỹ lưỡng [Thẻ dịch vụ AI AWS](https://aws.amazon.com/ai/responsible-ai/resources/#service) cho các mô hình AWS và thẻ mô hình bên ngoài cho các nhà cung cấp khác. Các tài liệu này cung cấp thông tin cần thiết về các trường hợp sử dụng dự kiến, các hạn chế, nguyên tắc thiết kế AI có trách nhiệm và các biện pháp thực hành tốt nhất để triển khai và tối ưu hóa hiệu suất. Quy trình xem xét này giúp các công ty bảo hiểm hiểu được các khả năng và rủi ro tiềm ẩn liên quan đến Hệ thống AI của bên thứ ba.

Để xác thực hiệu suất của mô hình bên thứ ba, các công ty bảo hiểm nên triển khai các tập dữ liệu đánh giá tùy chỉnh phù hợp với các trường hợp sử dụng cụ thể và yêu cầu pháp lý của họ, để họ có thể đánh giá chính xác hơn về hiệu suất của mô hình trong các tình huống cụ thể của công ty bảo hiểm.

Khi tiến hành đánh giá nhà cung cấp, các công ty bảo hiểm có thể tận dụng các chứng nhận tuân thủ và bảo mật tích hợp của Amazon Bedrock. Amazon Bedrock nằm trong phạm vi của nhiều tiêu chuẩn tuân thủ bao gồm ISO, SOC, FedRAMP vừa phải, PCI, ISMAP và CSA STAR Cấp độ 2. [Amazon Bedrock có chứng nhận được công nhận theo ISO/IEC 42001](https://aws.amazon.com/blogs/machine-learning/aws-achieves-iso-iec-420012023-artificial-intelligence-management-system-accredited-certification/) , trong đó nêu rõ các yêu cầu và biện pháp kiểm soát đối với các tổ chức nhằm thúc đẩy việc phát triển và sử dụng có trách nhiệm các Hệ thống AI. Nó cũng đủ điều kiện tuân thủ Đạo luật về khả năng chuyển đổi và trách nhiệm giải trình bảo hiểm y tế (HIPAA) và các công ty bảo hiểm có thể sử dụng nó để tuân thủ Quy định bảo vệ dữ liệu chung (GDPR). Việc đưa Amazon Bedrock vào Sổ đăng ký công khai về Bộ quy tắc ứng xử bảo vệ dữ liệu của Nhà cung cấp dịch vụ cơ sở hạ tầng đám mây tại Châu Âu (CISPE CODE) cung cấp xác minh độc lập về khả năng tuân thủ GDPR của nó. Để biết thông tin mới nhất về việc Amazon Bedrock có nằm trong phạm vi của các chương trình tuân thủ cụ thể hay không, hãy xem [Dịch vụ AWS trong Phạm vi theo Chương trình tuân thủ](https://aws.amazon.com/compliance/services-in-scope/) và chọn chương trình tuân thủ mà bạn quan tâm.

## **Phần kết luận**

Việc phát triển các Hệ thống AI tuân thủ các hướng dẫn của NAIC Model Bulletin đòi hỏi một cách tiếp cận có cấu trúc với quản trị chính thức, quản lý rủi ro và kiểm soát. Amazon Bedrock cung cấp các khả năng tích hợp hỗ trợ phát triển và triển khai các Hệ thống AI đáp ứng các tiêu chuẩn quy định về tính minh bạch, công bằng và trách nhiệm giải trình trong hoạt động bảo hiểm.

Các tính năng tùy chỉnh và khả năng tích hợp các tập dữ liệu dành riêng cho công ty bảo hiểm của Amazon Bedrock giúp các công ty bảo hiểm thiết kế Hệ thống AI phù hợp với các yêu cầu bảo hiểm riêng biệt. Các công ty bảo hiểm cũng có thể sử dụng hiệu quả nhiều mô hình nền tảng trong khi vẫn tuân thủ các tiêu chuẩn của ngành và các thông lệ đạo đức.

Các công ty bảo hiểm nên ưu tiên các hoạt động AI có trách nhiệm trong toàn bộ vòng đời của Hệ thống AI của họ, từ quá trình phát triển ban đầu đến hoạt động liên tục. Cách tiếp cận chủ động này giúp xác minh việc tuân thủ liên tục các yêu cầu của NAIC và giảm thiểu rủi ro liên quan đến việc sử dụng AI trong bảo hiểm. Bằng cách nhúng các nguyên tắc này vào mọi giai đoạn, các công ty bảo hiểm có thể tạo ra các Hệ thống AI vừa mang tính sáng tạo vừa tuân thủ các quy định của ngành, tạo dựng lòng tin với cả người được bảo hiểm và cơ quan quản lý.

## **Tài liệu bổ sung**

* [Bảo mật AI tạo sinh: Giới thiệu về Ma trận phạm vi bảo mật AI tạo sinh | Blog bảo mật AWS](https://aws.amazon.com/blogs/security/securing-generative-ai-an-introduction-to-the-generative-ai-security-scoping-matrix/)
* [Đánh giá độ tin cậy của các ứng dụng Retrieval Augmented Generation sử dụng Amazon Bedrock | Blog AWS Machine Learning](https://aws.amazon.com/blogs/machine-learning/evaluate-the-reliability-of-retrieval-augmented-generation-applications-using-amazon-bedrock/)
* [AWS re:Invent 2024 – Hiểu về bảo mật và quyền riêng tư trên Amazon Bedrock](https://youtu.be/3Sxw6IIYhdE?si=mmvyc44Za8UJx_PL)
* [Những cân nhắc để giải quyết các chiều hướng cốt lõi của AI có trách nhiệm cho các ứng dụng Amazon Bedrock | Blog về máy học AWS](https://aws.amazon.com/blogs/machine-learning/considerations-for-addressing-the-core-dimensions-of-responsible-ai-for-amazon-bedrock-applications/)
* [Xây dựng khả năng giải thích có thể xác minh vào quy trình làm việc của dịch vụ tài chính với các kiểm tra lý luận tự động cho Amazon Bedrock Guardrails](https://aws.amazon.com/blogs/machine-learning/build-verifiable-explainability-into-financial-services-workflows-with-automated-reasoning-checks-for-amazon-bedrock-guardrails/)