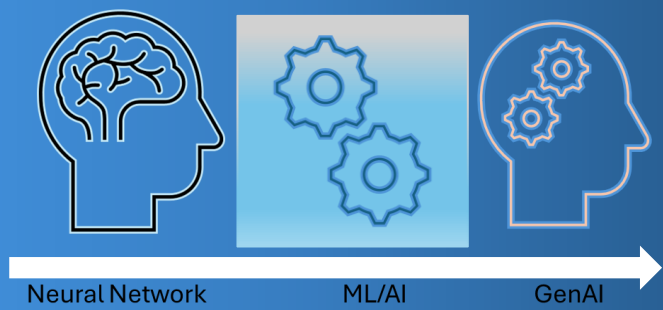




“As the  ears turn”

Advancing AI Model Trust and Security with AIBOMs



 **NeuStratos ProCap360**
“Beyond the BOM knowledge graph”

John Cavanaugh, President
Internet Infrastructure Services Corp

Agenda



- Understanding AI Model Elements
- Ongoing Human Brain Discovery
- AI Techniques and Trends
- The Birth of Generative AI, RAG, Graph RAG
- Automating Model Elements with SBOM and Cyclone DX / SPDX Evolving Automation
- The Future of Generative AI



John Cavanaugh



Education BSEE, MBA

2024 Advanced Technology Academic Research Center Future Secure Work Working Group

2024 DHS CISA SBOM Working Groups

2024 DHS CISA Systems of SBOM Working Group

2024 DHS CISA AIBOM Working Group

2024 OASIS Cybersecurity Automation Sub Project CASP

Certifications ITIL, CISSP

Key Career Notes

Military Power Supply – LAMPS – Navy Aegis Combat System

Precious Metals Refinery – Electronic Pass Key System

Warehouse Management System – Unix to Main Frame

Chase Credit Card – SSO and Electronic Banking

JP Morgan Treasury - Server Virtualization and CMBD enhancement

NSA – Simon/Speck Microprocessor Side Channel Analysis of Key Size

Knowledge Graph research – 2020 COVID Problem

ProCap360 Knowledge Graph Platform for SBOM/RBOM/AIBOM (2021 EO)

“protecting our assets with automation”

Whitepapers

Framing Software Component Transparency: Establishing a Common Software Bill of Materials (SBOM)

<https://www.cisa.gov/resources-tools/resources/framing-software-component-transparency-2024>

Data from Wearables: Emerging Technologies and Emerging Challenges

https://atarc.org/wp-content/uploads/2023/03/ATARC-WP_Wearable-FINAL.pdf

Annex 1: How to Implement Secure Mobile Operational Security White Paper

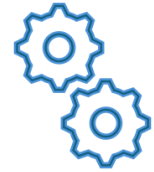
<https://atarc.org/project/annex-1-how-to-implement-secure-mobile-operational-security-white-paper/>

Annex 2: How to Enable the Secure Workforce of the Future by Implementing Classified Tablets and Phones White Paper

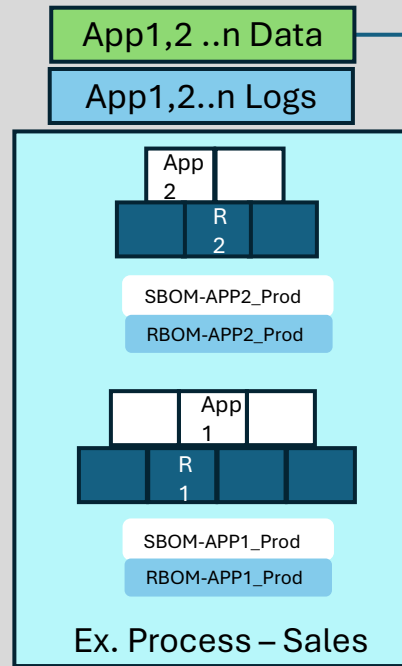
<https://atarc.org/project/white-paper-annex-2-how-to-enable-the-secure-workforce-of-the-future-by-implementing-classified-tablets-and-phones/>

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

The Evolution of AI Models



AI Model X
AIBOM

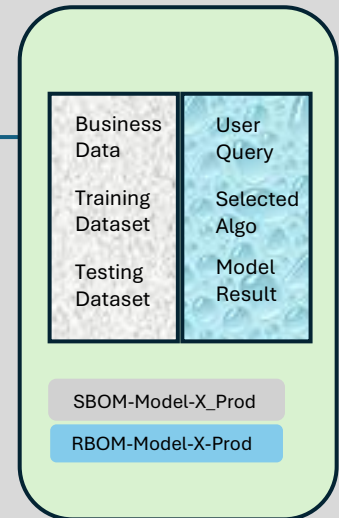


Ex. Process – Sales

Production

Customer Users

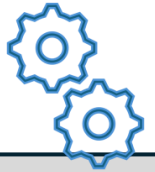
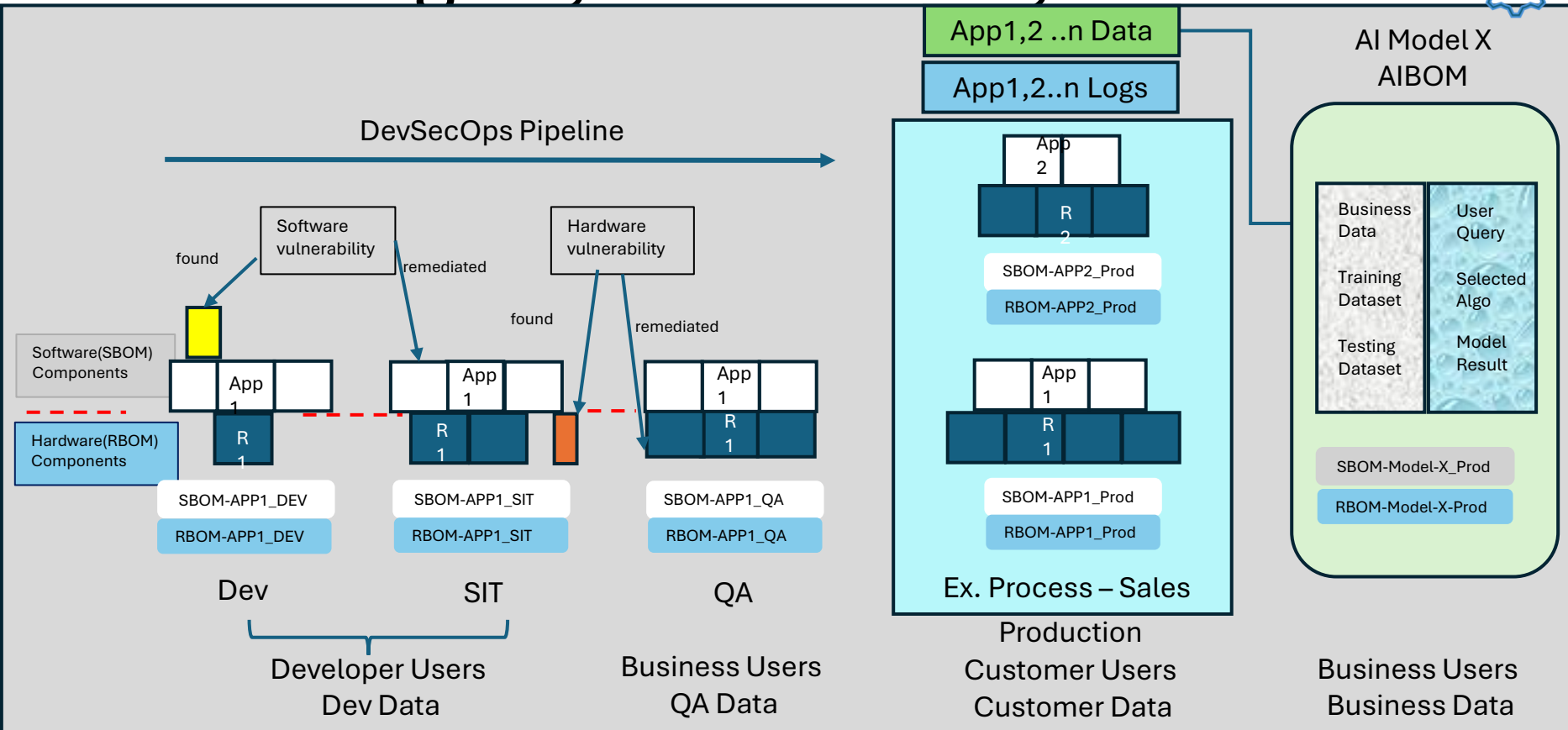
Customer Data



