

SERC RESEARCH REVIEW 2023 | NOVEMBER 15, 2023

DAU Credential Development and Workforce Development in Artificial Intelligence and Data Analytics

WRT-1073

Sponsor: Defense Acquisition University

Presenter: A. Emrah Bayrak, Assistant Professor



Training the Workforce

- Growing need to catch up with the advances in the cyber domain

S.2551 - AI Training Act

117th Congress (2021-2022)

LAW

Hide Overview ✕

Sponsor: [Sen. Peters, Gary C. \[D-MI\]](#) (Introduced 07/29/2021)

Committees: Senate - Homeland Security and Governmental Affairs

Committee Meetings: [08/04/21 10:30AM](#)

Committee Reports: [S. Rept. 117-40](#)

Latest Action: 10/17/2022 Became [Public Law No: 117-207](#). ([All Actions](#))

Roll Call Votes: There has been [1 roll call vote](#)

Tracker:  Introduced > Passed Senate > Passed House > To President > **Became Law**



create a training program to help Federal employees responsible for purchasing AI technologies better understand its risks and benefits



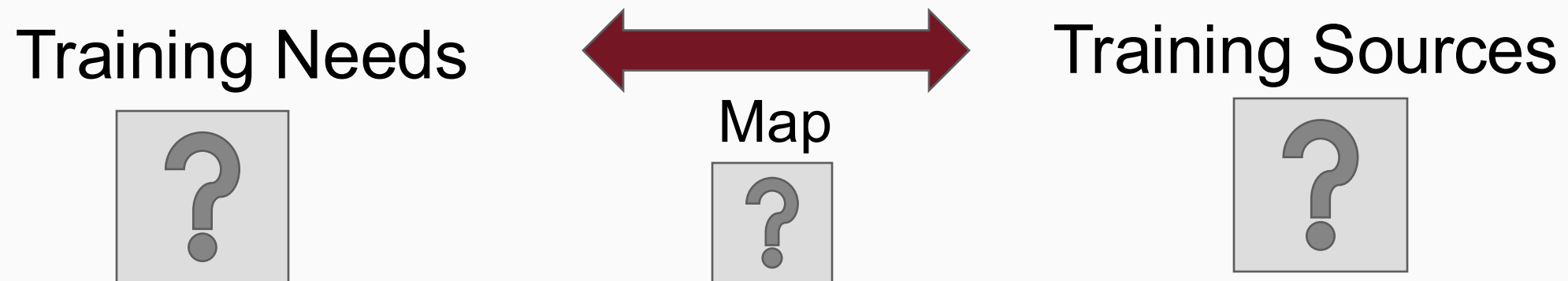
CHIEF DIGITAL AND ARTIFICIAL INTELLIGENCE OFFICE



responsible for the acceleration of the DoD's adoption of data, analytics, and AI

Project Goal:

- Provide micro-credentials to the defense acquisition workforce for the competencies relating to **Data Analytics** and **Artificial Intelligence**
- Identify training paths tailored for the needs of the defense acquisition workforce

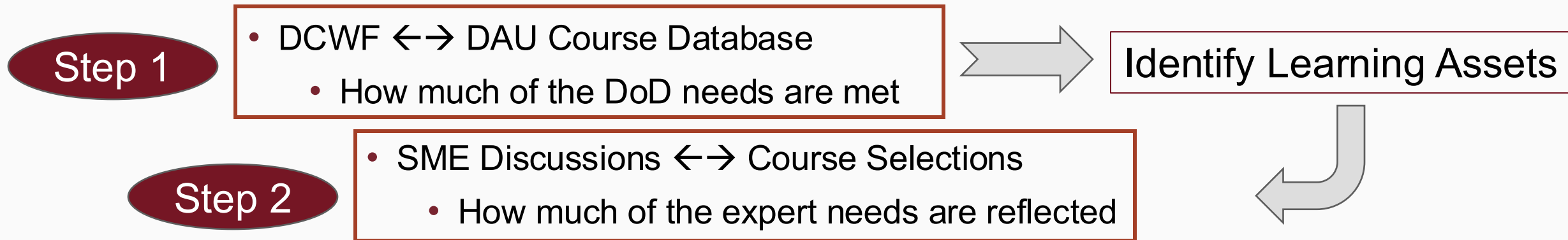


Information Sources

Three sources of information:

