PRESENTED BY NC DOA Department of Administration Purchase & Contract

Navigating AI in IT Procurement

2025 PEAK BREAKOUT SESSION

ELEVATING NC PROCUREMENT TOGETHER

Navigating AI in IT Procurement

Lauren Clemmons - Special Deputy Attorney General
NC Department of Justice

Sarah Robey- Privacy Legal Advisor and Analyst
NC Department of Information Technology

Keith Briggs- State Chief Architect & Innovation Director
NC Department of Information Technology

What is Generative Artificial Intelligence?

Generative AI is a specific application of **machine learning** and is a subset of artificial intelligence (AI).

GenAI uses machine learning models to **generate new content**.

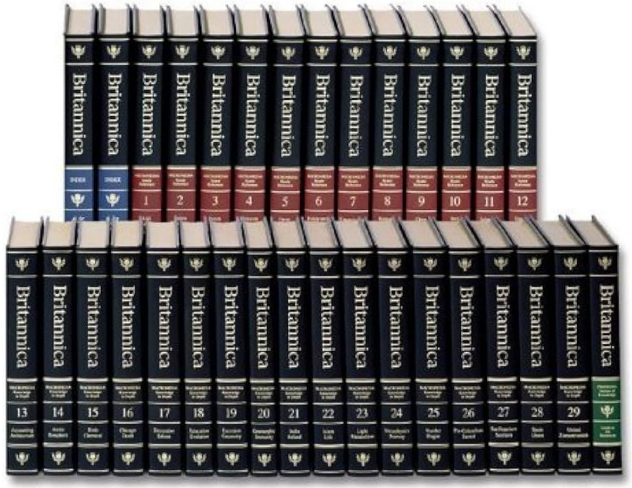
GenAI does this by using **machine learning models trained on large data sets**, enabling it to produce text, images, or music that is similar to the training data.

Machine learning is a subset of artificial intelligence that involves training algorithms to learn from and to make predictions or decisions based on data. It relies on mathematical models to identify patterns within the data, allowing the system to improve its performance over time without explicit programming for each task.

What are examples of GenAI Technology?

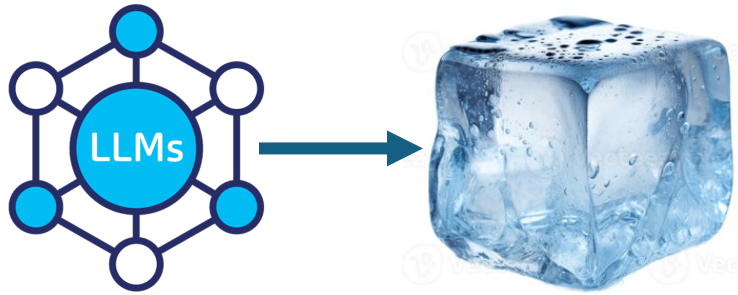
- **Style GAN** by NVIDIA (creates high-quality realistic images)
- **DALL-E** by OpenAI (generates images from text descriptions)
- **OpenAI's GPT-2** (text generation and translation)
- **OpenAI's GPT-3** (text generation, summarization and translation)
- **Microsoft CoPilot** (text generation, code generation; conversational interfaces (i.e., chatbot))
- **StableDiffusion** (open-source AI tool)(generates high-quality images from text prompts)

How is Data Used in AI models?



2010 edition of Encyclopedia Britannica contained around **50 million words** and **300 million characters** across 32 volumes, which is about **one gigabyte (GB)** of data

Large Language Model (LLM)
ChatGPT 3.5 – 176 Billion Parameters
ChatGPT 4 – 1.76 Trillion Parameters, one petabyte (PB) of data (1M encyclopedias)
ChatGPT 5 - ???



The "P" in GPT = Pre-trained
Still need to be aware of "**next version**"

Common Understanding:
Average MP4 Music File: **4** Minutes, **3** megabytes
A one petabyte MP4 would play continuously for **2,000 years**

Data Source Matters in AI models

Does Knowing Data Source matter in IT procurement?

Yes. ☺

The data source directly impacts the integrity, reliability, and security of the data being used.

There are six primary risk issues to consider with respect to the data source.

Knowing the data source lets the purchasing agency assess whether the GenAI solution is appropriate for the agency's "Use Case."

What is a “Use Case”?

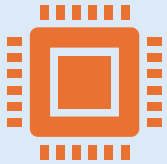
A “use case” refers to the specific situation in which a product or service may be used to achieve a particular goal or to solve a problem.

TRANSLATION: How is this AI thing going to be used?

AI Use Case is Important

Does knowing the Use Case matter in IT Procurement?

