Artificial Intelligence-Enabled Cyber Training

An Approach to Accelerated Training Development

CPT Zachary Szewczyk

3rd Multi Domain Task Force, U.S. Army Pacific

The overall classification of this brief is: **UNCLASSIFIED**

Version: **9.0**

Disclaimer: The views expressed here are those of the author and do not necessarily reflect the official position of the Department of the Army or Department of Defense.

Agenda & Purpose

Agenda

- Title Slide
- **Agenda & Purpose**
- Cyber Training and Certification Pipeline
- Manual to AI-Enabled to AI-Driven Training Development
- Accelerated Training Development Process
- AI- Versus Human-Generated Training Material
- Questions, Comments, & Closing

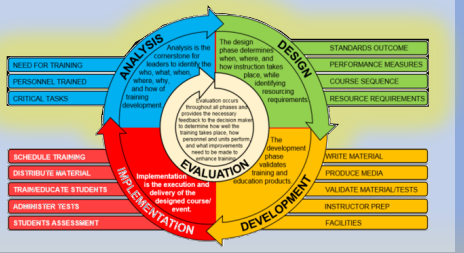
Purpose

The purpose of today's brief is to explain how artificial intelligence (AI) has accelerated the development of defensive cyber analyst training. I'll briefly frame the problem, then delve into the process, key inputs and outputs at each stage, and close with a look at the final products of this initiative.

Cyber Training and Certification Pipeline

LEGEND
- Resourced / Completed
- Resourcing / In Progress
- Not Available / Not Started

"The Army learning Concept for 2030- 2040 focuses on individual, tailored, and **career-long** learning integrated seamlessly along a learning continuum with **unit training systems and processes** to support the conduct of **multidomain operations**." – TRADOC Pamphlet 525-8-2



Proficiency

Master: Supervise training

Senior: Operate without supervision

Basic: Operate with supervision

Industry Training

- AND Practical Threat Hunting
- AND Investigation Theory
- Cisco CCNA
- (ISC)2 CISSP
- AND Practical Packet Analysis
- AND CyberChef for Analysts
- SANS FOR 578
- SANS FOR 508
- SANS FOR 572
- SANS SEC 555
- SANS SEC 595
- SANS SEC 599
- SANS FOR 610
- ???
- ???

Unit-Internal Training

- Active Directory
- Enterprise info sys mgmt
- Advanced host attack techniques
- Detection evasion and counter evasion
- Advanced host analysis
- Advanced host forensic collection
- Advanced network attack techniques
- Advanced network analysis
- Exploratory data analysis
- Advanced security concepts
- Indicator-based analysis
- Privilege escalation path enumeration
- Host forensic platform
- Forensic host analysis
- Basic malware analysis
- Endpoint enumeration
- Cyber attribution
- Network attack path enumeration
- Advanced report writing
- Metrics
- Tactic, technique, and procedure-based analysis
- Information collection matrices
- Cueing
- Report writing
- Basic Network Analysis
- Cyber threat frameworks
- Information systems mgmt
- The analysis process
- The intelligence cycle
- DCO Mission Planning
- Windows operating systems
- Scripting
- Incident handling
- Structured analytic techniques
- Capturing, storing, and viewing network data
- Capturing, viewing, and storing host data
- Basic analysis techniques
- Network composition and topology
- Basic host attack techniques
- Cyber risk
- Basic host analysis
- Basic network attack techniques
- Unix operating systems
- Intrusion detection and prevention systems
- The RFI process
- Routing
- Introduction to the CMF
- Basic host forensic collection
- Sources for cyber threat intel
- Basic security concepts
- Basic host forensic collection
- Common protocols and ports
- Elastic Fundamentals
- DDS-M Operator / Admin
- Splunk Fundamentals
- Data 301
- Microsoft Technology Associate
- Data 201
- Red Hat Administration
- Data 101
- CompTIA SEC+
- CompTIA NET+
- AIT

B — S — M

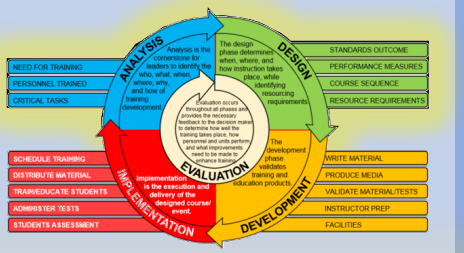
Time

POC: CPT Zachary Szewczyk. (N) zachary.j.szewczyk.mil@army.mil (S) zachary.j.szewczyk.mil@mail.smil.mil (J) zachary.szewczyk.mil@army.ic.gov

Cyber Training and Certification Pipeline

LEGEND
- Resourced / Completed
- Resourcing / In Progress
- Not Available / Not Started

"The Army learning Concept for 2030- 2040 focuses on individual, tailored, and **career-long** learning integrated seamlessly along a learning continuum with **unit training systems and processes** to support the conduct of **multidomain operations**." – TRADOC Pamphlet 525-8-2



