ahia

Assoc. of Healthcare Internal Auditors

Auditing Artificial Intelligence

CAITLIN HOLLERAN, CHIEF COMPLIANCE OFFICER
CHASE FRANZEN, CHIEF INFORMATION SECURITY OFFICER
SHARP HEALTHCARE

CASEY KACERIK, SENIOR MANAGER
DELOITTE

Meeting with you today

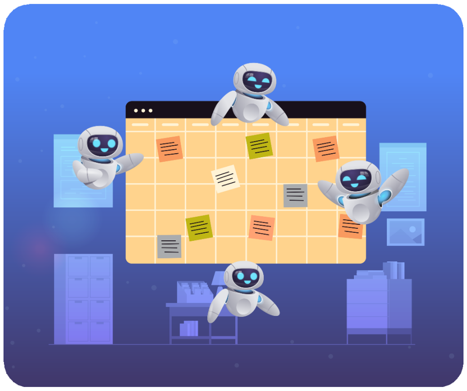
Casey Kacerik

Chase Franzen

Caitlin Holleran

Sharp GPT

Agenda & Objectives



AI Landscape

AI in Healthcare

Health System AI

AI Risk Domains

AI Governance & Best Practices – IA Specific View

Discussion, Q&A

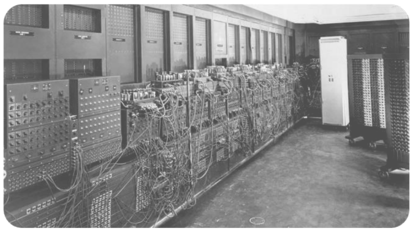
By the end of this session, you should:

- ✓ Develop a comprehensive understanding of the AI landscape, including its definition, significance, and current state.
- ✓ Demystify the development and applications of AI, enabling participants to grasp the process and identify common uses.
- ✓ Recognize the specific applications of AI in the healthcare industry and understand their benefits and challenges.
- ✓ Gain awareness of the various risk domains associated with AI and understand the importance of AI Governance and best practices.
- ✓ Equip internal auditors with the skills to effectively assess AI-related risks within the organization by learning appropriate questioning techniques.

AI Landscape

Brief History of AI

Timeline demonstrating the advances in artificial intelligence and how it continues to revolutionize various industries.



1950s

Turing Test: Alan Turing invents the Computer and invents the "Turing Test"

1970s

AI Winter: A period of reduced funding and lowered interest in the field as there were many false starts and dead ends.

1980s

Narrow AI: Research shifted toward creating "expert systems" focused on performing a single task instead of trying to create general intelligence.

2000s

Deep Blue: IBM's chess-playing computer defeats world chess champion, Garry Kasparov.

2010s

Siri, Google Now, Cortana: Personal assistants use speech recognition to answer questions and perform simple tasks.

2017

"Attention is all you need": Paper published by Google research introducing the Transformer model, revolutionizing NLPs.

2022

Release of ChatGPT: "AI chatbox that uses natural language processing to create humanlike conversational dialogue".

AI is inherently about augmenting humans with machines to reach greater heights

AI mimics the way humans perceive information, devise insights based on experience, and make decisions accordingly

OXFORD DICTIONARY DEFINITION:

ARTIFICIAL INTELLIGENCE (AI) is the theory and development of computer systems able to perform tasks normally requiring human intelligence

AI encompasses many technologies that work together to build innovative solutions that transform society and business…

Machine Learning

Natural Language Processing

Neural Networks

Predictive Analytics

Virtual Assistants

Deep Learning

Speech Recognition

Intelligent Automation

Robotics Process Automation

Unsupervised Learning

Computer Vision

Unlocking the Power of Language: Understanding Generative AI

WHAT is Generative AI | artificial intelligence that creates **original content across various modalities** *(e.g., text, images, audio, code, voice, video)* that would have previously taken human skill and expertise to create

HOW does it work | Generative AI is powered by **foundation models** such as OpenAI's GPT-4 , NVIDIA's Megatron, and Google's PaLM, which are trained on **vast amounts of data and computation** to perform a broad range of downstream tasks

WHY now | innovations in **machine learning** and the **cloud tech stack**, coupled with the **viral popularity** of publicly released applications have propelled Generative AI into the zeitgeist

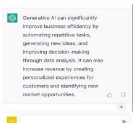
WHO is involved | **Big Tech** is building—and enabling access to—foundation models; **start-ups** are developing user applications on these underlying models; and **companies** are beginning to adopt

POTENTIAL BUSINESS IMPACT | the **marginal cost of producing initial versions of knowledge-intensive content**—such as IT code, marketing copy, and creative design—**can decrease significantly**

EXAMPLE MODALITIES

Text Generation

Prompt: *Explain my colleagues the business impact of generative AI in 50 words*



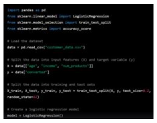
Generative AI can significantly improve business efficiency by automating repetitive tasks, generating new ideas, and improving decision-making through data analysis. It can also increase revenue by creating personalized experiences for customers and identifying new market opportunities.

Image Generation

Prompt: *A bowl of soup that is a portal to another dimension as digital art*

Code Generation

Prompt: *In python, code a program that predicts the likelihood of customer conversion*



```
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Load the dataset
data = pd.read_csv("customer_data.csv")

# Split the data into input features (X) and target variable (y)
X = data[["age", "income", "num_products"]]
y = data["converted"]

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a logistic regression model
model = LogisticRegression()
```

Video Generation

Prompt: *A teddy bear painting a portrait*

Audio Generation

Prompt: *Play 'we have to reduce the number of plastic bags' in a sleepy tone*

Speaker's Emotion Maintenance

E can synthesize personalized speech while maintaining the emotion in the speaker prompt. The audio prompts are sampled from the Emotional Voices base.

Text	Emotion	Speaker Prompt	VALL-E
ave to reduce the number of plastic bags.	Anger	▶	▶
	Sleepy	▶	▶
	Neutral	▶	▶
	Amused	▶	▶
	Disgusted	▶	▶

Generative AI comes with risks and limitations

There are several limitations to consider when using Generative AI

Bias in; bias out. If the training data is biased (e.g., over/under-representation of a population cohort, sexism, racism), then outputs generated could exhibit biases as well. Bias reductions in the training data and/or human supervision during model training is needed

Bias

Foundation Models generally offer a pay-as-you-go billing mechanism, and the cost per use of sophisticated models is materially significant. Fine tuning the biggest model and running large documents through several times could easily run up a bill of tens of US $1000s

Cost

Is the AI being used in a manner consistent with the purpose of the overall exercise? Is a human being brought into the loop to decide whether the AI's suggestion needs adjustment before actual use? Submitting an AI-generated essay for a high-school assignment may not be ethical

Ethical Use

Models might output facts that are factually false. Sources and citations are unavailable for most models. Users should be conscious that outputs could be inaccurate and should perform due diligence to validate generated content.

Hallucination

SaaS-AI companies may save some or all of prompt payloads for future training. Therefore, confidential data will be used to train future versions of the base model – how will this affect your organization's competitiveness in the market?

IP Protection

It is critical to proactively minimize risk from malicious behavior on the network to maintain operations and customer trust. For example, a customer service bot revealing confidential information to a hacker either by prompt or unintentionally

Malicious behavior

Foundation Models are comprised of billions of parameters (model size) and trained on petabytes of data. In theory, the larger the model, the better the output. Foundation Models take time to produce outputs, which may limit real-time use cases

Model Performance

SaaS-AI companies require to submit text as a payload to users' API call. The data could be crossing borders. Is this in accordance with data privacy laws and with your company's policies? Many cloud service providers offer market-leading controls to manage data privacy of Foundation Models

Privacy

Models are good at understanding text but struggle when the data are in irregular formats, or when the position of the text on the page (e.g., infographic, PPT presentation slide) is relevant to the context and understanding. Other emphasis generators such as bolded text, font color, etc., don't play a role yet

Text Formatting

Most models have a 2k token size limit. Some larger ones can process 4k tokens in a single call. 2k tokens are approximately 2-2.5 pages. This limit makes it difficult to process larger documents

Token Size Limits

GenerativeAI Use Cases by function

SALES & MARKETING	HUMAN RESOURCES	SUPPLY CHAIN & PROCUREMENT
1. Video editing and generation	13. Personal onboarding assistant	25. Demand planning (Consumer Sentiment Analysis)
2. Metaverse 3D experience	14. Compensation analysis	26. Inventory analysis
3. Product descriptions and reviews	15. Workforce skill analysis	27. Global trade-logistics analysis
4. Personalized consumer advertisements	16. 3D avatar creation	28. Contract Adherence & Anomaly Detection
5. Recommender systems for e-commerce	17. Metaverse 3D workforce experience	29. Scenario simulation
6. Chatbot / virtual assistant dialogue generation	18. Metaverse 3D workforce upskilling	30. Language translation for global trade

GOVERNANCE & OPERATIONS	INFORMATION TECHNOLOGY	FINANCE & ACCOUNTING
7. Intranet search (knowledge management)	19. Code generation across languages/frameworks/CSPs	31. Fraud, waste, and abuse prevention
8. Process analysis	20. Development lifecycle documentation	32. Regulation and oversight analysis
9. Training for new team members	21. Test automation and test scenario creation	33. Financial report analysis
10. Document inventory analysis	22. Training on new technologies	34. Proactive value opportunity identification
11. News and media summaries	23. Peer review for optimized code writing	35. Budget and ROI analysis
12. Sentiment Analysis for Workforce	24. Legacy code summarization & translation	35. Divestment recommendations