Yes. ☺



The Use Case provides the **Facts** that the IT Procurement Professionals will need for the solicitation document and that the IT security, enterprise architecture, privacy, and legal professionals will need to process the solicitation.

Possible Use Case Questions

What is the task to be accomplished?

Who needs to accomplish the task?

What steps need to be taken to achieve the goal?

Use Case Questions Continued

How does the agency intend to use the AI application?

What government business function or operation will the AI serve?

Is the AI service a business-to-business application that the purchasing agency will use internally?

Use Case Questions Continued

Is the use of the AI product intended to be consumer, or citizen focused?

Will the output of the AI feature be shared with consumers or citizens?

Will the AI product/service make or affect decisions impacting individuals that are subject to specific laws?

Risk issues Associated with Data Sources

1. Quality and Reliability

2. Compliance with Laws and Regulations

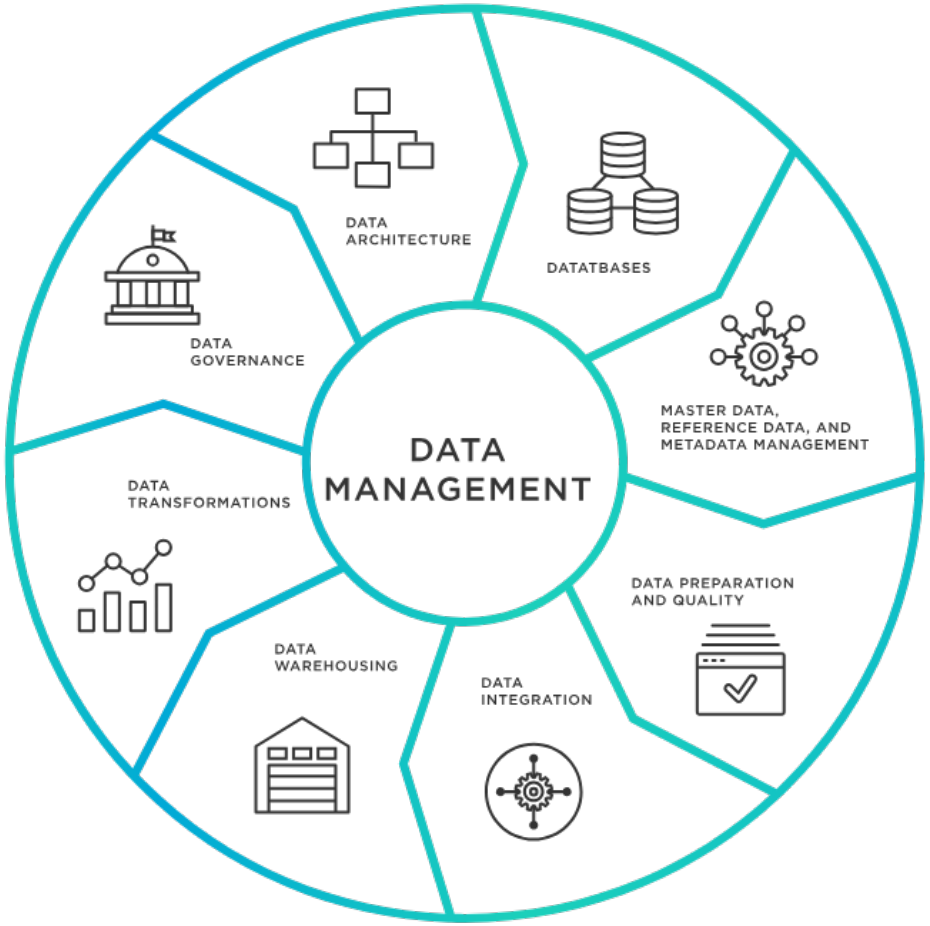
3. Bias and Fairness

4. Transparency and Accountability

5. Security and Integrity

6. Ethical considerations

Mitigating Risks



Bad Data In=Bad Data Out

1. Identify the source of the training data.

2. Consider the use case.

3. Identify the potential risk issue for the data (see slide 12), then mitigate.

AI Use case for Medical diagnosis (example)