Business Users

Business Data

Data Pedigree, Provenance, Precision

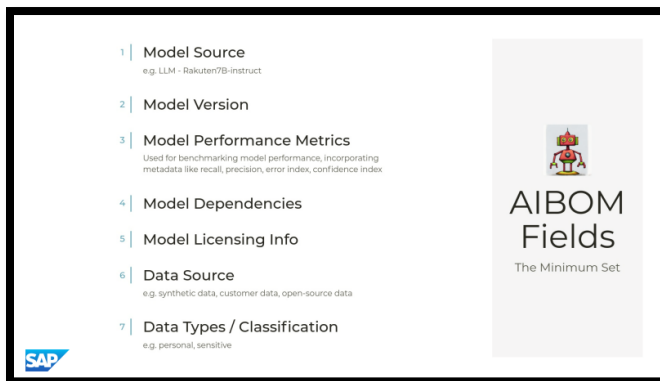


Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

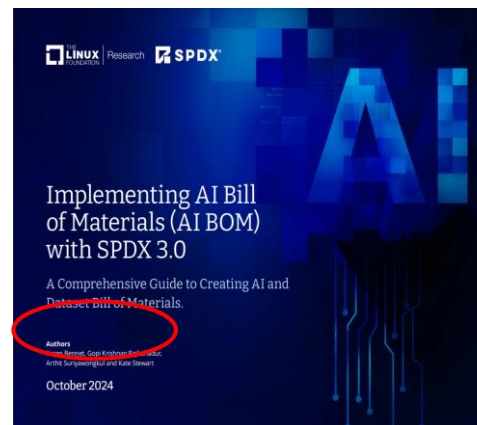
AI Model Elements (AIBOM)



- **Data:** Data Source, Type, Classification,
- **Architecture:** Explain the different components of a generative model (e.g., encoder, decoder, latent space).
- **Training:** Describe the process of training a generative model, including loss functions and optimization techniques.
- **Evaluation:** Discuss metrics used to evaluate the performance of generative models. **Explainability, Explainability accuracy, Meaningful, Knowledge Limits**



DHS CISA AIBOM Tiger Team
Sept 2024



https://www.linuxfoundation.org/research/ai-bom?hss_channel=lcpc-208777

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Field comparison	Category - 27	#ofFields - 84
SPDX original Datasheets	Motivation	2
	Composition	12
	Collection Process	7
	Preprocessing	1
	Uses	3
	Distribution	3
	Maintenance	5
SPDX Model Card	Model Details	8
	Intended Use	2
	Factors	2
	Metrics	2
	Evaluation Data	3
	Training Data	2
	Ethical Consideration	1
SPDX Factsheets + IBM Fields	Recommendation & Caveats	1
	General	5
	Usage	2
	Domain & Application	3
	Basic Performance-Test by Service Provider	2
	Basic Performance-Test by 3rd Party	3
	Safety General	2
	Safety Explainability	1
	Safety Fairness	1
	Safety Concept Drift	3
	Security	4
	Lineage - Training Data	1
	Lineage - Training Model	3

The Human Brain



... The human brain contains billions of neurons, each forming thousands of connections. A single byte comprises 8 bits, and the human brain can store more than one quadrillion bytes of data – a petabyte. As mentioned in an article in Scientific American, the memory capacity of a human brain was testified to have equal to 2.5 petabytes of memory capacity. A “petabyte” means 1024 terabytes or a million gigabytes so that the average adult human brain can accumulate the equivalent of 2.5 million gigabytes of memory. ... The cerebral cortex in the human brain is proficient in possessing 125 trillion synapses, which may stock as much as 2.5 petabytes of overall memory capacity. <https://www.medanta.org/patient-education-blog/what-is-the-memory-capacity-of-a-human-brain>

...The waking adult brain generates only about 20 watts of continuous power—as much as a very dim light bulb. The Salk discovery could help computer scientists build ultraprecise, but energy-efficient, computers, particularly ones that employ “deep learning” and artificial neural nets...

<https://www.salk.edu/news-release/memory-capacity-of-brain-is-10-times-more-than-previously-thought/>emory capacity of brain is 10 times more than previously thought - Salk Institute for Biological Studies

A Thousand Brains: A New Theory of Intelligence



1. **Cortical Columns**

- The brain is organized into thousands of individually computing units called cortical columns³.
- Each column processes information and builds a complete model of the world³.

2. **Thousand Brains Theory**

- Multiple models of the same object or concept are created by different columns³.
- These models work together to form a comprehensive understanding.

3. **Implications for AI**

- Insights from the Thousand Brains Theory can revolutionize AI development⁴.
- AI systems can be designed to mimic the brain's model-building process.

Cortical Columns

- Each column builds a model of the world
- Example: Different columns model various aspects of a cat (shape, color, movement).

Integration

- Columns integrate their models to form a unified perception.
- Example: Combining shape, color, and movement to recognize a cat.

Learning

- Continuous updating and refining of models based on new information
- Example: Learning to distinguish between different breeds of cats.

Conclusion

- The Thousand Brains Theory offers a new perspective on intelligence
- Potential applications in AI and neuroscience.

Jeff Hawkins is an American businessman, neuroscientist, and engineer. He co-founded Palm Computing, where he co-created the PalmPilot and Treo. Later, he founded the Redwood Center for Theoretical Neuroscience in 2002 and Numenta in 2005. His work focuses on understanding the brain's neocortex to develop true artificial general intelligence¹²³.

Source: Conversation with Copilot, 9/17/2024

(1) Is this how your brain works? | Bill Gates.

<https://www.gatesnotes.com/A-Thousand-Brains>.

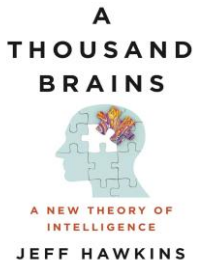
(2) A Thousand Brains Summary - BookBrief.

<https://bookbrief.io/books/a-thousand-brains-jeff-hawkins/summary>.

(3) Jeff Hawkins - Wikipedia.

https://en.wikipedia.org/wiki/Jeff_Hawkins.

[Synaptic Information Storage Capacity Measured With Information Theory | Neural Computation | MIT Press](#)



Source: Conversation with Copilot, 9/17/2024

(1) A Thousand Brains : A New Theory of Intelligence - Google Books. https://books.google.com/books/about/A_Thousand_Brains.html?id=FQ-pzQEACAAJ.