EmergingAI Regulatory Framework

Emerging Regulatory Environment

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

- Risk Management Framework

CHAI™ Coalition for Health AI

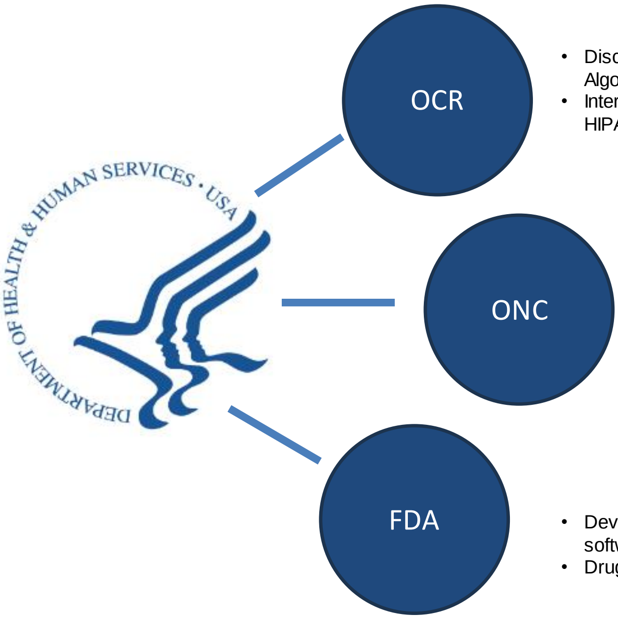
- Trustworthy AI in healthcare

BLUEPRINT FOR AN AI BILL OF RIGHTS
MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE

- 5 Principles
- Guide use, design and deployment

EU AI Act

- First AI law globally



OCR

- Discriminatory Algorithms
- Interplay with HIPAA

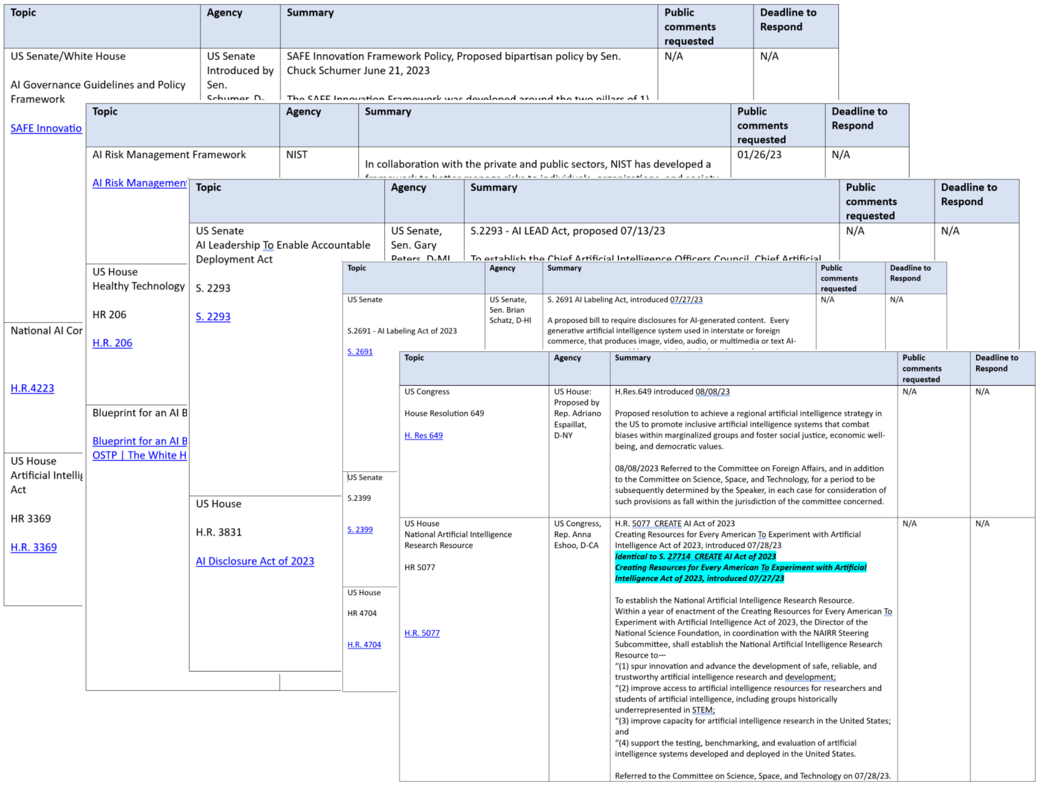
ONC

- Predictive Decision Support

FDA

- Devices- incl software
- Drugs

Federal Regulations



Topic	Agency	Summary	Public comments requested	Deadline to Respond
US Senate/White House AI Governance Guidelines and Policy Framework SAFE Innovatio	US Senate Introduced by Sen. Schumer, D-	SAFE Innovation Framework Policy, Proposed bipartisan policy by Sen. Chuck Schumer June 21, 2023 The SAFE Innovation Framework was developed around the two pillars of 1)	N/A	N/A
National AI Cor H.R.4223				
US House Artificial Intellig Act HR 3369 H.R. 3369				

Topic	Agency	Summary	Public comments requested	Deadline to Respond
AI Risk Management Framework AI Risk Managemen	NIST	In collaboration with the private and public sectors, NIST has developed a framework to better manage risks to individuals, organizations, and society	01/26/23	N/A
US House Healthy Technology HR 206 H.R. 206				
Blueprint for an AI B Blueprint for an AI B OSTP | The White H				

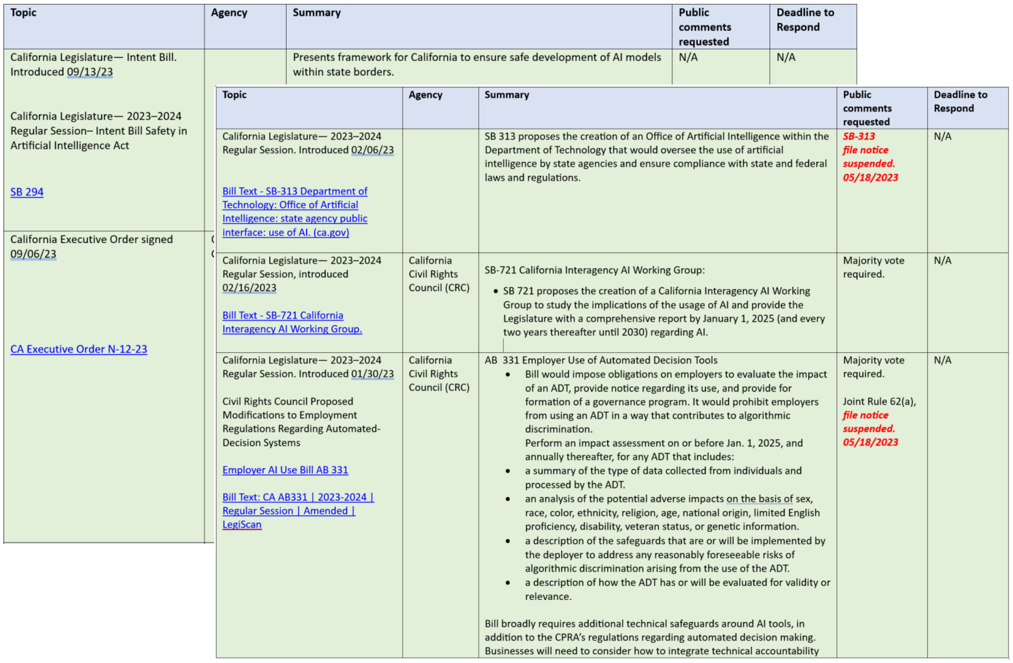
Topic	Agency	Summary	Public comments requested	Deadline to Respond
US Senate AI Leadership To Enable Accountable Deployment Act S. 2293 S. 2293	US Senate, Sen. Gary Peters, D-MI	S.2293 - AI LEAD Act, proposed 07/13/23 To establish the Chief Artificial Intelligence Officers Council, Chief Artificial	N/A	N/A
US House H.R. 3831 AI Disclosure Act of 2023				

Topic	Agency	Summary	Public comments requested	Deadline to Respond
US Senate S.2691 - AI Labeling Act of 2023 S. 2691	US Senate, Sen. Brian Schatz, D-HI	S. 2691 AI Labeling Act, introduced 07/27/23 A proposed bill to require disclosures for AI-generated content. Every generative artificial intelligence system used in interstate or foreign commerce, that produces image, video, audio, or multimedia or text AI-	N/A	N/A
US Senate S.2399 S. 2399				
US House HR 4704 H.R. 4704				