Proficiency

Master: Supervise training

Senior: Operate without supervision

Basic: Operate with supervision

Industry Training

- ???
- ???
- SANS FOR 610
- SANS SEC 599
- SANS SEC 595
- SANS SEC 555
- SANS FOR 572
- SANS FOR 508
- SANS FOR 578
- AND CyberChef for Analysts
- AND Practical Packet Analysis
- (ISC)2 CISSP
- Cisco CCNA
- AND Investigation Theory
- AND Practical Threat Hunting
- Elastic Fundamentals
- DDS-M Operator / Admin
- Splunk Fundamentals
- Data 301
- Microsoft Technology Associate
- Data 201
- Red Hat Administration
- Data 101
- CompTIA SEC+
- CompTIA NET+
- AIT

Unit-Internal Training

- Endpoint enumeration
- Basic malware analysis
- Metrics
- Forensic host analysis
- Advanced report writing
- Host forensic platform
- Network attack path enumeration
- Privilege escalation path enumeration
- Cyber attribution
- Indicator-based analysis
- Advanced security concepts
- Exploratory data analysis
- Advanced network analysis
- Advanced network attack techniques
- Advanced host forensic collection
- Advanced host analysis
- Detection evasion and counter evasion
- Report writing
- Advanced host attack techniques
- Cueing
- Enterprise info sys mgmt
- Information collection matrices
- Active Directory
- Tactic, technique, and procedure-based analysis
- Basic security concepts
- Basic host forensic collection
- Common protocols and ports
- Intrusion detection and prevention systems
- The RFI process
- Routing
- Introduction to the CMF
- Basic host forensic collection
- Sources for cyber threat intel
- Basic host analysis
- Basic network attack techniques
- Unix operating systems
- Network composition and topology
- Basic host attack techniques
- Cyber risk
- Capturing, storing, and viewing network data
- Capturing, viewing, and storing host data
- Basic analysis techniques
- DCO Mission Planning
- Windows operating systems
- Scripting
- Incident handling
- Structured analytic techniques
- Basic Network Analysis
- Cyber threat frameworks
- Information systems mgmt
- The analysis process
- The intelligence cycle

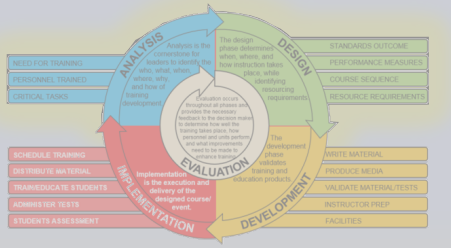
B S M

Time

Cyber Training and Certification Pipeline

LEGEND
- Resourced / Completed
- Resourcing / In Progress
- Not Available / Not Started

"The Army learning Concept for 2030- 2040 focuses on individual, tailored, and **career-long** learning integrated seamlessly along a learning continuum with **unit training systems and processes** to support the conduct of **multidomain operations**." – TRADOC Pamphlet 525-8-2



Proficiency
- Master: Supervise training
- Senior: Operate without supervision
- Basic: Operate with supervision

Industry Training

Unit-Internal Training

M
- ???
- ???
- SANS FOR 610
- SANS SEC 599
- SANS SEC 595
- Endpoint enumeration
- Basic malware analysis
- Metrics
- Forensic host analysis
- Advanced report writing
- Host forensic platform
- Network attack path enumeration
- Privilege escalation path enumeration
- Cyber attribution

S
- SANS SEC 555
- SANS FOR 572
- SANS FOR 508
- SANS FOR 578
- AND CyberChef for Analysts
- AND Practical Packet Analysis
- (ISC)2 CISSP
- Cisco CCNA
- AND Investigation Theory
- AND Practical Threat Hunting
- Indicator-based analysis
- Advanced security concepts
- Exploratory data analysis
- Advanced network analysis
- Advanced network attack techniques
- Advanced host forensic collection
- Advanced host analysis
- Detection evasion and counter evasion
- Report writing
- Advanced host attack techniques
- Cueing
- Enterprise info sys mgmt
- Information collection matrices
- Active Directory
- Tactic, technique, and procedure-based analysis