(2) Jeff Hawkins Announces the Thousand Brains Project for AI - IEEE Spectrum. <https://spectrum.ieee.org/jeff-hawkins>.

(3) A Thousand Brains: A New Theory of Intelligence - amazon.com. <https://www.amazon.com/Thousand-Brains-New-Theory-Intelligence/dp/1541675819>.

(4) A Thousand Brains: A New Theory of Intelligence - Goodreads. <https://www.goodreads.com/book/show/54503521-a-thousand-brains>.

(5) A Thousand Brains: A New Theory of Intelligence. https://www.harvard.com/book/a_thousand_brains/.

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Why do we create software applications?

We create application software for several reasons:

1. **Solving Specific Problems**: Application software is designed to address particular needs or problems. For example, word processors help with document creation, while accounting software assists with financial management.
2. **Enhancing Productivity**: Applications like project management tools, email clients, and office suites help users perform tasks more efficiently and effectively.
3. **Facilitating Communication**: Software such as messaging apps, video conferencing tools, and social media platforms enable people to connect and communicate easily.
4. **Entertainment**: Games, music players, and streaming services provide entertainment and leisure activities.
5. **Business Operations**: Enterprise applications like Customer Relationship Management (CRM) systems, Enterprise Resource Planning (ERP) systems, and supply chain management software help businesses streamline their operations and improve efficiency.
6. **Data Management**: Database management systems and data analysis tools help organizations store, manage, and analyze large amounts of data.
7. **Innovation and Development**: Development tools and environments enable programmers to create new software, driving innovation and technological advancement.

In essence, application software is created to make our lives easier, more productive, and more enjoyable. Do you have a specific type of application software in mind that you're interested in?

Copilot response October 2024

“Applications are written to create a digital record of a human activity...”

John Cavanaugh, 1970s

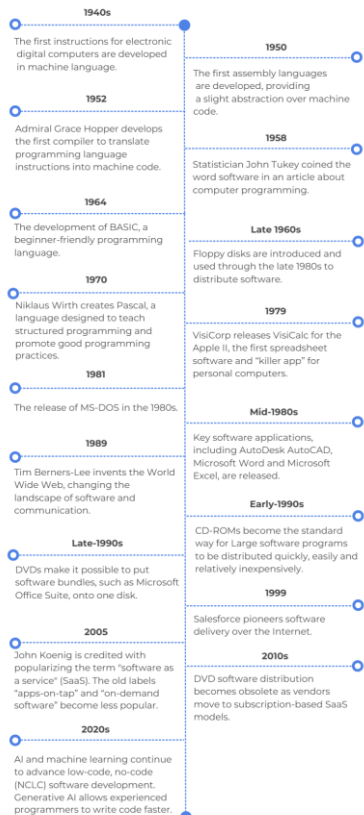
Left of Boom.... Before the incident

Existing tools are Right of Boom.

Computer Science Ontology



History of Software



© 2024

Techopedia

The IEEE in 2019 identified **300 computer programming languages in active use**. They have surveyed several sources and suggest 47 languages have a significant presence. https://iacis.org/iis/2020/2_iis_2020_259-269.pdf

Let's assume that computer applications are written to capture human events. The application output contains application data and logs.

- Application Software

This type of software allows end-users to perform specific tasks. [Customer relationship management \(CRM\)](#) and [enterprise risk management \(ERP\)](#) software are examples of application software.

Software	Proprietary	Open Source
Word Processor	Microsoft Word	LibreOffice Writer
Web Browser	Google Chrome	Mozilla Firefox
Email Client	Microsoft Outlook	Mozilla Thunderbird
Game	The Sims	0 A.D.
Photo Editing Software	Adobe Photoshop	GIMP

+ Development Software

+ System Software

Let's add to these examples the **evolution of mobile smartphones and tablets with separate application data and logs**.

Important Internet of Things (IIOT) can also be added.

<https://www.techopedia.com/definition/4356/software>

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Gen AI



Generative AI, sometimes called *gen AI*, is [artificial intelligence](#) (AI) that can **create original content**—such as text, images, video, audio or software code—**in response to a user’s prompt or request.**

<https://www.ibm.com/topics/generative-ai>

Generative AI relies on **sophisticated machine learning models called [deep learning models](#)**—algorithms that **simulate the learning and decision-making processes of the human brain.** These models work by identifying and encoding the patterns and relationships in huge amounts of data, and then using that information to understand users' natural language requests or questions and respond with relevant new content.

https://www.ibm.com/blog/ai-model-lifecycle-management-deploy-phase/?mhsrc=ibmsearch_a&mhq=ai%20model%20lifecycle%20management%26colon%3B%20deploy%20phase

Application Software- SBOMs



2018 – 2021 NTIA framing and use case documents were published defining SBOMs to be a list of components used in the release of application software. 12 minimum elements were defined for SBOMs that are still in effect to this day. [sbom_formats_survey-version-2021.pdf \(ntia.gov\)](https://www.ntia.gov/sites/default/files/2021-09/sbom_formats_survey-version-2021.pdf)

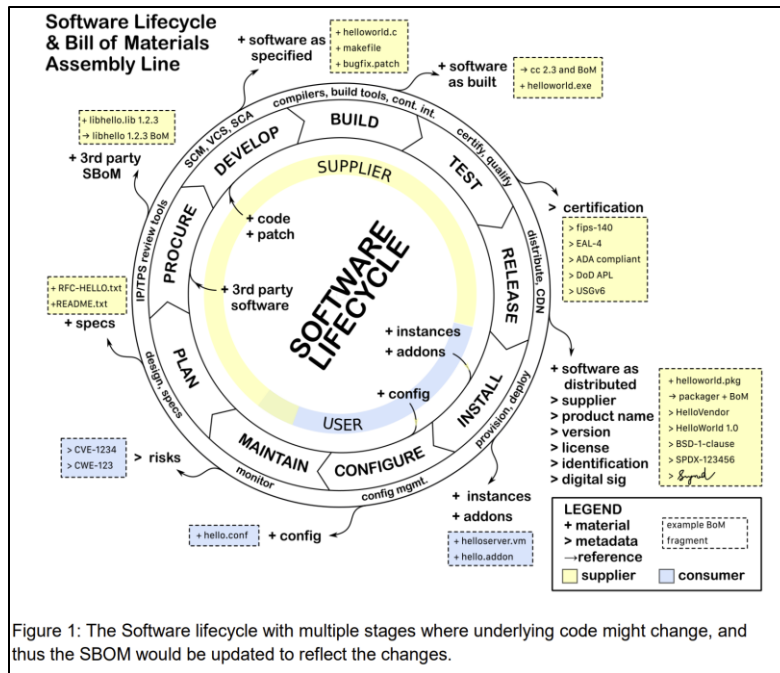


Figure 1: The Software lifecycle with multiple stages where underlying code might change, and thus the SBOM would be updated to reflect the changes.

Executive Order 14028 was signed in 2021 directing Federal Government agencies to request and adopt SBOMs for Commercial software deployments.