- SME Discussions (Voice of experts)
- Defense Cyber Workforce Framework (Voice of DoD)
- DAU Course Database

Analysis Overview:



Voice of Experts

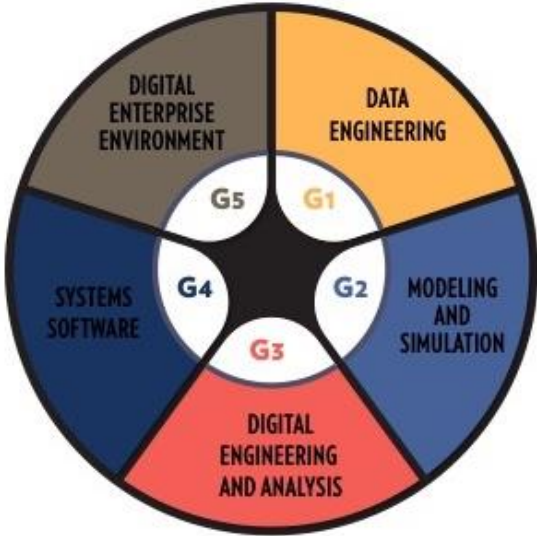
- 13 SMEs from academia, industry and government
 - Experts working for / interacting with several agencies across DoD
 - Varying expertise levels in the area of DA/AI/ML
- 1hr-long discussions focusing on:
 - How DA/AI/ML is used
 - Critical skillsets needed in the area of DA/AI/ML
 - Training preferences

Voice of DoD

AI Workforce Framework

Archetype	Description	Concentration
 Lead AI	Decides policy and doctrine, including how AI tools can or will be used; builds AI vision and plan	Policy Command Agency/Function Lead Acquisitions Manager
 Drive AI	Ensures appropriate AI tools and capabilities are developed and delivered across DOD	Capability Manager Technical Manager Product Manager AI Researcher
 Create AI	Creates AI tools to meet current and future needs	AI/ML Engineer Testing & Evaluation Engineer Data Scientist Deployment Engineer
 Embed AI	Embedded with Employ AI, establishes AI systems and provides end-user support at tactical edge	Technician
 Facilitate AI	Represents users to ensure appropriate AI tools are developed and delivered to address use cases	Product Owner UI/UX Other Technical Experts
 Employ AI	End-users of AI tools, provide feedback on and requirements for AI tools	Operations Intelligence Logistics & Maintenance Health Support

Digital Engineering Competency Framework



Defense Cyber Workforce Framework



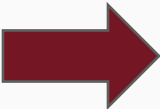
Workforce Elements



Defense Cyber Workforce Framework

Work Role	# of KSATs
AI Adoption Specialist	53
AI Innovation Leader	72
AI Risk & Ethics Specialist	50
AI Test & Evaluation Specialist	75
AI/ML Specialist	71
Data Analyst	38
Data Architect	39
Data Officer	50
Data Operations Specialist	53
Data Scientist	47
Data Steward	39

Total : 587



Tasks
Read, interpret, write, modify, and execute scripts, macros, and functions.
Manipulate and clean large, disparate datasets for bulk analysis to identify connections.
Utilize different programming languages to write code, open files, read files, and write output to different files.
Present technical information to technical and non-technical audiences.
...
Knowledge, Skills, and Abilities
...
Knowledge of how to collect, store, and monitor data.
Knowledge of how to structure and display data.
Knowledge of how to use data to tell a story.
Knowledge of laws, regulations, and policies related to AI, data security/privacy, and use of publicly procured data

DAU Course Database

STAGE

HomeLearningPerformanceConnectReportsHelp Desk

ONLINE CONTENT

(COURSE) Data Science: Data Science Tools

Last Updated 05/01/2023 Duration 47 minutes

Details

Explore a variety of new data science tools available today; the different uses for these tools; and the benefits and challenges in deploying them in this 12-video course. First, examine a data science platform, the nucleus of technologies used to perform data science tasks. You will then explore the analysis process to inspect, clean, transform, and model data. Next, the course surveys integrating and exploring data, coding, and building models using that data, deploying the models to production, and delivering results through applications or by generating reports. You will see how a great data science platform should be flexible and scalable, and it should combine multiple features and capabilities that effectively centralize data science efforts. You will learn the six sequential steps of a typical data science workflow, from defining the objective for the project to reporting the results. Finally, explore DevOps, resources that allow developers and IT to work together in harmony which includes people, processes, and infrastructure; and its typical functionalities including integration, testing, packaging, as well as deployment.