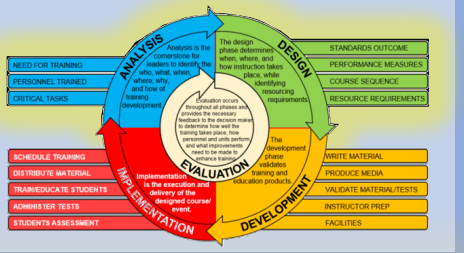
B
- Elastic Fundamentals
- DDS-M Operator / Admin
- Splunk Fundamentals
- Data 301
- Microsoft Technology Associate
- Data 201
- Red Hat Administration
- Data 101
- CompTIA SEC+
- CompTIA NET+
- AIT
- Basic security concepts
- Basic host forensic collection
- Common protocols and ports
- Intrusion detection and prevention systems
- The RFI process
- Routing
- Introduction to the CMF
- Basic host forensic collection
- Sources for cyber threat intel
- Basic host analysis
- Basic network attack techniques
- Unix operating systems
- Network composition and topology
- Basic host attack techniques
- Cyber risk
- Capturing, storing, and viewing network data
- Capturing, viewing, and storing host data
- Basic analysis techniques
- DCO Mission Planning
- Windows operating systems
- Scripting
- Incident handling
- Structured analytic techniques
- Basic Network Analysis
- Cyber threat frameworks
- Information systems mgmt
- The analysis process
- The intelligence cycle

Time

Cyber Training and Certification Pipeline

LEGEND
Resourced / Completed
Resourcing / In Progress
Not Available / Not Started

"The Army learning Concept for 2030- 2040 focuses on individual, tailored, and **career-long** learning integrated seamlessly along a learning continuum with **unit training systems and processes** to support the conduct of **multidomain operations**." – TRADOC Pamphlet 525-8-2



Proficiency

Master: Supervise training

Senior: Operate without supervision

Basic: Operate with supervision

Industry Training

Unit-Internal Training

S

M

B

???
???
SANS FOR 610
SANS SEC 599
SANS SEC 595
SANS SEC 555
SANS FOR 572
SANS FOR 508
SANS FOR 578
AND CyberChef for Analysts
AND Practical Packet Analysis
(ISC)2 CISSP
Cisco CCNA
AND Investigation Theory
AND Practical Threat Hunting
Endpoint enumeration
Basic malware analysis
Metrics
Forensic host analysis
Advanced report writing
Host forensic platform
Network attack path enumeration
Privilege escalation path enumeration
Cyber attribution
Indicator-based analysis
Advanced security concepts
Exploratory data analysis
Advanced network analysis
Advanced network attack techniques
Advanced host forensic collection
Advanced host analysis
Detection evasion and counter evasion
Report writing
Advanced host attack techniques
Cueing
Enterprise info sys mgmt
Information collection matrices
Active Directory
Tactic, technique, and procedure-based analysis
Elastic Fundamentals
DDS-M Operator / Admin
Splunk Fundamentals
Data 301
Microsoft Technology Associate
Data 201
Red Hat Administration
Data 101
CompTIA SEC+
CompTIA NET+
AIT
Basic security concepts
Basic host forensic collection
Common protocols and ports
Intrusion detection and prevention systems
The RFI process
Routing
Introduction to the CMF
Basic host forensic collection
Sources for cyber threat intel
Basic host analysis
Basic network attack techniques
Unix operating systems
Network composition and topology
Basic host attack techniques
Cyber risk
Capturing, storing, and viewing network data
Capturing, viewing, and storing host data
Basic analysis techniques
DCO Mission Planning
Windows operating systems
Scripting
Incident handling
Structured analytic techniques
Basic Network Analysis
Cyber threat frameworks
Information systems mgmt
The analysis process
The intelligence cycle

Time

Manual to AI-Enabled to AI-Driven Training Development

Key Data Points & Results

- Started with ~5,000 words in module descriptions (prompts).
- Manual training development:
  - 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-enabled training development:
  - **Optimizing for the wrong constraint** led to similar 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-driven training development:
  - **Training development reduced to minutes.**
  - LLMs transformed 5,000 word module descriptions into 60,000 words of outlines into **284,000 words on 1,600** slides of course material.
  - Total cost of project: **$34.68**