SBOM Lifecycle

Plan

Procure/Develop/Test

Release (SBOM)

NEW- Release Bill of Materials (RBOM)

NEW – AI Model Bill of Materials (AIBOM)

AI/ML Techniques



- **Machine Learning**
- **Supervised Learning:**
 - **Regression:** Predicting a continuous numerical value.
 - **Classification:** Categorizing data into discrete classes.
- **Unsupervised Learning:**
 - **Clustering:** Grouping similar data points together.
 - **Dimensionality Reduction:** Simplifying complex data.
- **Reinforcement Learning:**
 - Agent learns through trial and error, maximizing rewards.
- **Deep Learning**
- **Neural Networks:**
 - **Artificial Neural Networks (ANNs):** Simulate the human brain.
 - **Convolutional Neural Networks (CNNs):** Process images and videos.
 - **Recurrent Neural Networks (RNNs):** Process sequential data (text, time series).
- **Transfer Learning:** Reusing trained models on new tasks.
- **Generative AI**
- **Variational Autoencoders (VAEs):**
 - Encode data into a latent space and decode it back.
 - Used for image generation, data denoising, and anomaly detection.
- **Generative Adversarial Networks (GANs):**
 - Two neural networks compete: a generator and a discriminator.
 - Used for image generation, style transfer, and data augmentation.
- **Diffusion Models:**
 - Gradually add noise to data and then learn to remove it.
 - Used for high-quality image and audio generation.
- **Autoregressive Models:**
 - Generate data sequentially, predicting the next element based on previous ones.
 - Used for text generation, music composition, and time series forecasting.
- **Retrieval Augmented Generation**
- **Agent RAG**
- **Graph RAG**

AI Trends

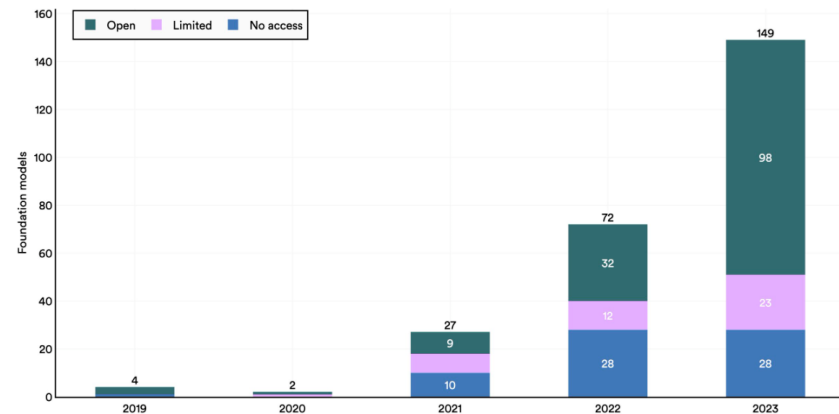
AI Index Report 2024 – Artificial Intelligence Index (stanford.edu)



1. AI beats humans on some tasks but not on all
2. Industry continues to dominate frontier AI research
3. Frontier Models get way more expensive
4. The US leads China, the EU, and the UK as the leading Source of top AI Models
5. Robust and standardized evaluations for LLM Responsibility are seriously lacking.
6. Generative AI investment skyrockets.
7. The data is in: AI makes workers more productive and Leads to higher quality of work.
8. Scientific progress accelerates even further, thanks to AI.
9. The number of AI regulations in the United States sharply increases.
10. People across the globe are more cognizant of AI's potential impact- And are more nervous.

Foundation models by access type, 2019–23

Source: Bommasani et al., 2023 | Chart: 2024 AI Index report



In 2023, a total of 149 foundation models were released, more than double the amount released in 2022. Of these newly released models, 65.7% were open-source, compared to only 44.4% in 2022 and 33.3% in 2021.

“The large, **proprietary models** are phenomenal at **advanced reasoning and breaking down ambiguous tasks**,” explains Salesforce EVP of AI, Jayesh Govindarajan. But for **tasks that are light on reasoning and heavy on crafting language**, for example drafting emails, creating campaign content, researching companies, “**open source models are at par and some are better**,” he said. Moreover, even the high reasoning tasks can be broken into sub-tasks, many of which end up becoming language tasks where open source excels, he said.

But **Meta’s openly available models have now been downloaded more than 400 million times**, the company told VentureBeat, at a rate **10 times higher than last year**, with usage **doubling from May through July 2024**.

<https://venturebeat.com/ai/the-enterprise-verdict-on-ai-models-why-open-source-will-win/>

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation



AI Trends

AI Index Report 2024 – Artificial Intelligence Index ([stanford.edu](https://aiindex.stanford.edu/))

Reported responsible AI benchmarks for popular foundation models

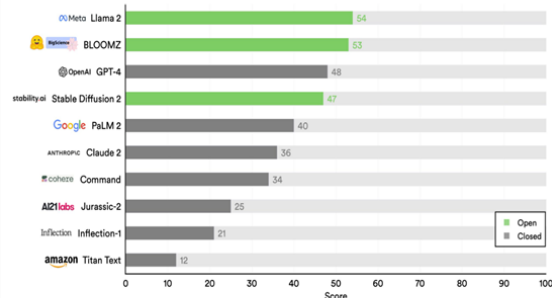
Source: AI Index, 2024 | Table: 2024 AI Index report

Responsible AI benchmarks	GPT-4	Llama 2	Claude 2	Gemini	Mistral 7B
TruthfulQA	✓	✓	✓		
RealToxicityPrompts	✓			✓	
ToxiGen		✓			
BOLD		✓			
BBQ			✓	✓	

New research from the AI Index reveals a significant lack of standardization in responsible AI reporting. Leading developers, including OpenAI, Google, and Anthropic, primarily test their models against different responsible AI benchmarks. This practice complicates efforts to systematically compare the risks and limitations of top AI models.

Foundation model transparency total scores of open vs. closed developers, 2023

Source: 2023 Foundation Model Transparency Index



The newly introduced Foundation Model Transparency Index shows that AI developers lack transparency, especially regarding the disclosure of training data and methodologies. This lack of openness hinders efforts to further understand the robustness and safety of AI systems.

Notable foundational and frontier AI models released between 2020 and 2024:

Older Models

1. Word2Vec (2013): Developed by Google, this model revolutionized natural language processing by using neural networks to create word embeddings.
2. BERT (2018): Developed by Google, BERT (Bidirectional Encoder Representations from Transformers) improved the understanding of context in NLP tasks.
3. GPT-2 (2019): OpenAI's Generative Pre-trained Transformer 2, known for its ability to generate coherent and contextually relevant text.
4. GPT-3 (2020): The successor to GPT-2, GPT-3 significantly increased the scale and capabilities of language generation.
5. DALL-E (2021): OpenAI's model that generates images from textual descriptions, showcasing the potential of multimodal AI.
6. CLIP (2021): Another OpenAI model, CLIP (Contrastive Language-Image Pre-training) connects images and text, enabling zero-shot learning for various tasks.
7. Codex (2021): OpenAI's model that powers GitHub Copilot, capable of generating code from natural language descriptions.
8. IBM Watson (2011): IBM's AI platform that gained fame for winning the Jeopardy! game show. It has since evolved to support various applications in healthcare, finance, and customer service.
9. IBM Project Debater (2019): An AI system capable of engaging in complex debates with humans, showcasing advanced natural language understanding and generation capabilities.