Topic	Agency	Summary	Public comments requested	Deadline to Respond
US Congress House Resolution 649 H. Res 649	US House: Proposed by Rep. Adriano Espaillat, D-NY	H.Res.649 introduced 08/08/23 Proposed resolution to achieve a regional artificial intelligence strategy in the US to promote inclusive artificial intelligence systems that combat biases within marginalized groups and foster social justice, economic well-being, and democratic values. 08/08/2023 Referred to the Committee on Foreign Affairs, and in addition to the Committee on Science, Space, and Technology, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned.	N/A	N/A
US House National Artificial Intelligence Research Resource HR 5077 H.R. 5077	US Congress, Rep. Anna Eshoo, D-CA	H.R. 5077 CREATE AI Act of 2023 Creating Resources for Every American To Experiment with Artificial Intelligence Act of 2023, introduced 07/28/23 ***Identical to S. 27714 CREATE AI Act of 2023*** ***Creating Resources for Every American To Experiment with Artificial Intelligence Act of 2023, introduced 07/27/23*** To establish the National Artificial Intelligence Research Resource. Within a year of enactment of the Creating Resources for Every American To Experiment with Artificial Intelligence Act of 2023, the Director of the National Science Foundation, in coordination with the NAIRR Steering Subcommittee, shall establish the National Artificial Intelligence Research Resource to— "(1) spur innovation and advance the development of safe, reliable, and trustworthy artificial intelligence research and development; "(2) improve access to artificial intelligence resources for researchers and students of artificial intelligence, including groups historically underrepresented in STEM; "(3) improve capacity for artificial intelligence research in the United States; and "(4) support the testing, benchmarking, and evaluation of artificial intelligence systems developed and deployed in the United States. Referred to the Committee on Science, Space, and Technology on 07/28/23.	N/A	N/A

- Frequent additions with federal regulations
- Many agencies are focused on creating frameworks to govern AI, evaluating and reducing risk
- Topics such as trust, security, privacy trend throughout proposed regs
- Some outliers of proposed regs are concerning, H.R. 206

California Regulations

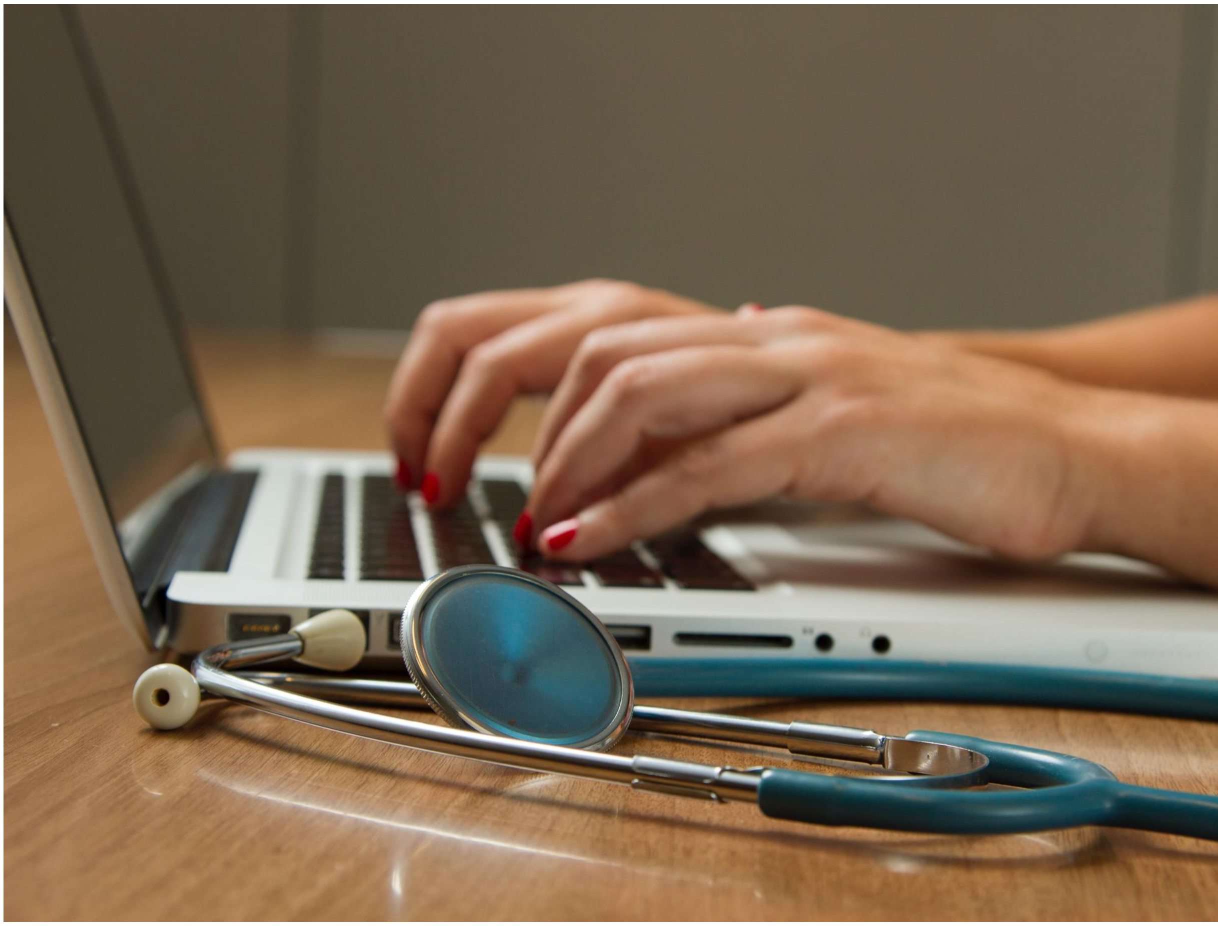


Topic	Agency	Summary	Public comments requested	Deadline to Respond
California Legislature— Intent Bill. Introduced 09/13/23 California Legislature— 2023–2024 Regular Session– Intent Bill Safety in Artificial Intelligence Act SB 294		Presents framework for California to ensure safe development of AI models within state borders.	N/A	N/A
California Executive Order signed 09/06/23 CA Executive Order N-12-23				

Topic	Agency	Summary	Public comments requested	Deadline to Respond
California Legislature— 2023–2024 Regular Session. Introduced 02/06/23 Bill Text - SB-313 Department of Technology: Office of Artificial Intelligence: state agency public interface: use of AI. (ca.gov)		SB 313 proposes the creation of an Office of Artificial Intelligence within the Department of Technology that would oversee the use of artificial intelligence by state agencies and ensure compliance with state and federal laws and regulations.	***SB-313 file notice suspended. 05/18/2023***	N/A
California Legislature— 2023–2024 Regular Session, introduced 02/16/2023 Bill Text - SB-721 California Interagency AI Working Group.	California Civil Rights Council (CRC)	SB-721 California Interagency AI Working Group: • SB 721 proposes the creation of a California Interagency AI Working Group to study the implications of the usage of AI and provide the Legislature with a comprehensive report by January 1, 2025 (and every two years thereafter until 2030) regarding AI.	Majority vote required.	N/A
California Legislature— 2023–2024 Regular Session. Introduced 01/30/23 Civil Rights Council Proposed Modifications to Employment Regulations Regarding Automated-Decision Systems Employer AI Use Bill AB 331 Bill Text: CA AB331 | 2023-2024 | Regular Session | Amended | LegiScan	California Civil Rights Council (CRC)	AB 331 Employer Use of Automated Decision Tools • Bill would impose obligations on employers to evaluate the impact of an ADT, provide notice regarding its use, and provide for formation of a governance program. It would prohibit employers from using an ADT in a way that contributes to algorithmic discrimination. Perform an impact assessment on or before Jan. 1, 2025, and annually thereafter, for any ADT that includes: • a summary of the type of data collected from individuals and processed by the ADT. • an analysis of the potential adverse impacts on the basis of sex, race, color, ethnicity, religion, age, national origin, limited English proficiency, disability, veteran status, or genetic information. • a description of the safeguards that are or will be implemented by the deployer to address any reasonably foreseeable risks of algorithmic discrimination arising from the use of the ADT. • a description of how the ADT has or will be evaluated for validity or relevance. Bill broadly requires additional technical safeguards around AI tools, in addition to the CPRA's regulations regarding automated decision making. Businesses will need to consider how to integrate technical accountability	Majority vote required. Joint Rule 62(a), ***file notice suspended. 05/18/2023***	N/A

- California – home to 36/51 major AI vendors
- State laws can conflict with federal (historical, and with AI framework)
- Some state laws well intentioned but concerning if poorly implemented (SB 294; establishes liability and penalties to damages caused by "foreseeable risk"