Manual

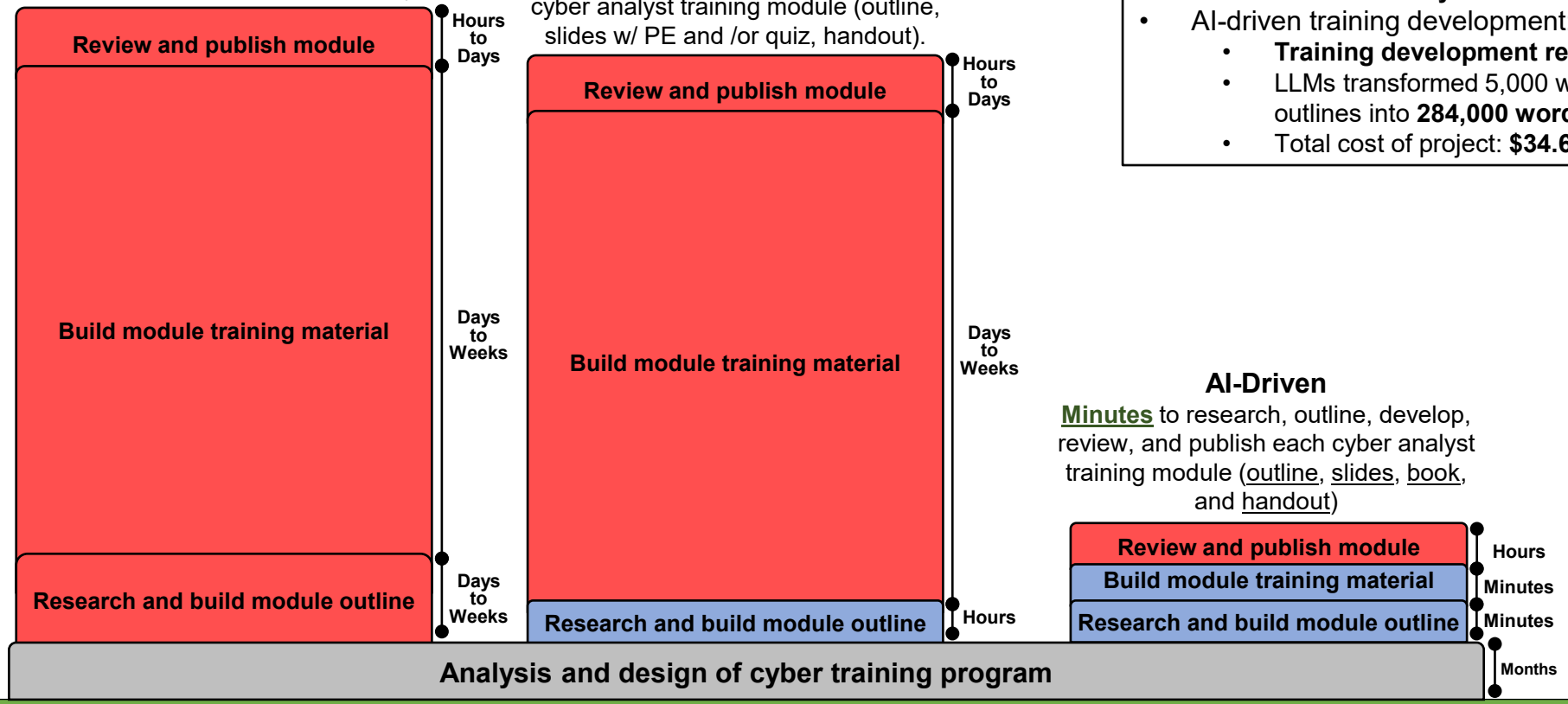
1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout)

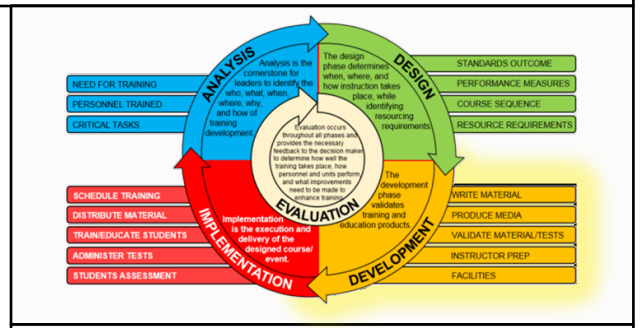
AI-Enabled

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout).

AI-Driven

Minutes to research, outline, develop, review, and publish each cyber analyst training module (outline, slides, book, and handout)





"Future Army forces require the capability to **rapidly understand, develop, and implement** training and education changes in order to meet shifting operational demands in the MDO environment. (paras 3-6.b and 3-8.b.)" – TRADOC Pamphlet 525-8-2

Manual to AI-Enabled to AI-Driven Training Development

Key Data Points & Results

- Started with ~5,000 words in module descriptions (prompts).
- Manual training development:
  - 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-enabled training development:
  - **Optimizing for the wrong constraint** led to similar 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-driven training development:
  - **Training development reduced to minutes.**
  - LLMs transformed 5,000 word module descriptions into 60,000 words of outlines into **284,000 words on 1,600** slides of course material.
  - Total cost of project: **$34.68**

Manual

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout)

- Review and publish module — Hours to Days
- Build module training material — Days to Weeks
- Research and build module outline — Days to Weeks

AI-Enabled

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout).

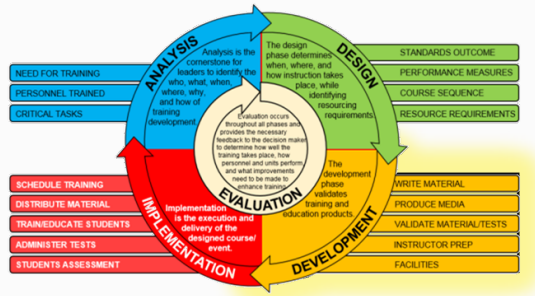
- Review and publish module — Hours to Days
- Build module training material — Days to Weeks
- Research and build module outline — Hours

AI-Driven

Minutes to research, outline, develop, review, and publish each cyber analyst training module (outline, slides, book, and handout)