Foundational Models

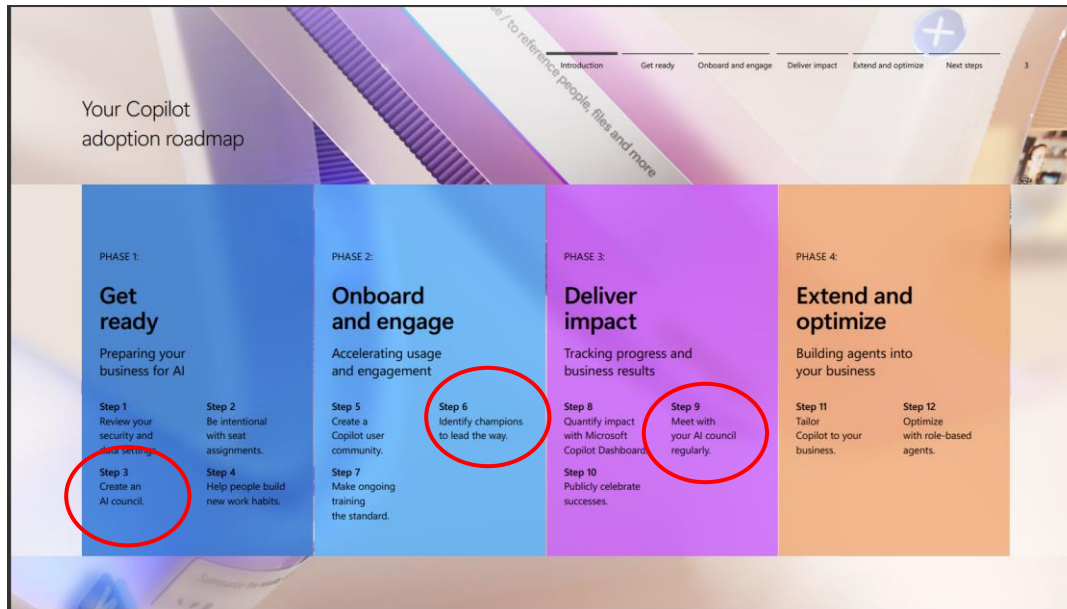
1. GPT-4 (2023): OpenAI's latest language model, pushing the boundaries of natural language understanding and generation.
2. PaLM (2022): Google's Pathways Language Model, designed to handle a wide range of NLP tasks with high accuracy.
3. Gato (2022): Developed by DeepMind, Gato is a generalist agent capable of performing multiple tasks across different domains.
4. Gemini (2024): A cutting-edge model developed by the Frontier Model Forum, representing the next generation of AI capabilities.
5. IBM Granite Models (2024): The Granite series is IBM's flagship collection of foundation models based on a decoder-only transformer architecture. These models are trained on diverse and trusted enterprise data, and are available on platforms like IBM watsonx, RedHat, and open-source platforms such as Hugging Face and GitHub.

Frontier Models

Transformer + Latent Space Focus + Inference Specific

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

The AI Horse has left the barn.....



The September release of **Llama 3.2** introduced Llama Guard Vision, adding to safety tools released in July. These tools can:

Detect potentially problematic text and image inputs before they reach the model.
Monitor and filter output responses for safety and compliance
Enterprise

AI providers have built upon these foundational safety features.

AWS's Bedrock service, for example, allows companies to establish consistent safety guardrails across different models. "Once customers set those policies, they can choose to move from one publicly available model to another without actually having to rewrite the application," explains AWS' Sridharan. This standardization is crucial for enterprises managing multiple AI applications.

Many closed sourced companies have offered users "indemnification," or protection against legal risks or claims lawsuits as a result of using their LLMs. Open source providers usually do not provide such indemnification.

<https://venturebeat.com/ai/the-enterprise-verdict-on-ai-models-why-open-source-will-win/>

Microsoft 365 Copilot Adoption Playbook (clouddamcdnprodep.azureedge.net)

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation



Model Data Provenance Good vs Bad

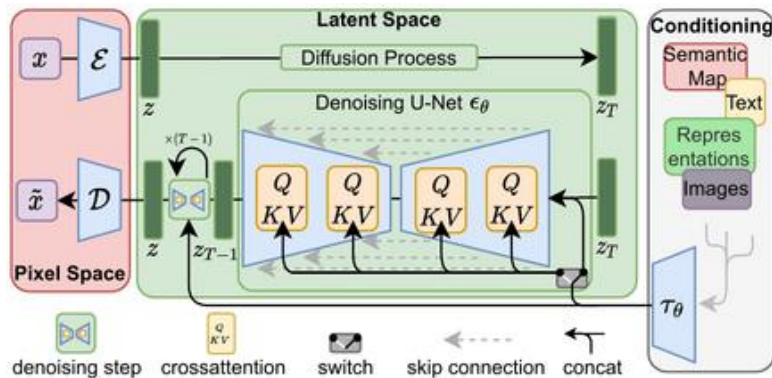
IMAGENET

ImageNet is an image dataset organized according to the WordNet hierarchy.

<https://image-net.org/>

Feature	ImageNet	LAION-5B
Size	14 million images	5.85 billion image-text pairs
Content	Hand-annotated images	CLIP-filtered image-text pairs
Language	Primarily English	Multilingual (2.32B English, 2.2B other languages)
Purpose	Image classification and object detection	Training large-scale multi-modal models
Accessibility	Requires acceptance of terms for download	Openly available for research purposes
Challenges	Limited to specific categories	Uncurated, potential for harmful content
Release Date	2009	15-Oct-22

LAION used Common Crawl for its scrape of the internet which includes billions of copyrighted images taken by photographers.



Stable Diffusion architecture - Stable Diffusion - Wikipedia

Provenance Example <https://purl.stanford.edu/kh752sm9123> Identifying and Eliminating CSAM in Generative ML Training Data and Models
Generative Machine Learning models have been well documented as being able to produce explicit adult content, including child sexual abuse material (CSAM) as well as to alter benign imagery of a clothed victim to produce nude or explicit content. In this study, we examine the LAION-5B dataset—parts of which were used to train the popular Stable Diffusion series of models—to attempt to measure to what degree CSAM itself may have played a role in the training process of models trained on this dataset. This methodology detected many hundreds of instances of known CSAM in the training set, as well as many new candidates that were subsequently verified by outside parties. We also provide recommendations for mitigating this issue for those that need to maintain copies of this training set, building future training sets, altering existing models and the hosting of models trained on LAION-5B.

Copyright

Content

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Generative Adversarial Networks (GAN)



Election Impact of AI.

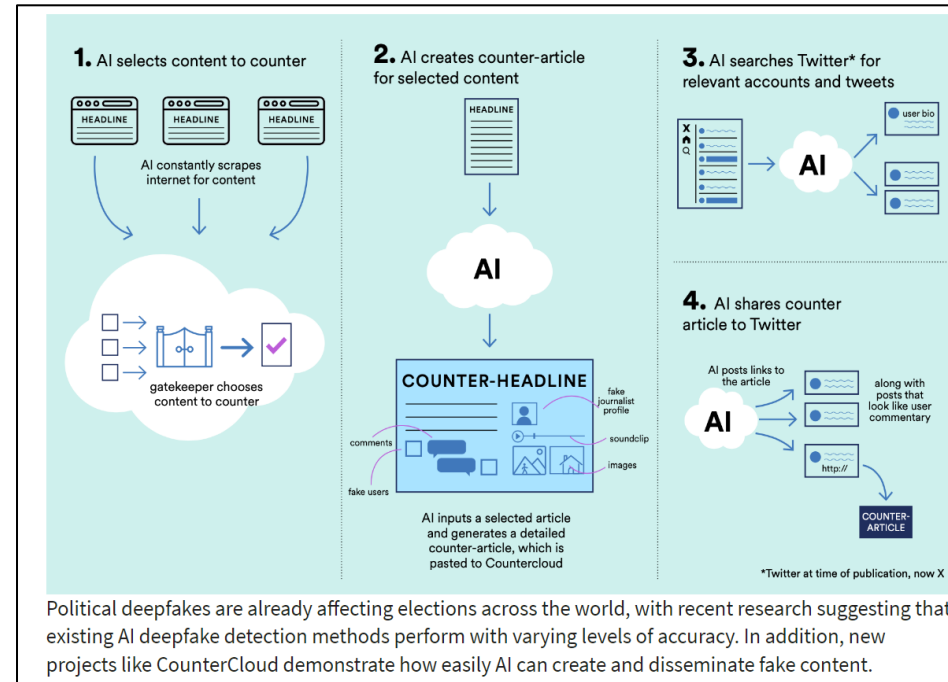
Given that 4 billion people are expected to vote globally in 2024, explores the potential impact of AI on political processes.