Provider

Percipio-US

Available Language(s)

English (US)

Show Less

Assets from Three Course Providers

- Coursera
- LinkedIn Learning
- Skillsoft (formerly Percipio)

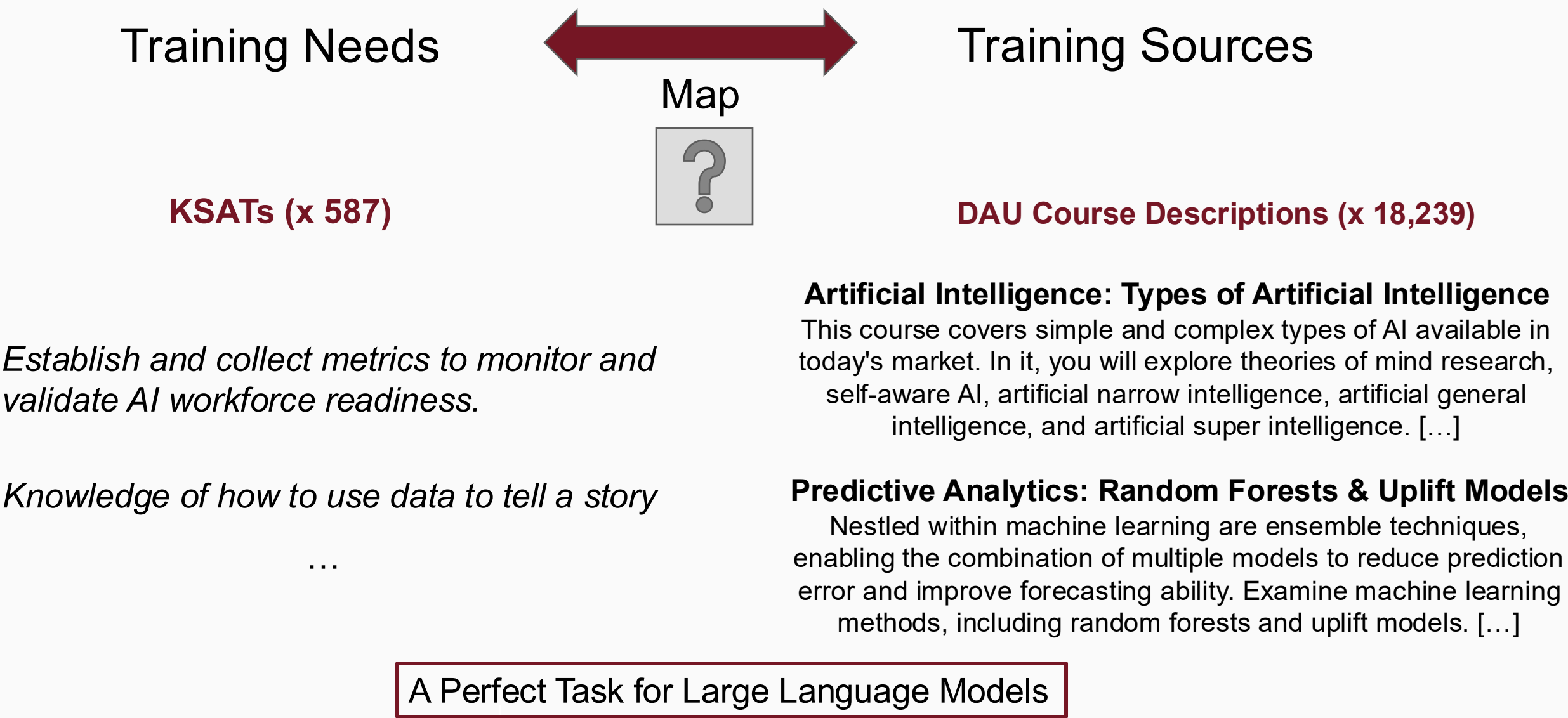
of Relevant Learning Assets > 18,000

Web Scraping

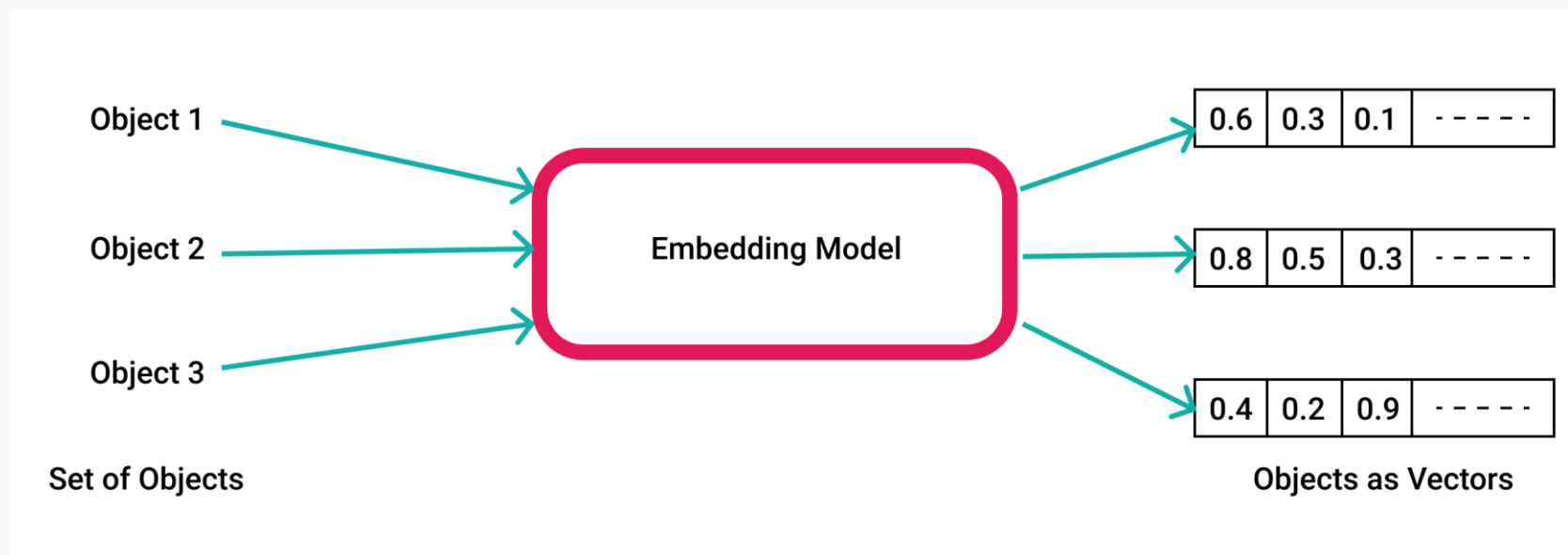


Keywords

Mapping DCWF to DAU Course Database



Mapping DCWF to DAU Course Database



- A whole paragraph of text or any other object can be reduced to a vector
- Allows translating semantic similarity as perceived by humans to proximity in a vector space
 - Similar words or sentences would lie close to each other on the vector space