- Review and publish module — Hours
- Build module training material — Minutes
- Research and build module outline — Minutes

Analysis and design of cyber training program — Months



"Future Army forces require the capability to **rapidly understand, develop, and implement** training and education changes in order to meet shifting operational demands in the MDO environment. (paras 3-6.b and 3-8.b.)" – TRADOC Pamphlet 525-8-2

Manual to AI-Enabled to AI-Driven Training Development

Key Data Points & Results

- Started with ~5,000 words in module descriptions (prompts).
- Manual training development:
  - 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-enabled training development:
  - **Optimizing for the wrong constraint** led to similar 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-driven training development:
  - **Training development reduced to minutes.**
  - LLMs transformed 5,000 word module descriptions into 60,000 words of outlines into **284,000 words on 1,600** slides of course material.
  - Total cost of project: **$34.68**

Manual

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout)

- Review and publish module — Hours to Days
- Build module training material — Days to Weeks
- Research and build module outline — Days to Weeks

AI-Enabled

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout).

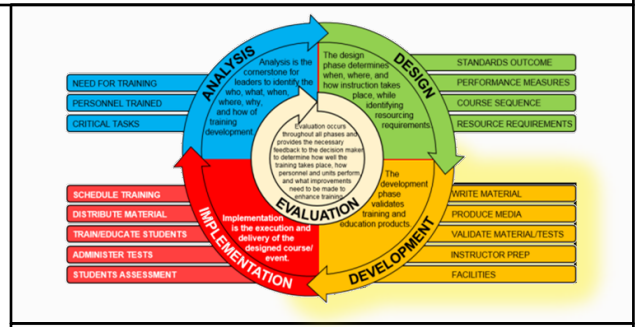
- Review and publish module — Hours to Days
- Build module training material — Days to Weeks
- Research and build module outline — Hours

AI-Driven

Minutes to research, outline, develop, review, and publish each cyber analyst training module (outline, slides, book, and handout)

- Review and publish module — Hours
- Build module training material — Minutes
- Research and build module outline — Minutes

Analysis and design of cyber training program — Months



"Future Army forces require the capability to **rapidly understand, develop, and implement** training and education changes in order to meet shifting operational demands in the MDO environment. (paras 3-6.b and 3-8.b.)" – TRADOC Pamphlet 525-8-2

Manual to AI-Enabled to AI-Driven Training Development

Key Data Points & Results

- Started with ~5,000 words in module descriptions (prompts).
- Manual training development:
  - 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-enabled training development:
  - **Optimizing for the wrong constraint** led to similar 1 to 2 months to research, outline, develop, review, and publish **each** cyber analyst training module.
  - **Estimated ~1 year to develop entire unit-internal training pipeline.**
- AI-driven training development:
  - **Training development reduced to minutes.**
  - LLMs transformed 5,000 word module descriptions into 60,000 words of outlines into **284,000 words on 1,600** slides of course material.
  - Total cost of project: **$34.68**

Manual

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout)

- Review and publish module — Hours to Days
- Build module training material — Days to Weeks
- Research and build module outline — Days to Weeks

AI-Enabled

1 to 2 months to research, outline, develop, review, and publish each cyber analyst training module (outline, slides w/ PE and /or quiz, handout).

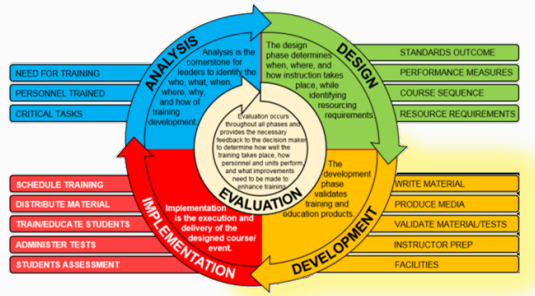
- Review and publish module — Hours to Days
- Build module training material — Days to Weeks
- Research and build module outline — Hours

AI-Driven

Minutes to research, outline, develop, review, and publish each cyber analyst training module (outline, slides, book, and handout)

- Review and publish module — Hours
- Build module training material — Minutes
- Research and build module outline — Minutes

Analysis and design of cyber training program — Months



"Future Army forces require the capability to **rapidly understand, develop, and implement** training and education changes in order to meet shifting operational demands in the MDO environment. (paras 3-6.b and 3-8.b.)" – TRADOC Pamphlet 525-8-2

Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

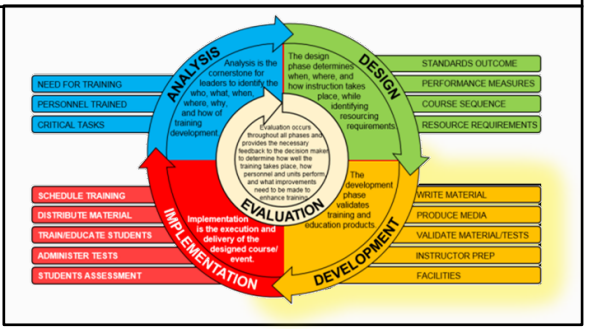
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