AI in Healthcare



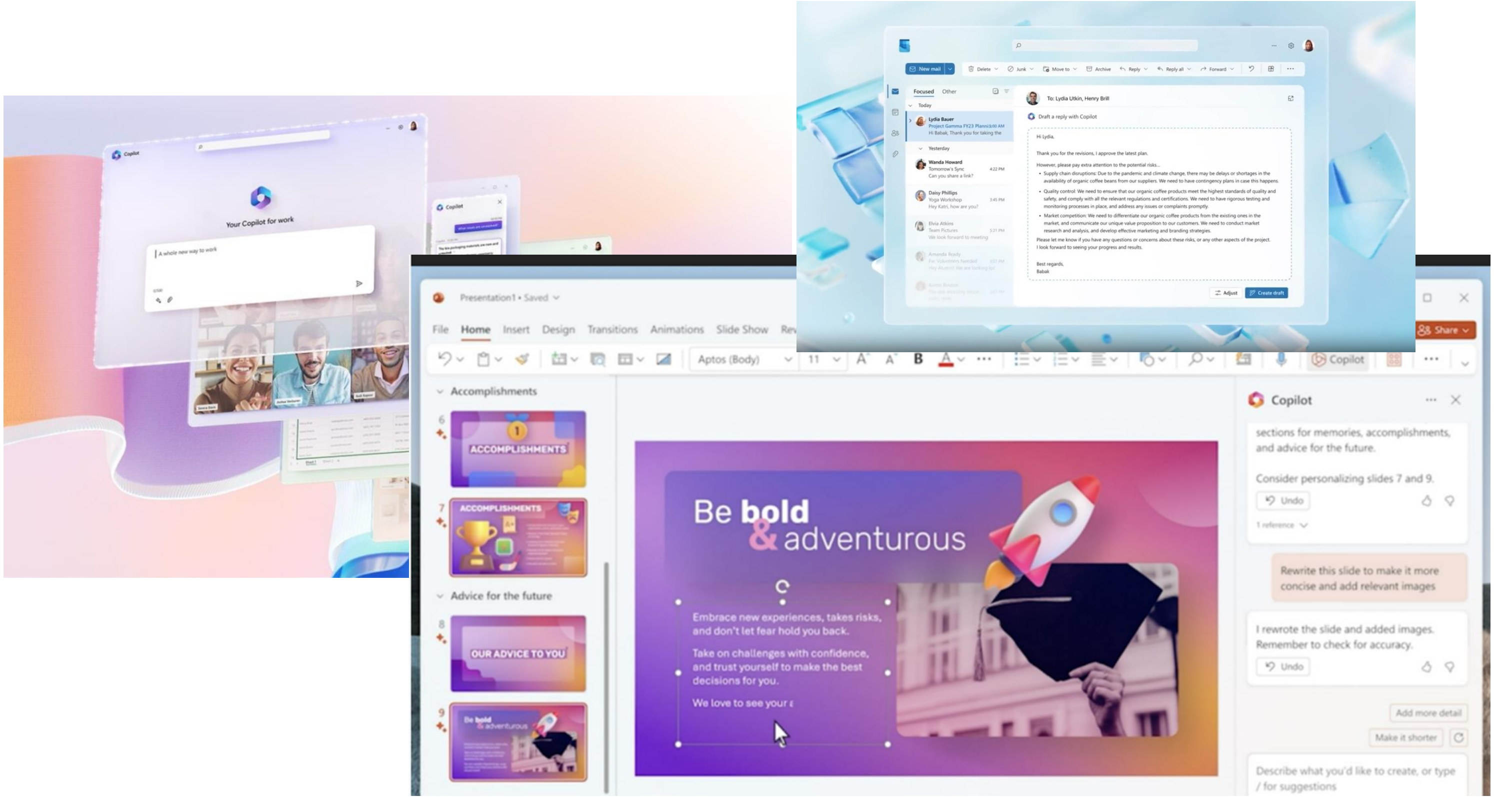
Numerous AI use cases across the industry

Current use cases include AI to improve patient engagement and health outcomes

	AI Use Cases				
LIFE SCIENCES	**Robotic-Assisted Therapy and Surgery** *Use robotics and AI to assist patients in their recovery, leveraging digital algorithms to detect motions that patients can't execute during therapy and guide them through execution*	**Drug Discovery** *Create machine learning powered models to process massive datasets to identify and validate targets, design molecules, and test in silico*	**Digital Data Flow for Clinical Trials** *Utilize cognitive automation to integrate trial data from multiple systems, create standardized digital elements, and generate trial artifacts*	**Precision Medicine** *Use AI to analyze genomic and phenotypic data to design individualized pharmacological treatments*	**Drug Manufacturing Intelligence** *Leverage sensor data and algorithmic models to predict manufacturing deviations and maximize factory yield and productivity*
	Self Healing Supply Chain *Apply AI to automate the analysis and aggregation of data to forecast supply and demand, and recommend the next best action to supply chain operators and autonomously perform activities*	**Drug Marketing Omnichannel Engagement** *Develop ML models based on promotional and longitudinal data to predict how, when, and with what message type to best engage with patients & HCPs, as well as optimize marketing spend across media mix channels*	**Voice of the Patient Insight** *Analyze patient social media feedback, complaints, and adverse events to identify insights that can improve product design, packaging, and educational materials*	**Proactive Risk and Compliance** *Apply AI to automate the analysis and aggregation of data to identify risk and compliance items, and recommend the next best action along with mitigation techniques*	**IT Quality / Data Integrity** *Apply AI to automate the analysis and aggregation of data to identify quality and data integrity issues, and recommend corrective and preventative actions to address gaps/control failures/non-conformances*
HEALTH CARE	**Patient Engagement** *Provide patient access to medical records, efficient appointment scheduling, and direct communication with staff and care coordination teams*	**Real-Time Monitoring of Regulations/Policy Impacting LS/HC Industry** *Use machine learning to assess real-time policy changes, and associated impact to HC providers*	**Precision Medicine/Personalized Health** *Leverage predictive insights to diagnose, prevent and treat a future illness based on an individual's lifestyle, real-world environment, biometric data and genomics*	**Hospital Management** *Forecast based on predictive insights, surges/lows in patient volumes, to support hospitals in staffing appropriately*	**Computer Assisted Diagnosis** *Leverage deep neural nets, machine learning and categorization technology to obtain a more efficient/accurate evaluation of imaging studies*
	Clinical Decision Support *Use complex clinical algorithms to aid and drive clinical decisions that will streamline, improve, and standardize medical practices*	**Care Claim Revenue Cycle Optimization and Efficiency** *Automate pre-care, day of care, and post-care claims submission and payment activities*	**Virtual Personal Health Assistants** *Use augmented reality, cognitive computing, sentiment analysis, and speech and body recognition to create a virtual encounter between a personal health assistant and patient*	**Provider Supply Chain Management** *Establish cognitive profiles of physicians and supplies utilized, detect anomalies in behavior patterns and sourcing, understand root causes, and provide recommendations for investigation and resolution*	**Provider Payment Integrity** *Detect and prevent fraud, waste, and abuse before it occurs by leveraging AI to increase access to information and uncover insights when analyzing claims, payments, and behavior trends*

Cost Reduction · *Speed to Execution* · *Reduced Complexity* · *Transformed Engagement* · *Fueled Innovation* · *Fortified Trust* · *Apt for AI Supercomputing* · *Use case where Deloitte is engaged*

It Is Not Just Clinical - Microsoft 365



To: Lydia Utkin, Henry Brill

Draft a reply with Copilot

Hi Lydia,

Thank you for the revisions, I approve the latest plan.

However, please pay extra attention to the potential risks...

- Supply chain disruptions: Due to the pandemic and climate change, there may be delays or shortages in the availability of organic coffee beans from our suppliers. We need to have contingency plans in case this happens.
- Quality control: We need to ensure that our organic coffee products meet the highest standards of quality and safety, and comply with all the relevant regulations and certifications. We need to have rigorous testing and monitoring processes in place, and address any issues or complaints promptly.
- Market competition: We need to differentiate our organic coffee products from the existing ones in the market, and communicate our unique value proposition to our customers. We need to conduct market research and analysis, and develop effective marketing and branding strategies.

Please let me know if you have any questions or concerns about these risks, or any other aspects of the project.
I look forward to seeing your progress and results.

Best regards,
Babak

Be bold & adventurous

Embrace new experiences, takes risks, and don't let fear hold you back.

Take on challenges with confidence, and trust yourself to make the best decisions for you.

We love to see your

sections for memories, accomplishments, and advice for the future.

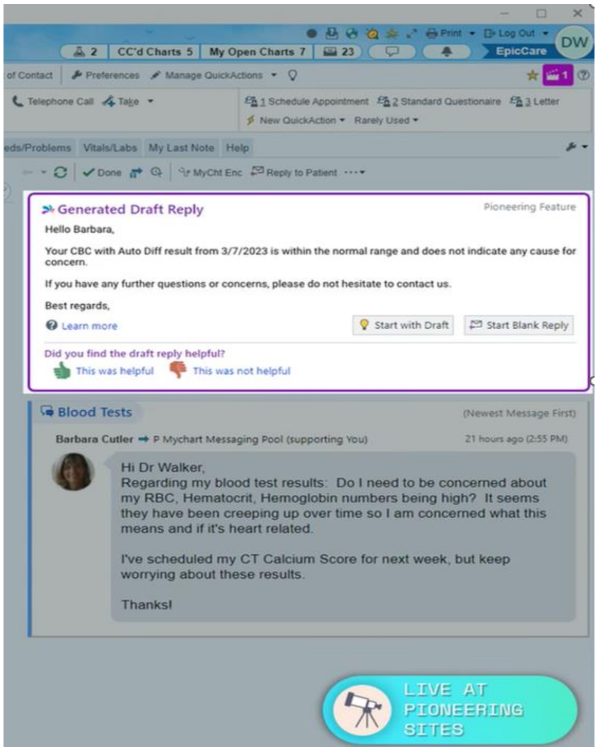
Consider personalizing slides 7 and 9.

Rewrite this slide to make it more concise and add relevant images

I rewrote the slide and added images. Remember to check for accuracy.

Describe what you'd like to create, or type / for suggestions

EMR Vendors – Epic Generative AI



Generated Draft Reply — Pioneering Feature

Hello Barbara,

Your CBC with Auto Diff result from 3/7/2023 is within the normal range and does not indicate any cause for concern.

If you have any further questions or concerns, please do not hesitate to contact us.

Best regards,

Did you find the draft reply helpful?