"Currently, deep fakes are already causing concern as the country heads into an election season. Eyre argues that there's an urgent need to design neuroscience-based policies to support citizens against AI—such as a "neuroshield."

"The way we interpret the reality around us, the way we learn and react, depends on the way our brains are wired," he writes. "It has been argued that, given the rapid rise of technology, evolution has not been given enough time to develop those regions of the neocortex which are responsible for higher cognitive functions. As a consequence, we are biologically vulnerable and exposed."

<https://techxplore.com/news/2023-07-neuroshield-citizens-artificial-intelligence-neuroscience.html>

Model
Transparency



Political deepfakes are already affecting elections across the world, with recent research suggesting that existing AI deepfake detection methods perform with varying levels of accuracy. In addition, new projects like CounterCloud demonstrate how easily AI can create and disseminate fake content.

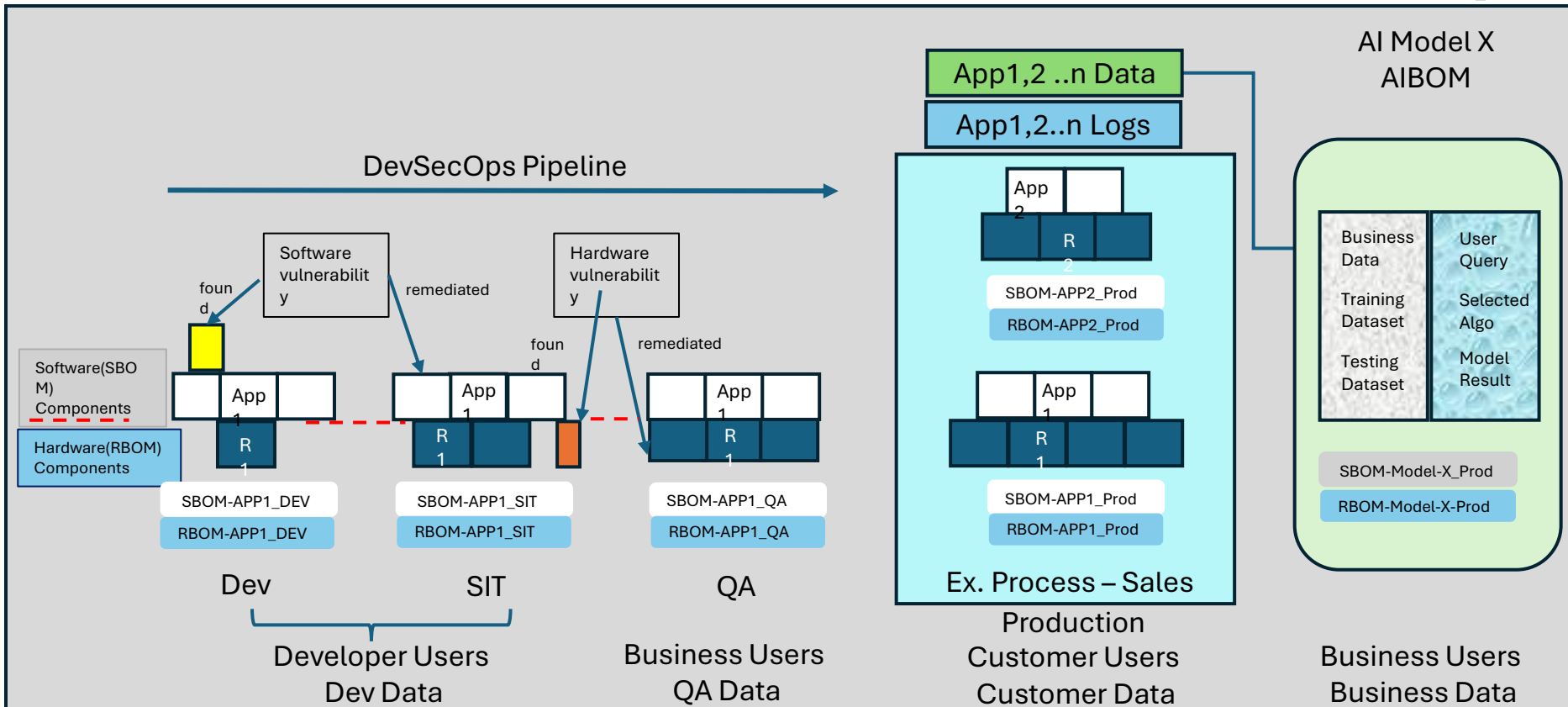
The C2PA is a [Joint Development Foundation](#) project to collectively build an end-to-end open technical standard to provide publishers, creators, and consumers with opt-in, flexible ways to understand the authenticity and provenance of different types of media. C2PA opposes efforts to make content provenance measures mandatory.