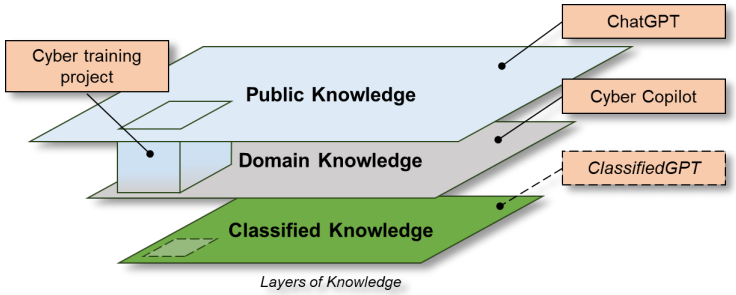
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

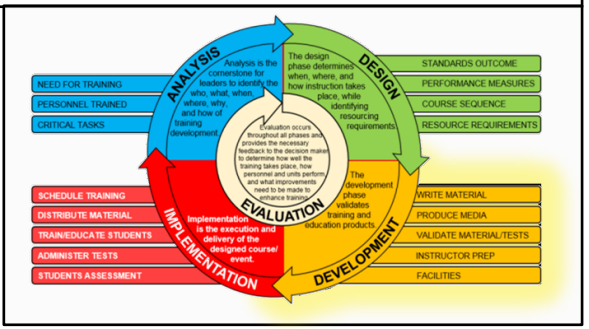
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

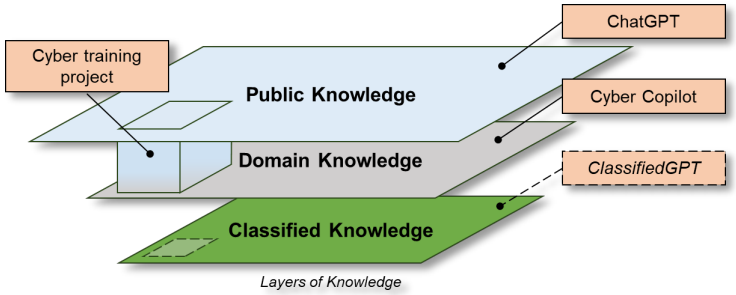
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

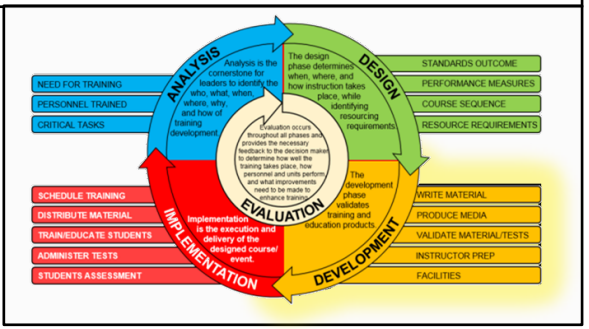
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

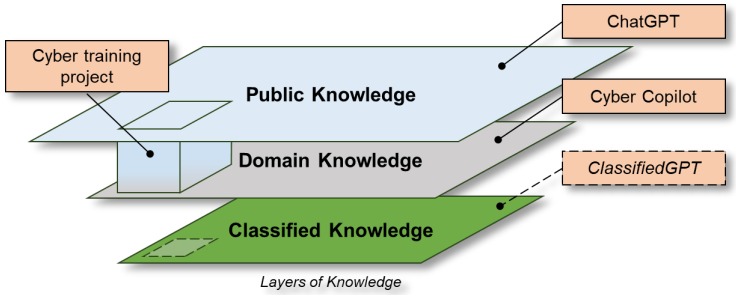
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

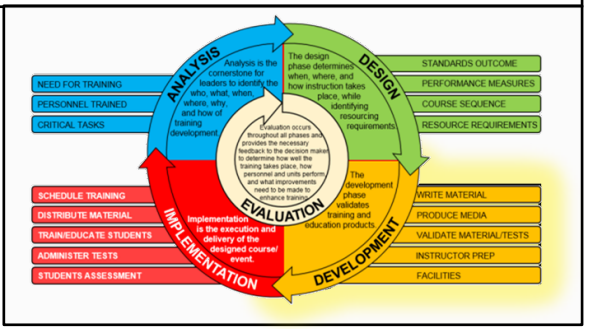
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

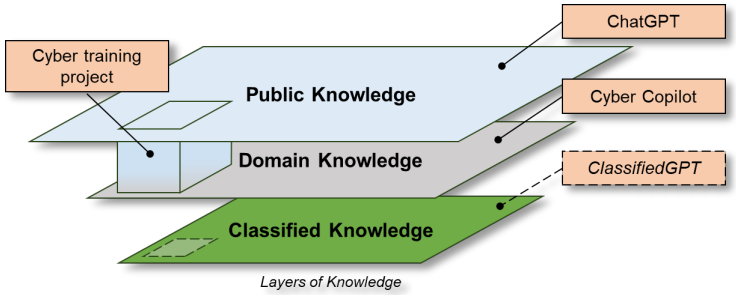
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