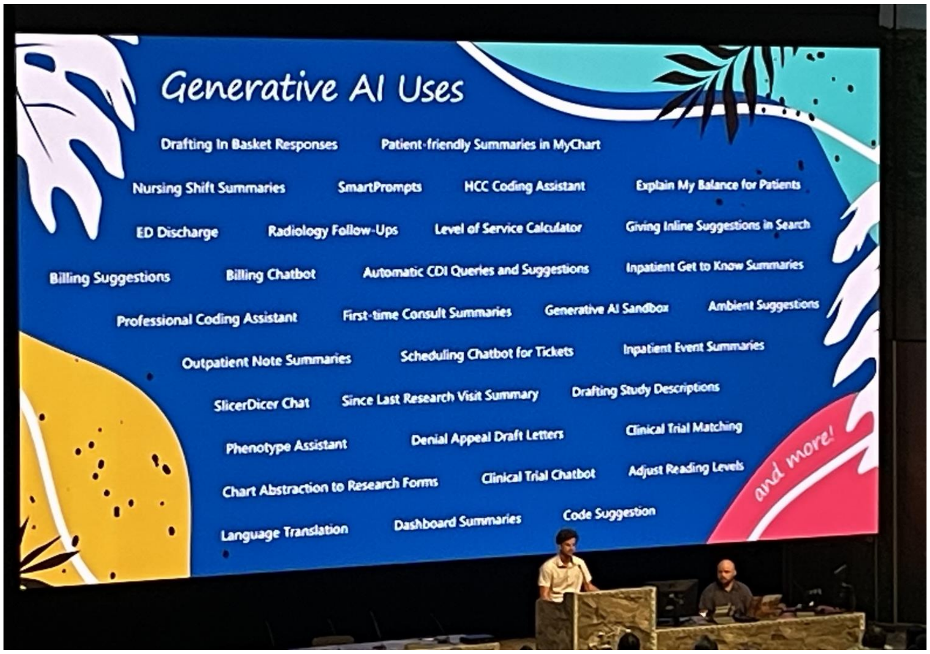
Blood Tests (Newest Message First)

Barbara Cutler ➡ P Mychart Messaging Pool (supporting You) — 21 hours ago (2:55 PM)

Hi Dr Walker,
Regarding my blood test results: Do I need to be concerned about my RBC, Hematocrit, Hemoglobin numbers being high? It seems they have been creeping up over time so I am concerned what this means and if it's heart related.

I've scheduled my CT Calcium Score for next week, but keep worrying about these results.

Thanks!

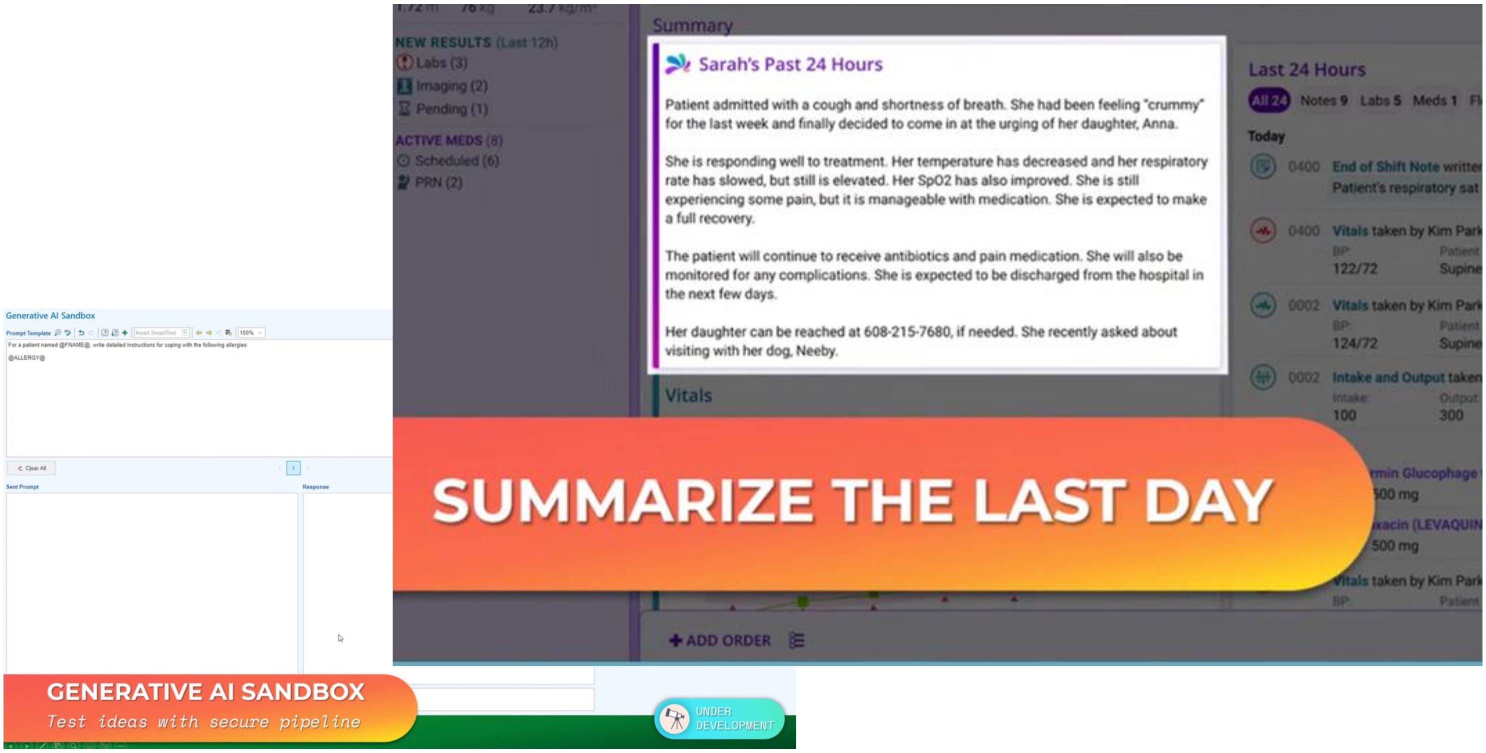


Generative AI Uses

Drafting In Basket Responses
Patient-friendly Summaries in MyChart
Nursing Shift Summaries
SmartPrompts
HCC Coding Assistant
Explain My Balance for Patients
ED Discharge
Radiology Follow-Ups
Level of Service Calculator
Giving Inline Suggestions in Search
Billing Suggestions
Billing Chatbot
Automatic CDI Queries and Suggestions
Inpatient Get to Know Summaries
Professional Coding Assistant
First-time Consult Summaries
Generative AI Sandbox
Ambient Suggestions
Outpatient Note Summaries
Scheduling Chatbot for Tickets
Inpatient Event Summaries
SlicerDicer Chat
Since Last Research Visit Summary
Drafting Study Descriptions
Phenotype Assistant
Denial Appeal Draft Letters
Clinical Trial Matching
Chart Abstraction to Research Forms
Clinical Trial Chatbot
Adjust Reading Levels
Language Translation
Dashboard Summaries
Code Suggestion

and more!

EMR Vendors – Epic Generative AI



Generative AI Sandbox

For a patient named @FNAME@, write detailed instructions for coping with the following allergies:

@ALLERGY@

GENERATIVE AI SANDBOX
Test ideas with secure pipeline

NEW RESULTS (Last 12h)
Labs (3)
Imaging (2)
Pending (1)
ACTIVE MEDS (8)
Scheduled (6)
PRN (2)

Summary

Sarah's Past 24 Hours

Patient admitted with a cough and shortness of breath. She had been feeling "crummy" for the last week and finally decided to come in at the urging of her daughter, Anna.

She is responding well to treatment. Her temperature has decreased and her respiratory rate has slowed, but still is elevated. Her SpO2 has also improved. She is still experiencing some pain, but it is manageable with medication. She is expected to make a full recovery.

The patient will continue to receive antibiotics and pain medication. She will also be monitored for any complications. She is expected to be discharged from the hospital in the next few days.

Her daughter can be reached at 608-215-7680, if needed. She recently asked about visiting with her dog, Neeby.