<https://c2pa.org/>

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

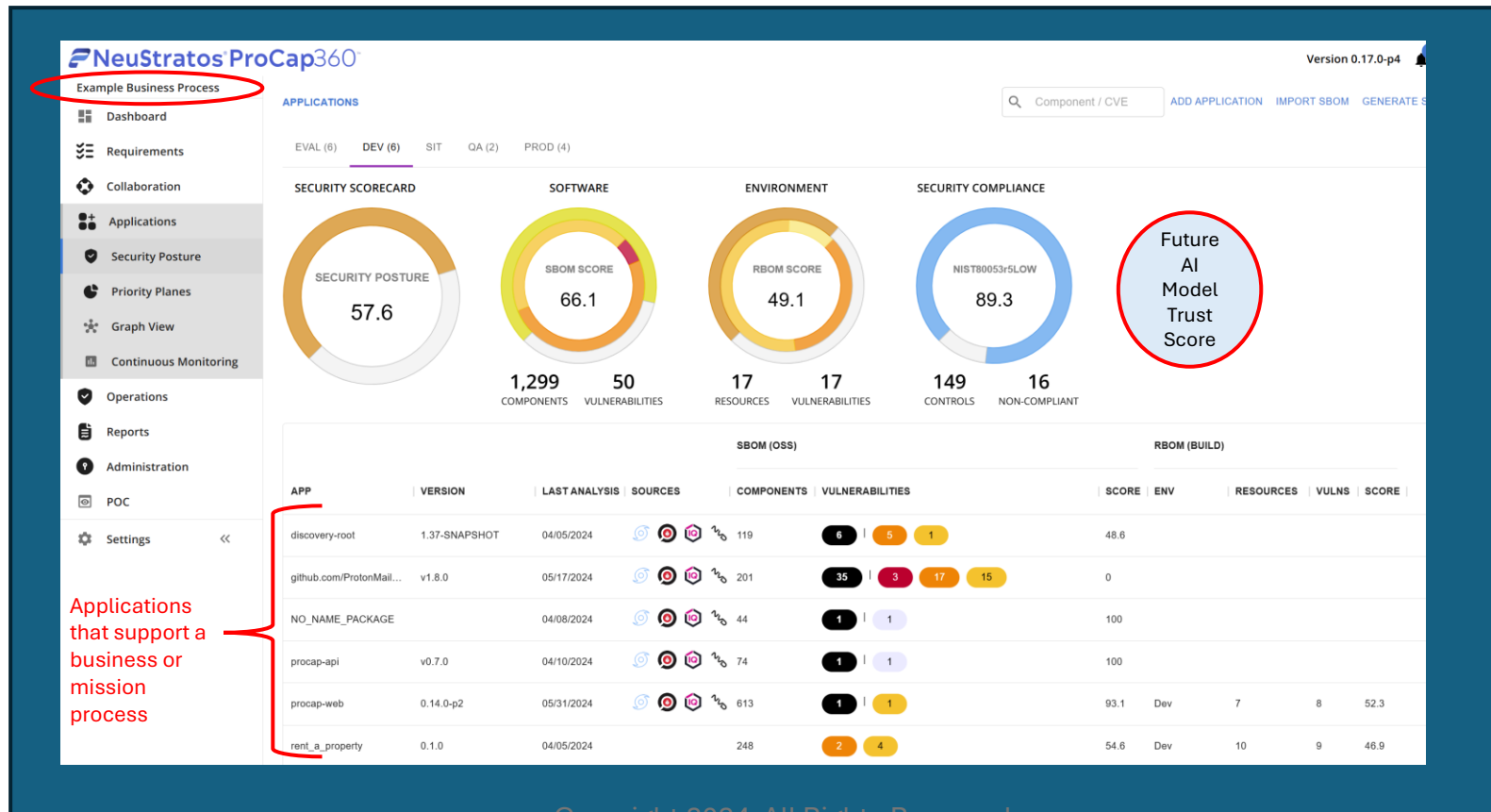
Copyright

SBOM + RBOM + AIBOM



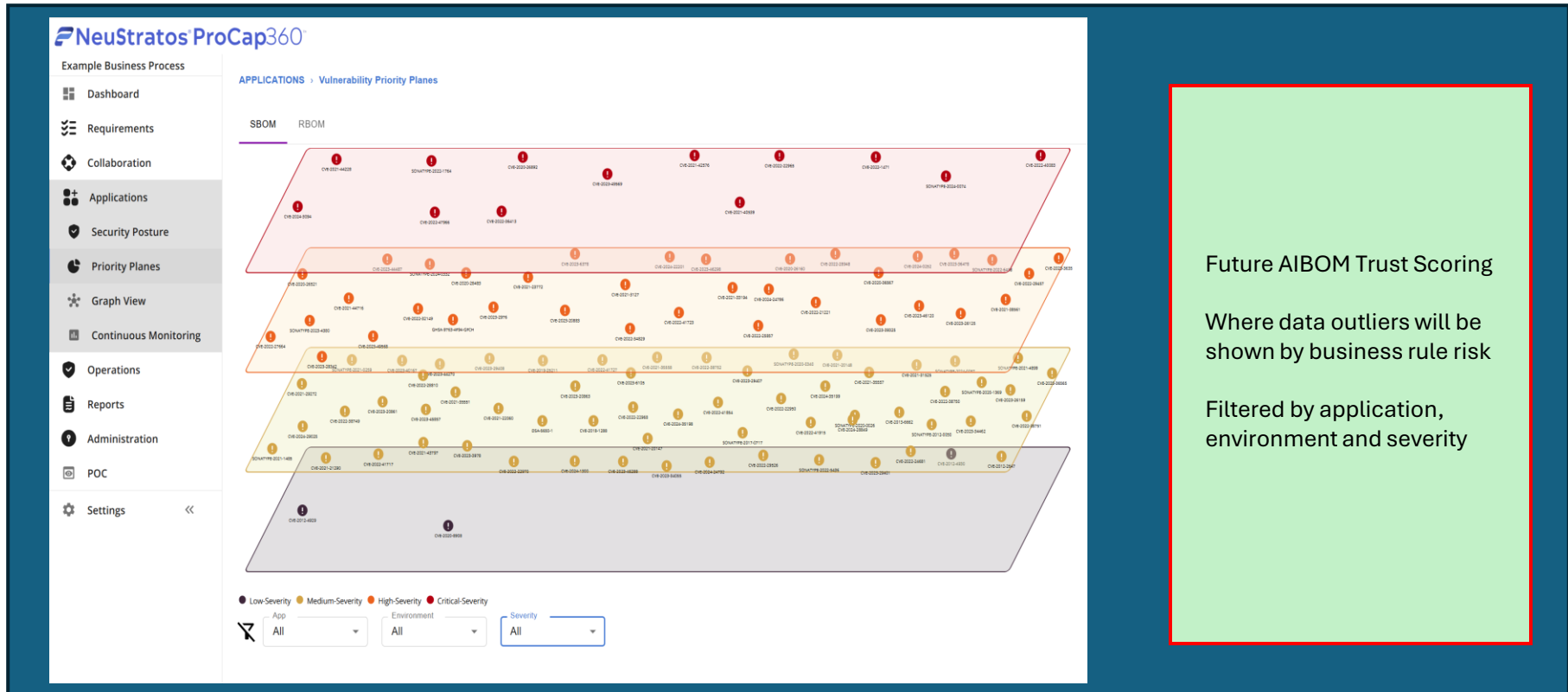
Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Business Process Assurance



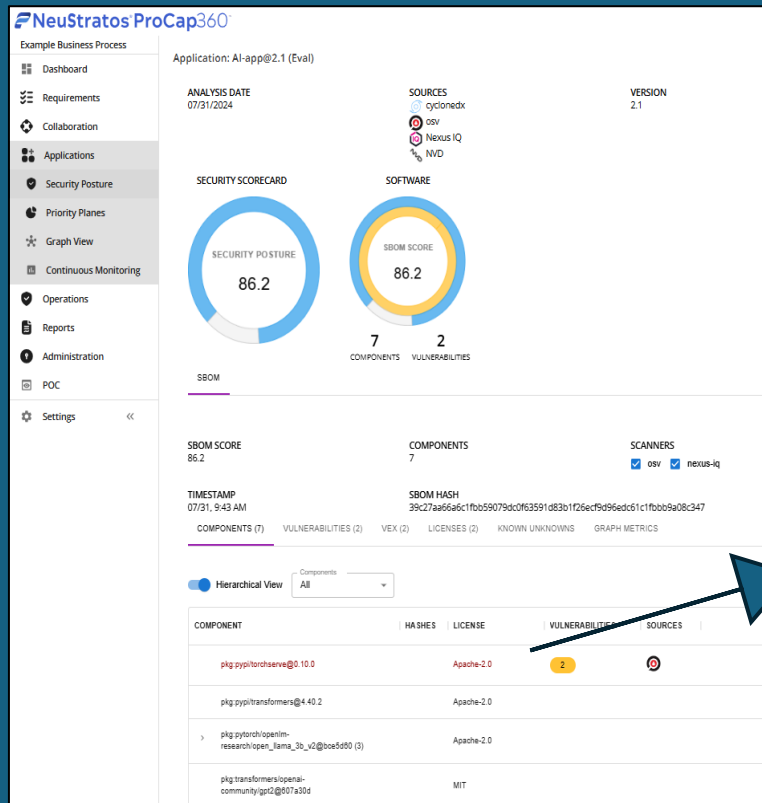
Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Risk Based Vulnerability Visualization



Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

AIBOM Software Component(s) with Vulnerabilities



pytorch

PACKAGE URL

pkg:pypi/torchserve@0.10.0

FORMAT

pypi

HASH

LICENSE

Apache-2.0

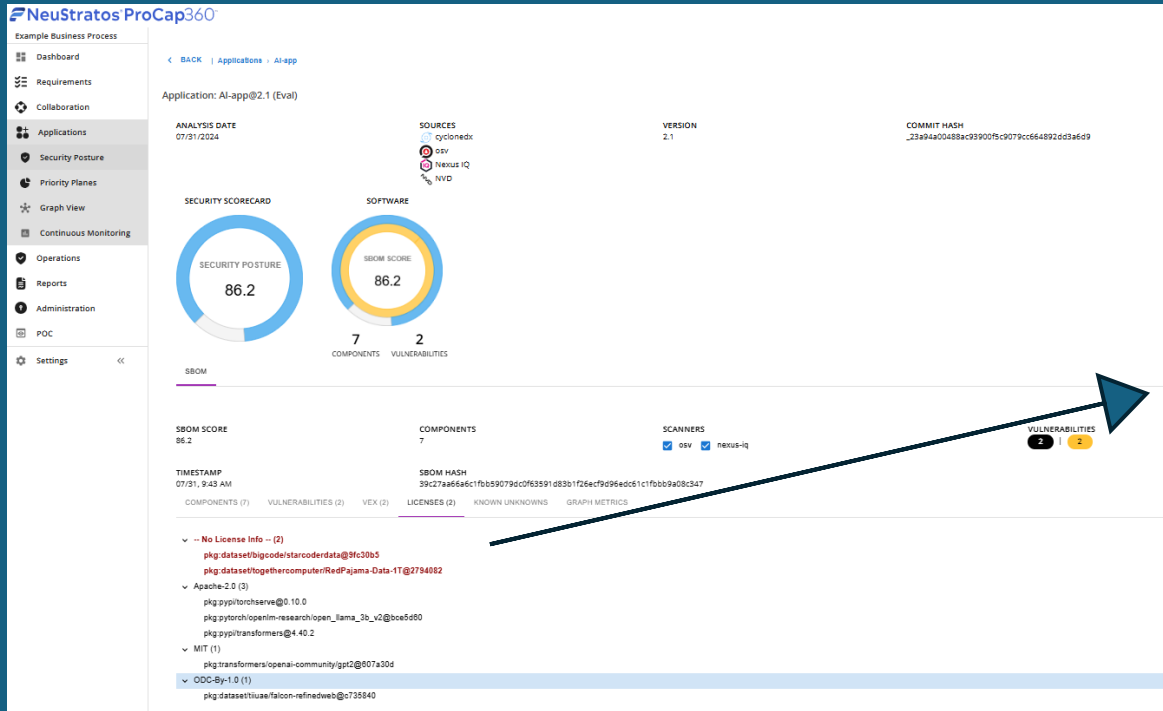
VULNERABILITIES (2)

DEPENDENCIES

ACTION	VULNERABILITY ID	DESCRIPTION	SEVERITY	CVSS	EPSS	KEV	KRI	CWE	UPGRADE PATH	VEX STATE	REFERENCES
<div><div>—</div><div>Analyse</div></div>	CVE-2024-35198	TorchServe is a flexible and easy-to-use tool for serving and scaling PyTorch models in production. TorchServe's...	medium	5.7	0.00045		0.002565	706	0.11.1		GHSA-WXCK-GG9C-FWP2
VULNERABILITY ID		CVSS		CWE		SEVERITY		SOURCE			
CVE-2024-35198		5.7		706		medium		NVD			
EXPLANATION			UPGRADE PATH								
<p>TorchServe is a flexible and easy-to-use tool for serving and scaling PyTorch models in production. TorchServe's check on <code>allowed_urls</code> configuration can be by-passed if the URL contains characters such as <code>..</code> but it does not prevent the model from being downloaded into the model store. Once a file is downloaded, it can be referenced without providing a URL the second time, which effectively bypasses the <code>allowed_urls</code> security check. Customers using PyTorch inference Deep Learning Containers (DLC) through Amazon SageMaker and EKS are not affected. This issue in TorchServe has been fixed by validating the URL without characters such as <code>..</code> before downloading see PR #3082. TorchServe release 0.11.0 includes the fix to address this vulnerability. Users are advised to upgrade. There are no known workarounds for this vulnerability.</p>			<div><div>▼ torchserve@0.10.0 (vulnerable)</div><div>torchserve@0.11.1</div><div>torchserve@0.11.0</div></div>								
RECOMMENDATION											

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

AIBOM Model License Provenance



- ✓ -- No License Info -- (2)
 - pkg:dataset/bigcode/starcoderdata@9fc30b5
 - pkg:dataset/togethercomputer/RedPajama-Data-1T@2794082
- ✓ Apache-2.0 (3)
 - pkg:pypi/torchserve@0.10.0
 - pkg:pytorch/openlm-research/open_llama_3b_v2@bce5d60
 - pkg:pypi/transformers@4.40.2
- ✓ MIT (1)
 - pkg:transformers/openai-community/gpt2@607a30d
- ✓ ODC-By-1.0 (1)
 - pkg:dataset/tiiuae/falcon-refinedweb@c735840

As the ears turn.....AIBOMs



Manual Risk Assessment



Automated Risk Assessment



Automated Operational
Assurance

AI Risk Repository

A Comprehensive Database of Risks from AI Systems

OVERVIEW

DATABASE

CAUSAL TAXONOMY

DOMAIN TAXONOMY

HOW TO USE

FAQ

NEWS

What is the AI Risk Repository?

The AI Risk Repository has three parts:

- The **AI Risk Database** captures 700+ risks extracted from 43 existing frameworks, with quotes and page numbers.
- The **Causal Taxonomy of AI Risks** classifies how, when, and why these risks occur.
- The **Domain Taxonomy of AI Risks** classifies these risks into seven domains (e.g., "Misinformation") and 23 subdomains (e.g., "False or misleading information").

<https://airisk.mit.edu/>

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

AI Regulations



https://airc.nist.gov/AI_RMF_Knowledge_Base/Roadmap

[TikTok to automatically label AI-generated user content in global first \(ft.com\)](https://www.ft.com/content/ai-generated-user-content-in-global-first)

<https://www.ai.gov/wp-content/uploads/2023/09/FAQs-on-Foundation-Models-and-Generative-AI.pdf>

<https://www.nextgov.com/artificial-intelligence/2024/07/fcc-proposes-requiring-advertisers-disclose-ai-political-ads-radio-tv/398341/>

Copyright 2024 All Rights Reserved
Internet Infrastructure Services Corporation

Conclusion – Gen AI

- ✓ Human Actions are recorded as digital data
- ✓ Data has x, y, z, t, pedigree/provenance dimensions
- ✓ Data Pedigree (Model Data)
- ✓ Data Provenance (Training and Testing Data)
- ✓ AI Model Risks (Data and Structural)
- ✓ AI Bill of Materials (AIBOM) is a machine-readable record of data used, query, result
- ✓ A knowledge graph repository of AIBOM model use in business processes with Application SBOMs and Release RBOM components supports Responsible Gen AI.