Training data=Medical records

Risk issue=Quality and reliability of the data

Mitigate= Assess the accuracy and consistency of the data

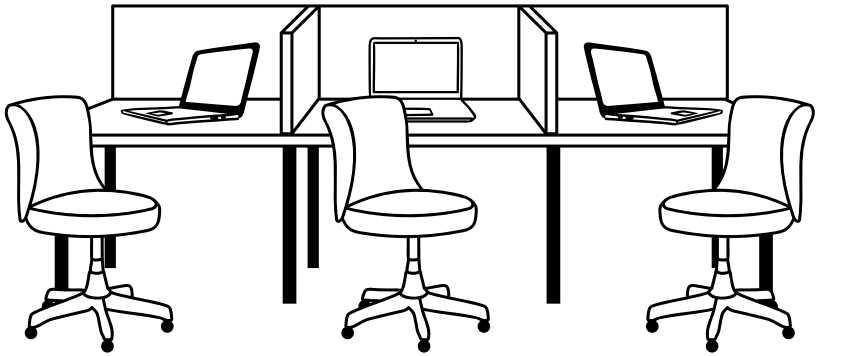
Outcome= Using high quality *verified* medical records to train an AI model for diagnosing diseases ensures accurate predictions and reliable patient care.

Example: AI Use for Hiring Decisions

Training data= Applicant data

Risk issue=Bias and Fairness

Mitigate= Identify potential biases in the data; is the data representative or are certain groups underrepresented?



Outcome= Using diverse applicant data for training the AI hiring tool helps mitigate biases and promotes fair hiring decisions.

User Input and User Output

User Input=Information, data, or instructions, such as text prompts or uploaded images, datasets, documents, or audio recordings, provide by user to the AI model.

User Output=Content or data, such as text, images, music, or code, generated by the AI model in response to the user Input.

NCDIT Terms and Conditions

Ownership, Confidentiality, And Security

The State owns all information, materials, and data provided to the Vendor.

Vendor must keep these materials confidential.

Vendor must keep secure all data collected, stored, and processed by Vendor's Services.

State owns all deliverables created by the Vendor, except those licensed to the State by the Vendor.

Issue Spotting: Confidentiality

Will the Vendor's AI model maintain the privacy or confidentiality of the Inputs?

How will the Vendor's AI model maintain the security of the Inputs?

Will the State's Input be used to the train the AI model?

Confidentiality Continued

Does the State's Input contain sensitive, confidential, or proprietary information that would be exposed if those inputs were used to train model?

Does the Vendor disclaim or decline to ensure the confidentiality or security of the user's Inputs?

Does the Vendor prohibit the input of confidential or private information?

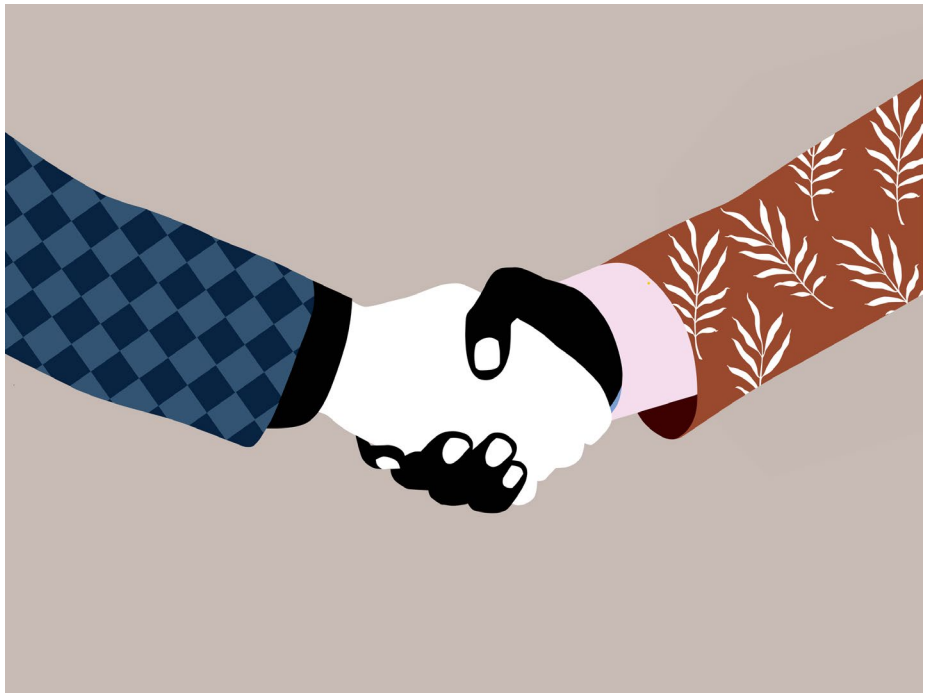
Example of Vendor Terms and Conditions

You retain ownership of all information you input into the AI Features ("Input") and any responses generated based on Input that you provide to the AI Features ("Output"). Notwithstanding the foregoing, you acknowledge that due to the nature of the AI Features and artificial intelligence generally, Output may not be unique, and other users may receive similar content from the AI Features. Responses that are requested by and generated for other users of the Features are not considered [your] Output.