We compared KSATs with DAU Course Descriptions using this embedding

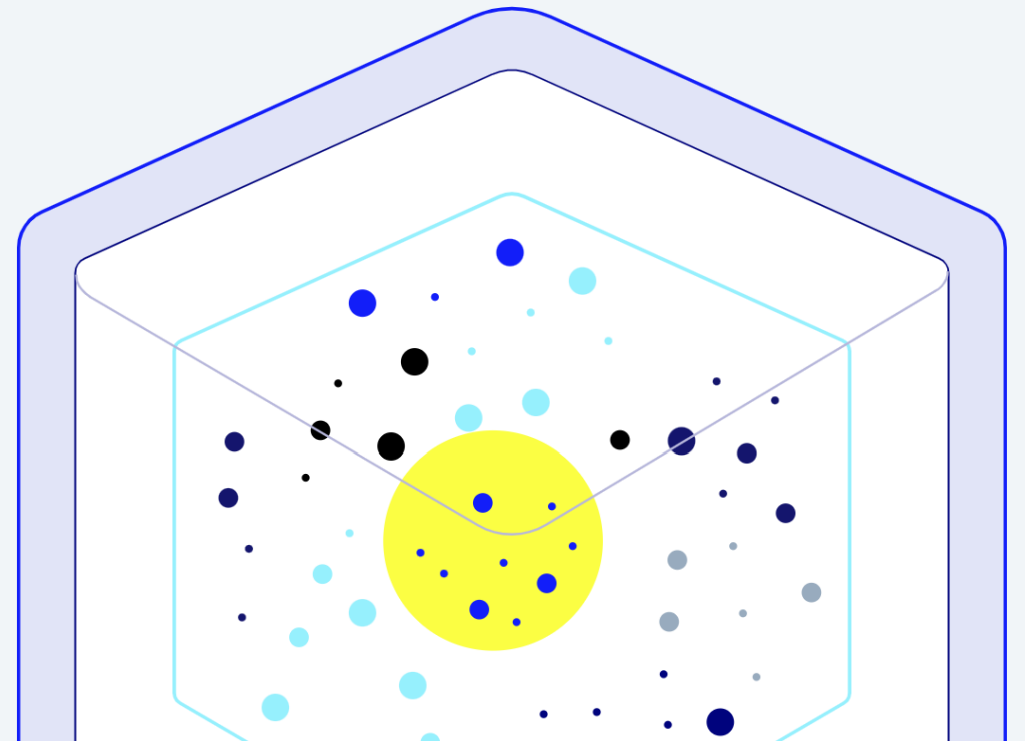
Mapping DCWF to DAU Course Database

- Existing vector database from Pinecone
 - Allows using pre-trained models to create the embeddings
 - Pre-trained models developed using public databases (not necessarily with technical database)

The Pinecone Vector Database

Storing and searching through AI representations of your data lets you build search apps that “get it” and chatbots that answer right. Without exposing your data or fine-tuning models.

[What is a vector database →](#)



Mapping DCWF to DAU Course Database

Each KSAT is mapped to the most similar five courses in the database

More similar 

KSAT		Course 1	Course 2	Course 3	Course 4	Course 5
<i>Establish and collect metrics to monitor and validate AI workforce readiness.</i>	→	<i>Introducing AI to Your Organization</i>	<i>Understand the importance of building an AI-ready culture</i>	<i>Artificial Intelligence: Planning AI Implementation</i>	<i>Responsible AI: Principles and Practical Applications</i>	<i>AI Accountability Essential Training</i>
		Score: 0.82	Score: 0.89	Score: 0.94	Score: 0.99	Score: 1.02
<i>Knowledge of how to use data to tell a story.</i>	→	<i>Visualizing Data for Impact: Data Storytelling</i>	<i>Data Visualization: Storytelling</i>	<i>Telling Stories with Data</i>	<i>Storytelling with Data: Introduction</i>	<i>Data Storytelling Basics</i>
		Score: 0.41	Score: 0.45	Score: 0.47	Score: 0.54	Score: 0.58
...		...	...	...	...	...

While some similarity scores may appear low, the course description still matches the KSAT

Developing Course Playlists

Work Role	# of KSATs
AI Adoption Specialist	53
AI Innovation Leader	72
AI Risk & Ethics Specialist	50
AI Test & Evaluation Specialist	75
AI/ML Specialist	71
Data Analyst	38
Data Architect	39
Data Officer	50
Data Operations Specialist	53
Data Scientist	47
Data Steward	39

Relevance filter (done manually)

Removed irrelevant KSATs:

- Knowledge of the organization's core business/mission processes.
- Knowledge of organization objectives, leadership priorities, and decision-making risks.
- ...



48 Relevant KSATs

Still a large number !

Developing Course Playlists

Example: AI Innovation Leader

- Overlapping courses across KSATs
- E.g., **Artificial Intelligence: Planning AI Implementation**