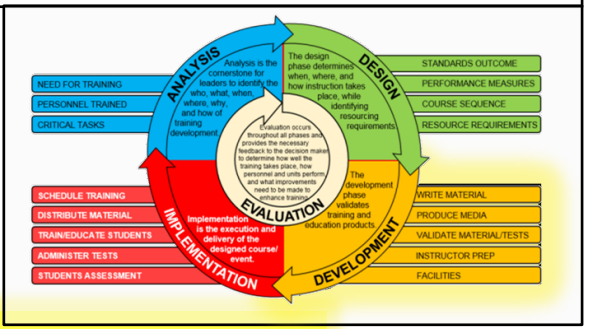
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

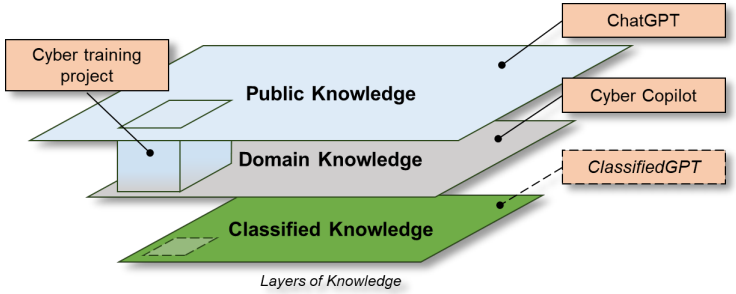
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

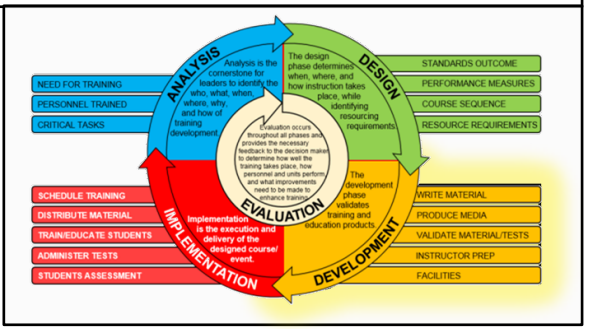
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

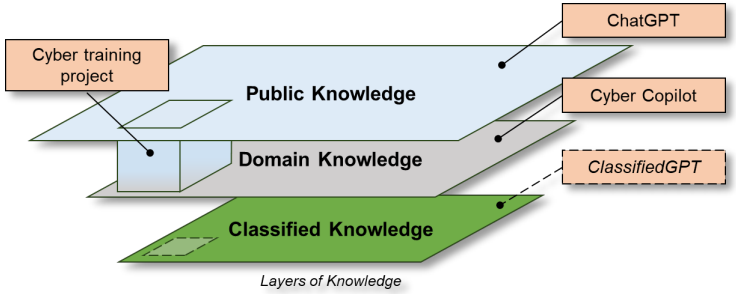
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





Accelerated Training Development Process

Step 1:
Manually create training module title and description.

- Individual training module concepts developed based on operational experience and prior industry training.
- Not necessarily *the* answer, but a good start.
- Current concept consists of **54** unique training modules.

Step 2:
Script extracts all module titles and descriptions, then prompts GPT 3.5 to create a module outline based on prompt.

- Module titles and brief, one paragraph descriptions provide enough context to create a good first draft of a training outline using the Large Language Model (LLM) GPT 3.5.

Step 3:
Script parses outline, then prompts GPT 4 to reorder, expand, and revise outline to satisfy JQR requirements. Final outline saved to disk.

- Manually extracted all **598** Job Qualification Record requirements for Host Analyst and Network Analyst, then mapped to individual training modules.
- Outline is revised using more capable LLM GPT 4 based on JQR requirements from manual mapping.

Step 4:
Script reads module outlines, then prompts GPT 4 to generate study guide for each module.

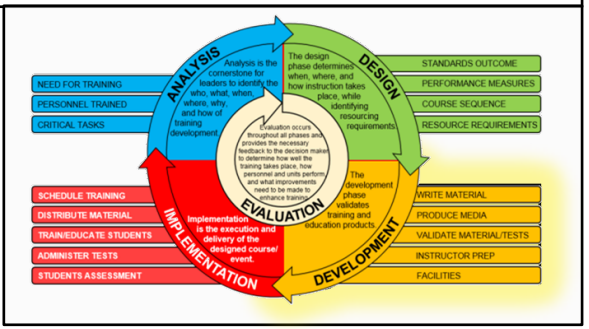
- Revised outline is fed to GPT 4 to develop study guide for each training module.
- Both training and handout are saved to disk.