Vitals

Last 24 Hours

Today

0400 End of Shift Note writter
Patient's respiratory sat

0400 Vitals taken by Kim Park
BP: 122/72 Patient Supine

0002 Vitals taken by Kim Park
BP: 124/72 Patient Supine

0002 Intake and Output taken
Intake: 100 Output: 300

rmin Glucophage
500 mg

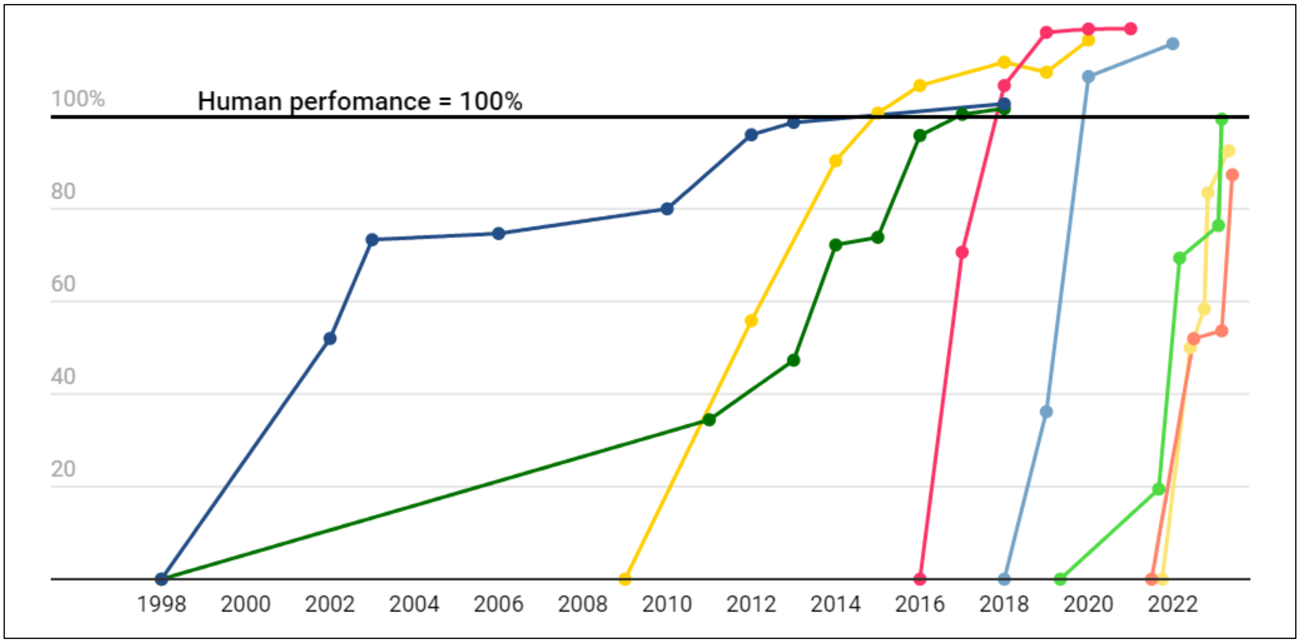
xacin (LEVAQUIN
500 mg

Vitals taken by Kim Park
BP: Patient

SUMMARIZE THE LAST DAY

Health System AI

WHY AI, Why Now?



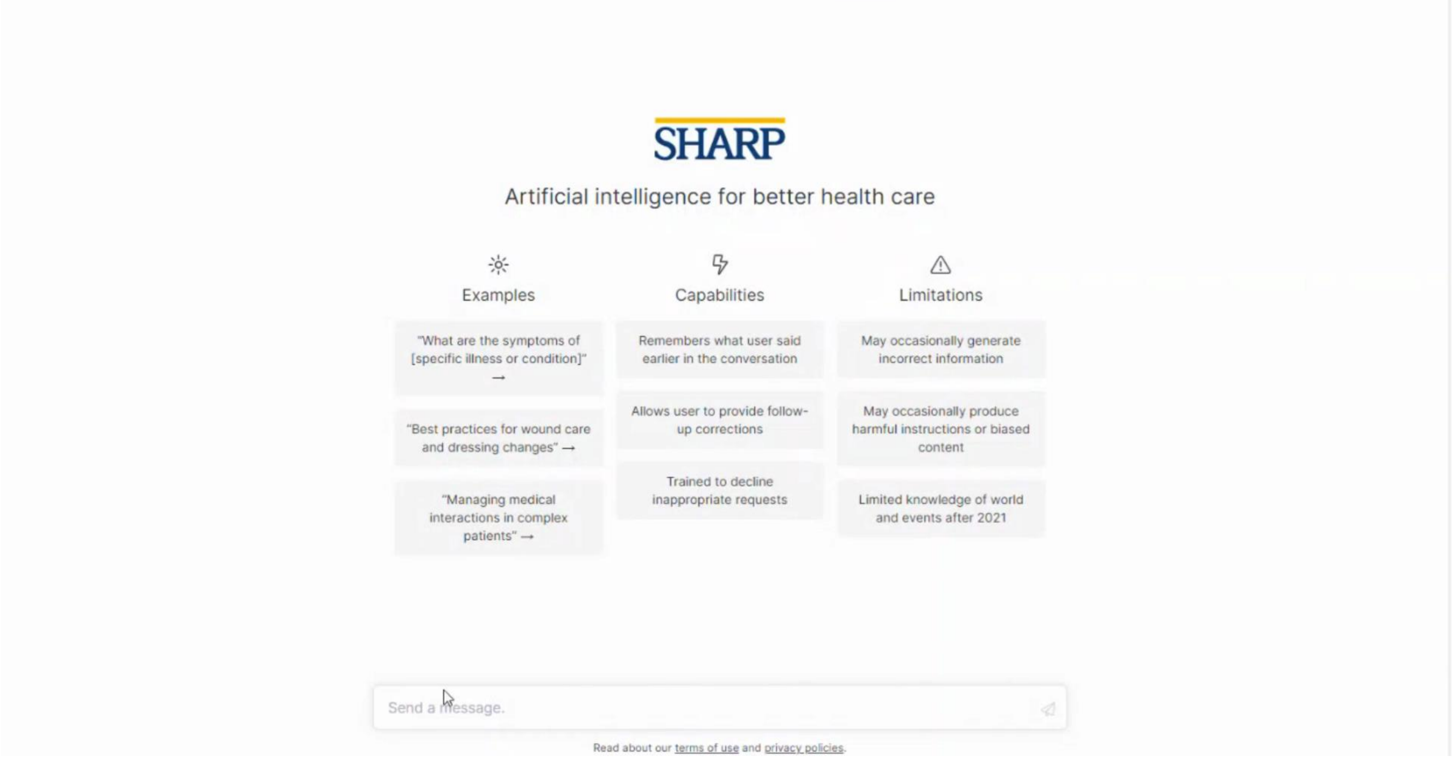
State-of-the-art AI performance on benchmarks, relative to human performance

● Handwriting recognition ● Speech recognition ● Image recognition ● Reading comprehension

● Language understanding ● Common sense completion ● Grade school math ● Code generation

Source: Why AI Progress Is Unlikely to Slow Down | Time

Meet Sharp GPT



Artificial intelligence for better health care

Examples	Capabilities	Limitations
"What are the symptoms of [specific illness or condition]" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Best practices for wound care and dressing changes" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"Managing medical interactions in complex patients" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

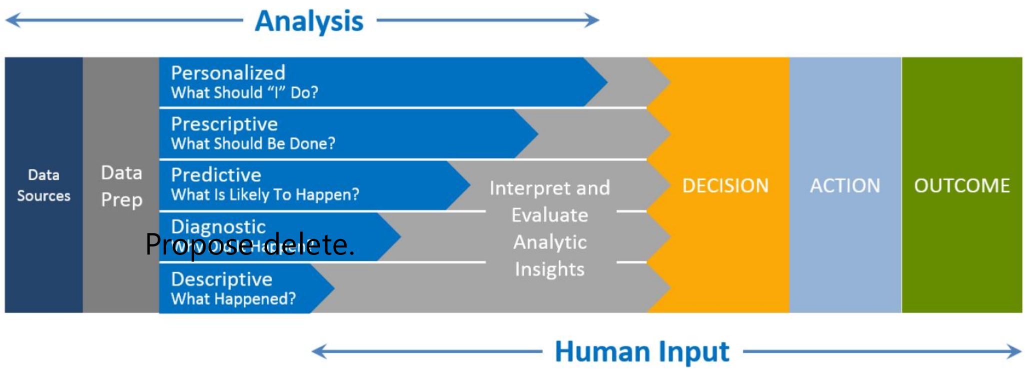
Read about our terms of use and privacy policies.

But with Great Technology…. Comes a Need for Great Oversight

A Statement of Purpose – AI Oversight Committee

This Committee will provide oversight of where and how artificial intelligence and data science assets are used within Sharp HealthCare. The Committee is responsible for leveraging individual subject matter expertise to try to anticipate and mitigate unintended consequences of AI. The committee will provide guidance and expertise to develop and implement standards, policies and process around this rapidly evolving discipline of artificial intelligence.

Human AND Machine Together



Propose delete.

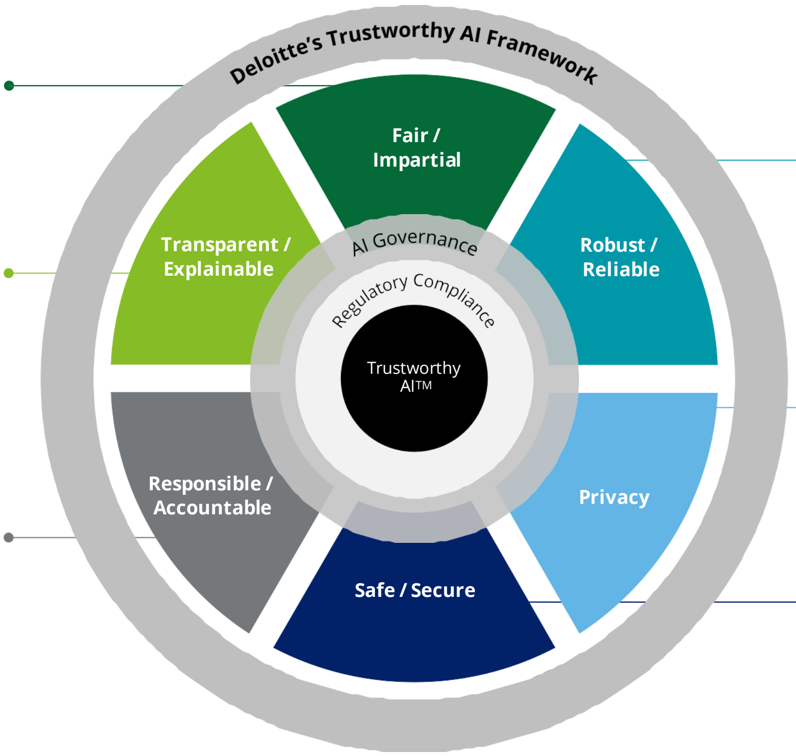
Combining computer and human analysis to make better decisions and achieve better outcomes.