Linear Programming for Selecting Courses

$$\max \sum_{j=1}^M \sum_{i=1}^N c_{ij} x_{ij}$$

$$\text{subject to } \sum_{i=1}^N x_{ij} = 1$$

$$x_{ij} \in \{0,1\} \forall i,j$$

where

x_{ij} : Decision to assign KSAT i to course j

c_{ij} : Weighted similarity score between KSAT i and course j

Similarity score is weighted by the number of occurrences

Courses

KSAT	C1	C2	C3	C4	C5
1	0.91	0.93	0.93	0.93	0.95
2	0.45	0.66	0.75	0.78	0.82
3	0.64	0.72	0.73	0.74	0.79
4	0.80	0.82	0.83	0.84	0.87
5	0.91	0.96	0.96	0.99	1.00
6	0.82	0.85	0.85	0.89	0.90
7	0.40	0.71	0.80	0.81	0.81
8	1.01	1.02	1.02	1.03	1.03
9	0.65	0.92	0.92	0.97	0.97
10	0.68	0.77	0.81	0.87	0.89
11	0.75	0.77	0.84	0.86	0.89
12	0.77	0.83	0.84	0.90	0.91
13	0.87	0.95	0.96	0.98	1.00
14	0.66	0.74	0.74	0.78	0.80
15	0.88	0.88	0.88	0.89	0.90
16	0.72	0.80	0.81	0.82	0.84
17	0.77	0.84	0.86	0.86	0.89
18	0.71	0.72	0.72	0.79	0.81
19	0.78	0.84	0.85	0.87	0.89
20	0.75	0.76	0.80	0.82	0.82
21	0.77	0.77	0.80	0.82	0.83
22	0.82	0.89	0.95	0.99	1.02
23	0.79	0.88	0.88	0.89	0.91
24	0.90	0.90	0.93	0.95	0.95
25	0.84	0.85	0.89	0.89	0.92
26	0.85	0.89	0.90	0.90	0.92
27	1.09	1.11	1.13	1.13	1.14
28	0.88	0.90	0.94	0.96	0.96
29	0.85	0.86	0.89	0.91	0.92
30	0.72	0.76	0.83	0.84	0.84
31	0.65	0.68	0.75	0.76	0.77
32	0.85	0.86	0.86	0.87	0.88
33	0.75	0.79	0.88	0.88	0.88
34	0.65	0.83	0.85	0.89	0.91
35	0.72	0.73	0.75	0.76	0.77
36	0.67	0.72	0.78	0.80	0.80
37	0.65	0.67	0.69	0.79	0.80
38	0.81	0.82	0.84	0.84	0.89
39	0.64	0.67	0.67	0.77	0.80
40	0.87	0.88	0.88	0.89	0.92
41	0.79	0.84	0.92	0.93	0.93
42	0.78	0.83	0.83	0.84	0.84
43	0.83	0.83	0.87	0.87	0.93
44	0.67	0.73	0.77	0.78	0.81
45	0.82	0.86	0.90	0.92	0.92
46	0.75	0.78	0.78	0.84	0.88
47	1.01	1.01	1.06	1.09	1.09
48	0.78	0.79	0.80	0.82	0.84

KSATs

Sample Playlists

Manually analyzed linear programming results to further refine the list

Seq.	Level	Course Title	Prov.	Duration
1	1	Data Curation Foundations	LL	4hr, 56min
2	1	Meaningful Predictive Modeling	CE	8hr, 33min
3	2	Artificial Intelligence: Types of Artificial Intelligence	SS	47min
4	2	Artificial Intelligence Foundations: Machine Learning	LL	1hr, 17min
5	2	World's Smartest Home – Artificial Intelligence	SS	N/A
6	3	Introducing AI to Your Organization	LL	53min
7	3	Understand the importance of building an AI-ready culture	SS	49min
8	3	Machine Learning and Artificial Intelligence Security Risk: Categorizing Attacks and Failure Modes	LL	1hr, 11min
9	3	Safeguarding AI	LL	36min
10	3	AI Accountability Essential Training	LL	2hr, 16min
Alternative Course Options				
6	3	Artificial Intelligence: Planning AI Implementation	SS	44min
10	3	Responsible AI: Principles and Practical Applications	LL	1hr, 6min