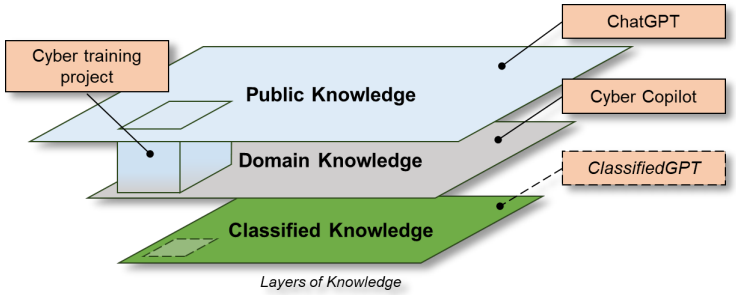
Step 5:
Script reads each outline, creates a template slide presentation, then iteratively prompts GPT 3.5 to explain each topic from the outline. Explanation is appended to the appropriate slide presentation to produce finished product.

- For each training module:
  - Read each topic in the outline.
  - Prompt GPT4 to explain topic.
  - Append output to LaTeX slide deck.
  - Generate slide deck with XeLaTeX engine.

Ensemble Model Approach

- **ChatGPT**: Web interface and long context window makes this particularly well-suited to **human-in-the-loop iteration.**
- **GPT 3.5 Turbo**: Accessible via API and ChatGPT. Fast, accurate, and well-suited to handling **explicit instructions**.
- **GPT 4**: Accessible via API and ChatGPT. Slower than GPT 3.5, more expensive, but well-suited to **tasks** that may require inference or reasoning.
- **Google Gemini**: Accessible via (free) API and web interface. Quality between GPT 3.5 and GPT 4.





AI- Versus Human-Generated Training Material

UNCLASSIFIED

Passive Tap

UNCLASSIFIED

Active Tap

UNCLASSIFIED

Switch SPAN Port

- A Switched Port Analyzer (SPAN) port, commonly referred to as a 'mirror port', is a designated port on a network switch that is configured to receive and replicate packets from one or more target ports. This functionality enables network administrators or security personnel to monitor network traffic entering or leaving these target ports.
- The primary goal of a SPAN port is to aid in troubleshooting, monitoring, and analyzing network issues without interrupting the normal flow of data. This makes it an invaluable tool for identifying anomalies or potential security threats within a network.
- Configuring a SPAN port typically involves selecting a source port (or ports) whose traffic will be mirrored and a destination port where the mirrored traffic will be sent. It's important to ensure that the destination port has adequate bandwidth to handle the additional traffic, as it will receive both its own traffic and the mirrored traffic.
- It is essential to be cautious when using SPAN ports. Overloading the destination port can lead to dropped packets and incomplete data capture. Additionally, the SPAN port itself does not introduce any latency or alter the packets, making it a transparent and reliable tool for network analysis.
- When using a SPAN port for capturing data, it's often paired with network analysis tools or intrusion detection systems. These tools can interpret, store, and further analyze the mirrored traffic for insights or alerts.

AI-Generated Training Material

Port Mirroring

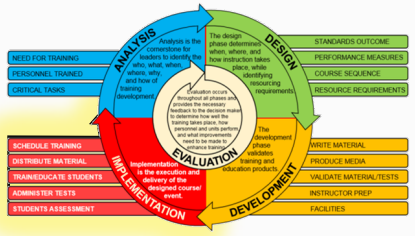
Port Mirroring

SPAN

- Limitations
  - Performance degradation can occur when mirrored traffic exceeds the capacity of the listening or monitor interface, to prevent this issue it is recommended to:
    - Limit the ports monitored to prevent duplicate data
    - Use a firewall filter to send specific traffic to a port supporting mirroring
  - If the device monitor interface capacity is insufficient to handle traffic from the source port(s), overflow packets are dropped
  - Both ingress and egress switched traffic not originating from the switch can be mirrored from interfaces, but this is not the case for VLAN traffic
  - Only traffic entering a VLAN can be mirrored, and you cannot copy packets exiting a VLAN with mirroring

Contractor-Generated Training Material

AI won't obviate the need for skilled course designers and knowledgeable instructors, but it can produce training material at least as good as what we have today.



Key Takeaways

1. This project has already had significant impact at the **tactical level** but could have a greater impact **across echelons** and **across the force**.
   - The Army has many programs to teach specialists to create lesson materials in the institutional domain, but few opportunities for that training in the operational force.

2. This project **automates the mundane work** of basic information gathering and product creation. It enables **focusing on higher order tasks.**
   - Evaluating the effectiveness of training programs.
   - Improving the quality of training.
   - Integrating emerging research from academia and lessons from the operational force in a far more rapid manner than is done today.

Questions, Comments, & Closing

💡 *All the code for this project is hosted on US Cyber Command's GitLab server, R2D2, here: https://code.levelup.cce.af.mil/3mdtf/idco/training*