A Human

In the Loop … Always

AI Risk Domains

Deloitte's Trustworthy AI framework



Fair / Impartial
AI applications include internal and external checks to help enable equitable application across all participants

Transparent / Explainable
All participants are able to understand how their data is being used and how AI systems make decisions; algorithms, attributes, and correlations are open to inspection

Responsible / Accountable
Policies are in place to determine who is held responsible for the output of AI system decisions

Robust / Reliable
AI systems have the ability to learn from humans and other systems and produce consistent and reliable outputs

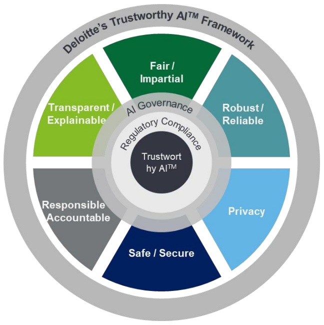
Privacy
Consumer privacy is respected and customer data is not used beyond its intended and stated use; consumers are able to opt in/ out of sharing their data

Safe / Secure
AI systems can be protected from risks (including cyber risks) that may cause physical and/or digital harm

Deloitte's Trustworthy AI™ Framework

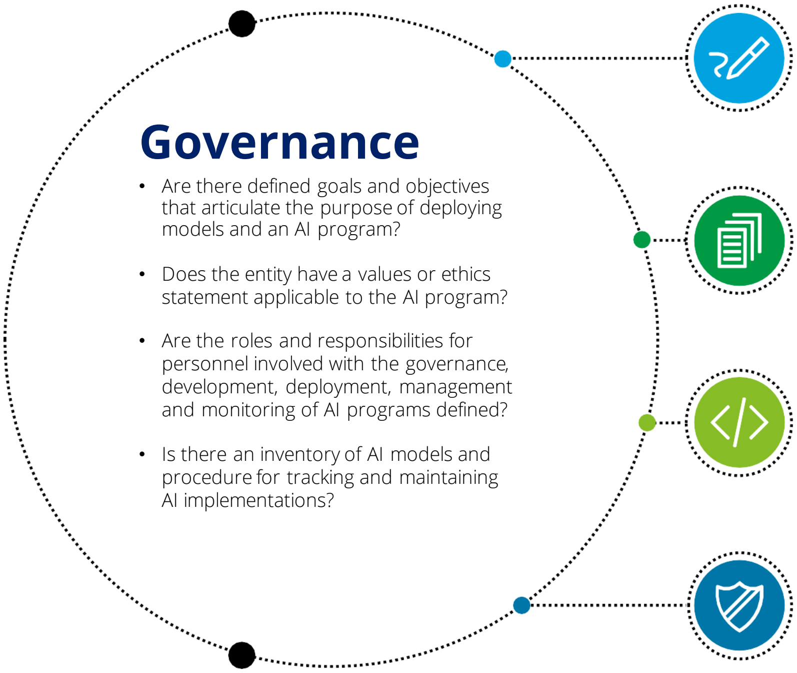
Below are example areas that our team has identified for Digital Program Assurance and Trustworthy AI™ assessments.

Ideation & Planning		Data Prep & Model Building			Deployment		Model Management	
A AI Strategy	A Roles and Responsibilities	F Fairness/Biasness Testing	F Data Completeness	F Feature Engineering	P Data Protection	P Third-Party Access	A Role and Responsibilities	A Change Management
R Effective Governance and Oversight	R Assess Risks of AI	F Hyper-parameter Selection	T Model Methodology	T Feature Design	S Open-Source Libraries	S Cyber Risk	A Enterprise-wide Controls	A Monitoring frequency
		T Hyper-parameters testing	T Vendor Solutions	T Documentation	S Storage/ Backup	T Intermediate Model Outputs	A Manging Model Inventory	R Monitor Data Applicability
		R Data Quality	R Conceptual Soundness and Model Testing	R Training/ Fitting, Model Stability	T Inter-dependencies of Components/ Technology		R Monitor model weakness	R Monitor model performance
		P Sensitive Data					R Thresholds/ escalation process	T Documentation



AI Governance and Best Practices

Internal Audit Considerations throughout the AI lifecycle



Governance

- Are there defined goals and objectives that articulate the purpose of deploying models and an AI program?
- Does the entity have a values or ethics statement applicable to the AI program?
- Are the roles and responsibilities for personnel involved with the governance, development, deployment, management and monitoring of AI programs defined?
- Is there an inventory of AI models and procedure for tracking and maintaining AI implementations?

Design Process

- Has a business case evaluation been performed (i.e., a broader group weighed in on the need, and pros and cons, for the AI application)?
- Have potential sources of risks (e.g., bias) been identified and addressed/mitigated?
- Is there a process or procedure in place that monitors current and emerging regulations and their applicability to AI implementations?

Data

- Is there a process for how data streams are selected and evaluated?
- Is bias and potential bias with AI implementations, reviewed, evaluated, and documented?
- Are data sets validated to ensure they are representative of the underlying populations and operational environment

Development & Implementation

- Is the complexity of the model commensurate with the use case and benefits of the model?
- Are assumptions and limitations for AI models evaluated?
- Are models testing to ensure they are consistent with goals and objectives set forth in the business case as well as principles to foster public trust

Validation & Review

- Are the test cases comprehensive, with appropriate pass/fail criteria, and is there appropriate statistical or other quantitative/qualitative testing of the modeled results performed by relevant stakeholders with appropriate expertise?
- Is the applicability and relevance of model policy, procedures and standards reviewed on a regular basis to ensure they are up to date and reflect evolving regulations and corporate requirements?

How Can Internal Audit Support AI?

Ways that Internal Audit can engage with operations in this space

AI strategy, governance and operating model

Provide a strategic cross-functional governance, roadmap and operating model for an effective AI risk management program.

- *Design and implement an AI strategy and framework to be supported by processes and controls over governance, deployment, and monitoring of AI based on our Trustworthy AI framework*
- *Design and implement a strategy for compliance with AI regulations*

Establishing an AI risk program and operational constructs in alignment with your business strategy and operations

AI data governance

Establish a data governance and risk management framework to safeguard for the security, privacy, integrity and ethics of data used for AI throughout its lifecycle.

- *Provide recommendations for remediation of data and controls over AI based on Trustworthy AI framework*
- *Assist stakeholders in the development, design and implementation of controls to address AI-specific risks*

Implementing a data-centric risk approach and framework for managing risks throughout the AI lifecycle

AI risk management operations

Build trust and resiliency in AI systems against anomalous activities that could compromise the data, models or outcomes. Develop robust and resilient infrastructure, operations and model development processes.

- *Provide independent testing on the design and operating effectiveness of AI controls, findings, and recommendations for deficiencies in the AI environment*
- *Conduct independent testing of AI models and related datasets for potential adverse outcomes*

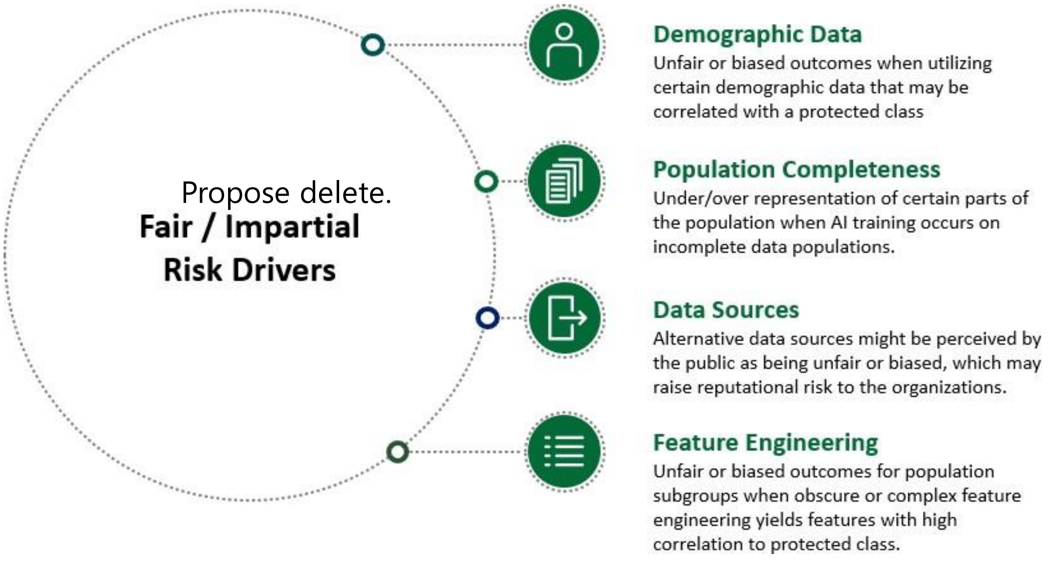
Analyzing and improving AI technology and related processes, to promote organizational trust for your AI solutions

Questions?

APPENDIX

Fair / Impartial

The risk of producing discriminatory bias, or the perception thereof, towards certain subgroups of the populations and thus against the organization’s ethical value.