Vendor Terms and Conditions Continued

- You agree not to input any sensitive, proprietary, personal information, or confidential data into the AI Features unless you have obtained all necessary rights and consents to do so.

- You agree that Vendor may use your Input, and Output generated from such Inputs, to train (including, without limitation, fine-tuning) or otherwise improve Vendor's underlying models; solely to the extent Output generated by such models do not contain, disclose any of your Confidential Information.



More Vendor Terms and Conditions

- THE AI FEATURES ARE PROVIDED "AS-IS" WITHOUT ANY WARRANTIES OR GUARANTEES WHATSOEVER. It is your responsibility to review and verify the accuracy of all Output before using it. You represent and warrant that you have the necessary rights and permissions to provide the Input and agree to indemnify and hold CompanyX, as well as its affiliates, officers, and directors, harmless against any third-party claims and associated losses, damages, costs or injury (including without limitation attorneys' fees) of any nature arising from or caused by a breach of this warranty.

NCDIT Terms and Conditions

Infringement and Indemnification

NCDIT Terms and Conditions require the Vendor to defend and indemnify the State against any claims that the Vendor's Software, Products, or Services infringe a patent, copyright, Trademark, or trade secret in the United States.

Plain Language Translation

If someone complains that the Vendor's "stuff" is copying someone else's "stuff" without permission, then the Vendor will take care of the problem, so the State does not have to.

The State does not bear the financial burden or legal responsibility for such claims.

Issue Spotting

Infringement Issues and Shifting Liability

Does the Vendor disclaim warranties of non-infringement for the AI tools?

Does the Vendor state that it will not defend and indemnify for infringement claims arising from the AI model or AI output?

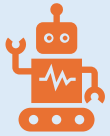
Does the Vendor refuse to indemnify if the user combines the AI output with other content?

Does the Vendor require the user to indemnify and hold harmless the Vendor for the user's Inputs or Outputs?

Shifting Liability Continued

Does the Vendor require the user to indemnify and hold harmless the Vendor for the user's Inputs or Outputs?

If Vendor promises to defend and indemnify for infringement, does the Vendor impose conditions or limitations on this promise?



Will Vendor defend and indemnify if the AI output is only used within the scope of the license?

Consider Your Use Case

Are Vendor's IP infringement conditions contrary to the agency-user's "use case"/objectives or goals?

How does the agency intend to use the Output?

Does the agency intend to combine the AI-Output with other content?

Can the agency comply with the scope of the license?

These questions relate back to gathering and understanding the **facts** for the GenAI purchase.

Example of Vendor Terms and Conditions

You represent and warrant that you have the necessary rights and permissions to provide the Input and **agree to indemnify and hold** Vendor, as well as its affiliates, officers, and directors, **harmless** against any third-party claims and associated losses, damages, costs or injury (including without limitation attorneys' fees) of any nature arising from or caused by a breach of this warranty.

The AI Features are not intended to process works of authorship. You agree to **hold** Vendor and its third-party providers **harmless** from any copyright infringement claims related to your Input into the AI Features.

Due Diligence

Researching the Vendor helps to identify any potential risks and ensures the Vendor's credibility.

Questions to Consider

Are there any lawsuits pending against the Vendor, such as copyright infringement actions?

Has the Vendor been the subject of regulatory investigations, such as international privacy regulators?

Protecting the State's Interests In the Procurement Process

- Understand the AI System, Feature or Tool.
- Gather Facts. Ask Questions.
- How Does the AI Work?
- How will the Agency use the AI?
- What is the source of the Training Data?
- What is the architecture of the System?
- How is Confidentiality and Security Maintained?
- What are the rights and obligations of the parties in the use of the AI under the Vendor's Terms and Conditions?

Protecting the State's Interests Continued

Consult Legal Counsel

Present Legal Counsel with the Facts of the AI purchase, including the Vendor's Terms and Conditions.

Protecting Continued

Use Terms and Conditions to manage the inherent risks and uncertainties of AI Systems. For example:

- Prohibit Vendor's use of AI absent authorization by the State.
- Secure Vendor's warranty that:
  - Vendor will monitor the AI to ensure accurate performance in accordance with specifications
  - Vendor has complied with all laws and regulations applicable to the development of the AI and the training of the AI (e.g., privacy law compliance)
  - Vendor will comply with the State's security standards, and
  - There are no claims against the Vendor's AI.

Protecting Continued

- Prohibit and limit the Vendor's use of the State's Inputs and Outputs for training or improving the AI.

- Require Vendor to defend and indemnify for infringement with respect to AI Outputs.

Where are we today?

Where are we going?

How we can use your help!