Total: 21 hrs 18 minutes

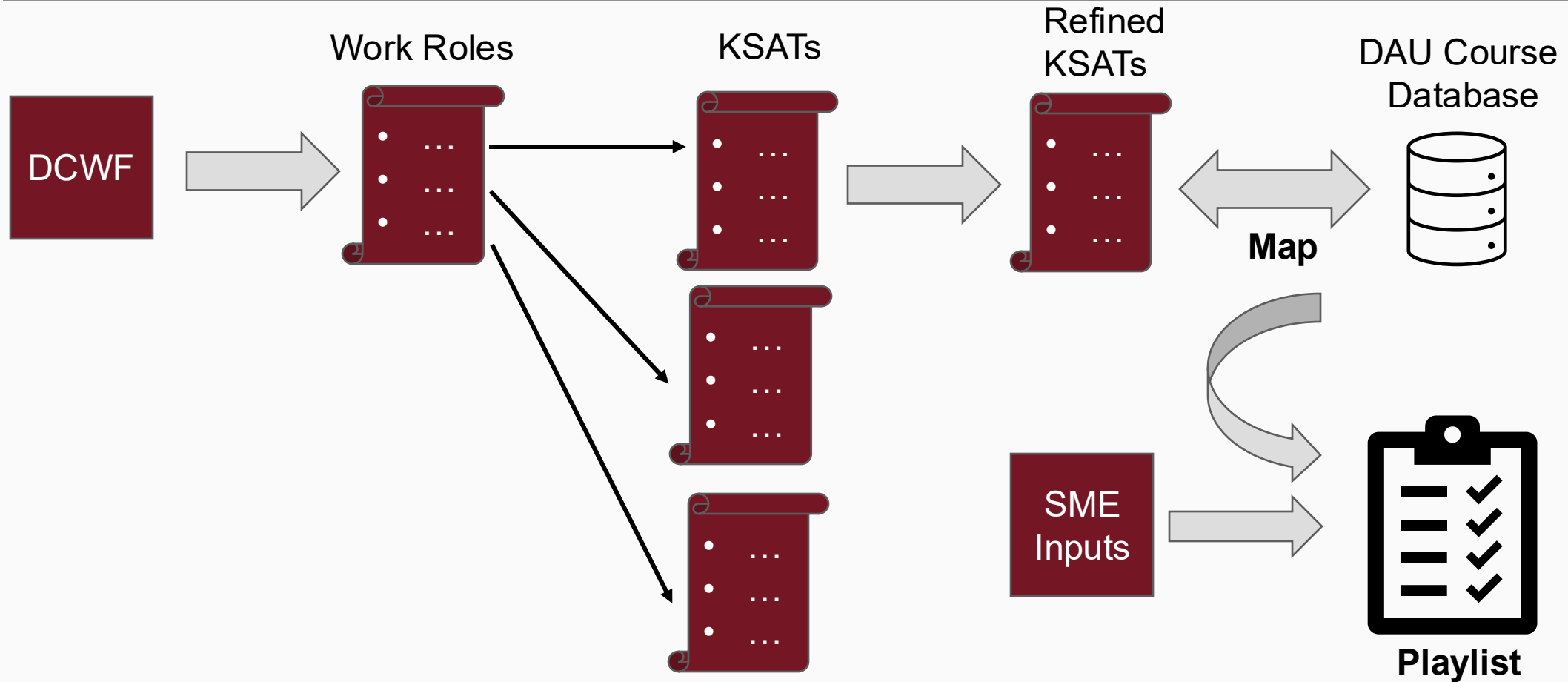
Similar courses listed as alternatives

SME Inputs on Training the Workforce

- Limited time for training
 - Few people have dedicated time for DA/AI/ML training
 - Most people do it at their own time
 - Short courses are preferred
- Many expressed need to define prerequisites for training paths
- A big discrepancy in data literacy between leadership and analysts
 - Leadership lacks critical training in broad understanding of the basics of DA/AI/ML
 - More detailed training is better for analysts

and more ...

Summary



A Generalized Approach to Develop Training Paths for an Organization

Future...

Default (GPT-3.5)



What can I cook with mushrooms, onions, tomatoes and potatoes



With mushrooms, onions, tomatoes, and potatoes, you have a great base for a variety of delicious dishes. Here are a few ideas for what you can cook:



1. **Mushroom and Potato Hash:**

- Heat some oil in a pan and sauté chopped onions until translucent.
- Add diced potatoes and cook until they are golden brown and crispy.
- Add sliced mushrooms and cook until they are tender.
- Season with salt, pepper, and your favorite herbs or spices.

2. **Stuffed Mushrooms:**

- Remove the stems from the mushrooms and chop them finely.
- Mix the chopped mushroom stems with sautéed onions, diced tomatoes, and some breadcrumbs.
- Stuff the mushroom caps with this mixture and bake until the mushrooms are tender.

We have the methods to
build Chat-GPT like
generative tools to suggest
training paths for a given
need

Thank you

Stay connected with SERC Online:



Email the presenter: A. Emrah Bayrak

✉ ebayrak@stevens.edu

Email the research team:

✉ clipizzi@stevens.edu
gvesonde@stevens.edu
nlong@stevens.edu



SYSTEMS
ENGINEERING
RESEARCH CENTER