Propose delete.

Demographic Data
Unfair or biased outcomes when utilizing certain demographic data that may be correlated with a protected class

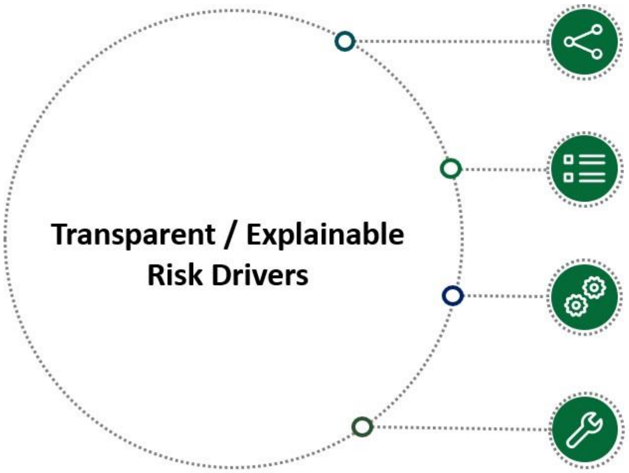
Population Completeness
Under/over representation of certain parts of the population when AI training occurs on incomplete data populations.

Data Sources
Alternative data sources might be perceived by the public as being unfair or biased, which may raise reputational risk to the organizations.

Feature Engineering
Unfair or biased outcomes for population subgroups when obscure or complex feature engineering yields features with high correlation to protected class.

Transparent / Explainable

The lack of ability to explain a particular behavior of the AI system due to model complexity or feature inexplicability.



Transparent / Explainable Risk Drivers

Model Methodology

Model complexity and architecture may make model behavior difficult to understand or explain (e.g. multi-layered neural nets).

Feature Design

Complex feature transformation can result in obfuscation of original data attributes yielding difficult to explain outcomes.

Hyperparameters

Difficulty explaining the impact of hyperparameter choices. Hyperparameters are typically derived based on a trial & error or optimization process.

Infrastructure

Potentially difficulty understanding behavior of complex or heavily inter-connected models.

Responsible / Accountable

Lack of responsibility or accountability at each stage of AI use can lead to increased risk, including regulatory and operational risks.

Responsible / Accountable Risk Drivers

AI Strategy

- Inability to apply AI across an organization consistently and within the ethical bounds of the organizational values.
- AI usage outside of the risk appetite of the organization.
- AI solutions inconsistent with overarching strategy.

Roles and Responsibilities

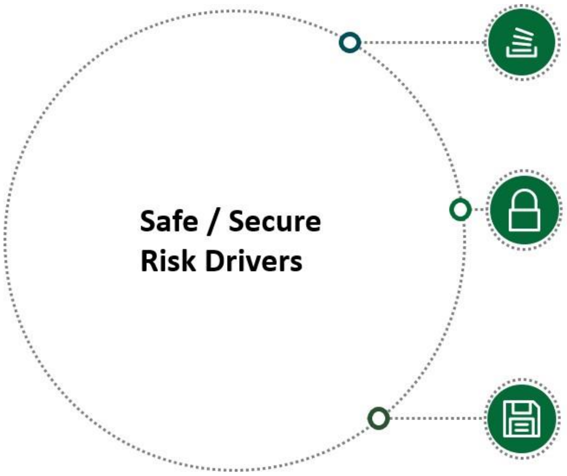
- Poor AI/data/model oversight driven from ambiguous and overlapping roles and responsibilities in the organization.
- Undesired AI outcomes can arise from poorly defined roles and responsibilities and lack of accountability across the 1st, 2nd and 3rd lines of defense
- Undesired outcomes arise from inadequate review.

Change Management

- Uncontrolled changes to models may lead to technology failures or poor model outcomes.

Safe / Secure

The internal and external threats that arise due to lack of consistent and cohesive security of AI systems can lead to multiple risks including loss of business, customers' and regulators' trust.



Open-Source Libraries

- Usage open-source packages or beta versions may pose inherent security risks.
- Open sources packages may transmit data to third party servers (e.g.: Plottly or Dash can transmit data to plot graphs on cloud)

Cyber Risks

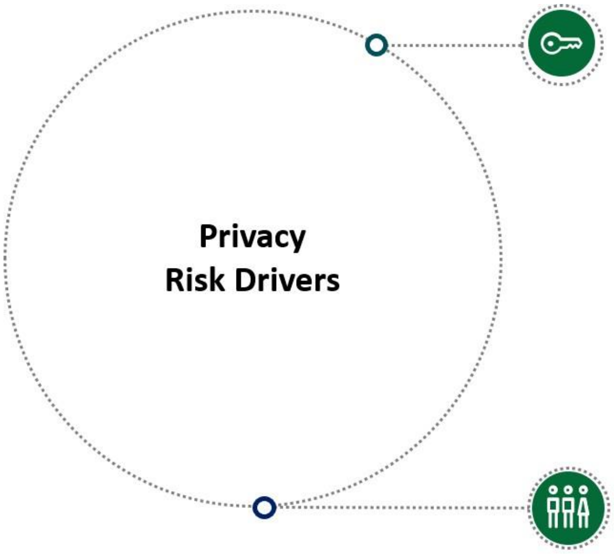
- Prototype code on unofficial infrastructure may be vulnerable to cyber attacks resulting in loss of IP.
- Automated systems may expose upstream/downstream AI/IT systems to security threats.
- Fragmented learning systems may be exploited to spread viral/malware infections throughout the AI network.

Storage/Backup

- Large-scale AI systems can take up to weeks of training. Any disruption during the training can lead to loss to critical training and compute time.
- Valuable training data collated from years of business experience may accidentally get deleted/lost.
- Lack of formalized storage infrastructure to maintain data and trained models (i.e. desktop storage)

Privacy

The risk in partial or complete failure to maintain privacy of data used or created within/by the AI systems or non-compliance to internal/external data protection rules.



Data Protection

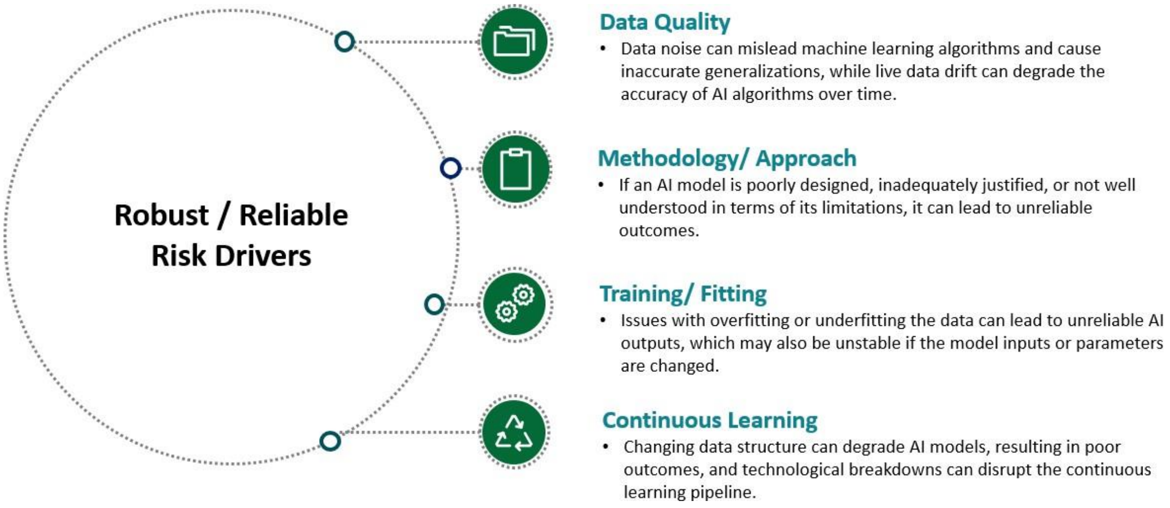
- Non-compliance to local/international data privacy rules (e.g. : GDPR) can lead to:
    - Prohibition of model usage leading to business risks.
    - Regulatory Penalties
    - Market Cap/ Reputation loss.
- Unauthorized usage of data in AI systems may lead to legal risks (e.g., usage of data beyond permissible/consented time period/applications may be perceived as unauthorized)
- Unapproved sensitive data used in ML pipelines.
- Lack of data privacy governance in AI strategy may lead to potential sensitive data risks
- Potential privacy risks when working with unstructured data as documents may include various elements of sensitive data.

Third Party Access

- Un-intended usage of data by third party vendors can lead to financial, regulatory and reputational risks.
- Lack of virtual boundaries in cloud infrastructure may leak sensitive data from one AI application into another.

Robust / Reliable

The risk that AI systems are not functioning as designed, resulting in unintended consequences over time or on new data sets.



Robust / Reliable Risk Drivers

Data Quality

- Data noise can mislead machine learning algorithms and cause inaccurate generalizations, while live data drift can degrade the accuracy of AI algorithms over time.

Methodology/ Approach

- If an AI model is poorly designed, inadequately justified, or not well understood in terms of its limitations, it can lead to unreliable outcomes.

Training/ Fitting

- Issues with overfitting or underfitting the data can lead to unreliable AI outputs, which may also be unstable if the model inputs or parameters are changed.

Continuous Learning

- Changing data structure can degrade AI models, resulting in poor outcomes, and technological breakdowns can disrupt the continuous learning pipeline.