Skills in the Era of Big Data and Artificial Intelligence

An Open-Source Approach to Linking Skills Data

Presented by: George Washington Institute for Public Policy
Kyle Albert, Mike Sanders, Thomas Weko

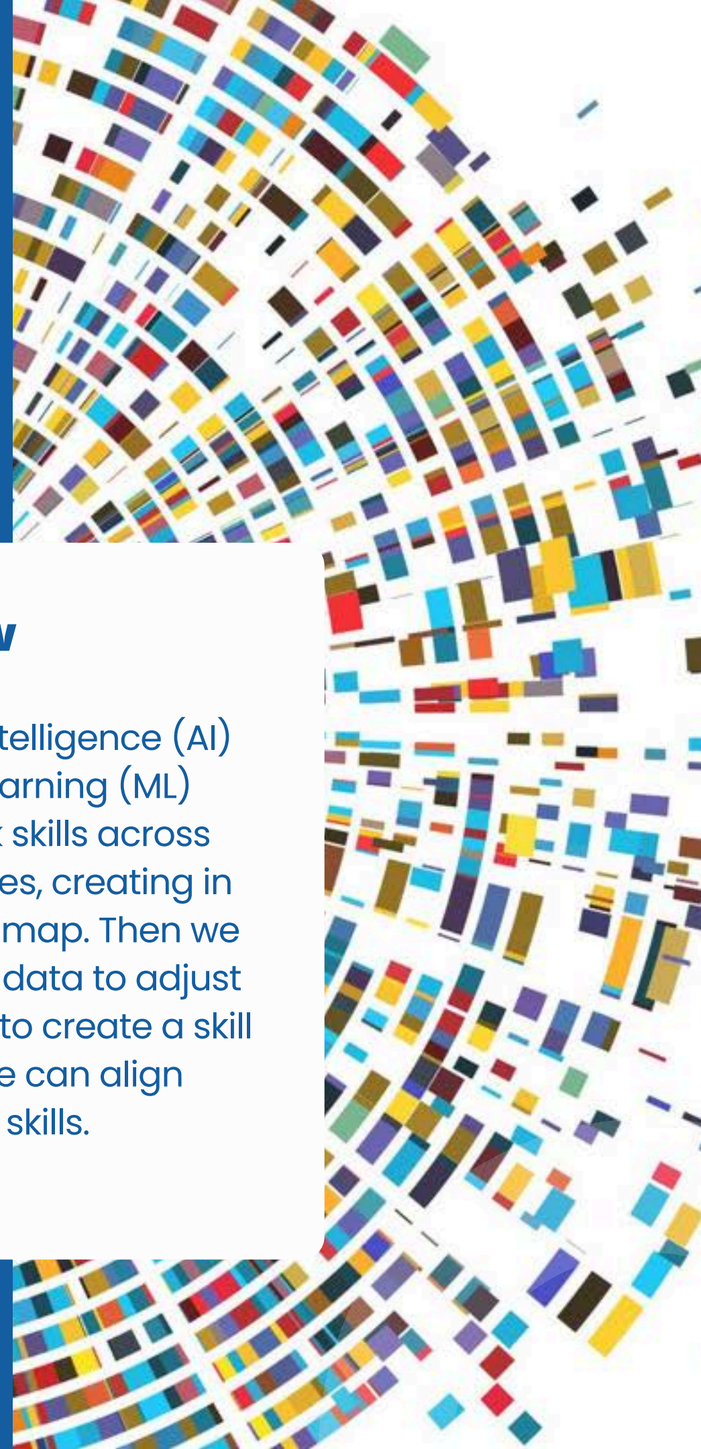
October 31, 2025

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC



Objective

Help learners, educators and employers share trusted and mutually intelligible information about skills.



Why

Skills are the language of education, training, and employment. Skills have increasingly become the focus of government policy, as leaders grapple with skill-biased technological change and aging working populations and aim to promote equity and inclusion through skill-based hiring.

Who

Educators, employers, learners, and policymakers often struggle to communicate in mutually intelligible ways across the diverse skill languages of their different disciplines, occupations and industries, and regions.

How

We use artificial intelligence (AI) and machine learning (ML) algorithms to link skills across multiple taxonomies, creating in effect a taxonomy map. Then we layer on real world data to adjust these relationships to create a skill grid to which we can align individual skills.

LAiSER



Skill Extraction

- Open-source
- Python based
- Pip install or .exe



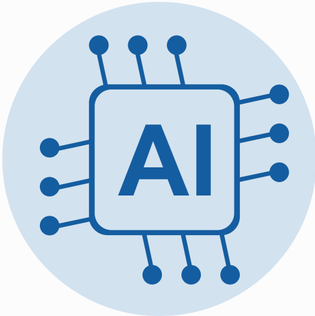
Linked Data

- Education & Workforce data
- Contextual data



SLM Skills Model

- Open Access
- Skill Maps



Data Navigation

- Agentic combination
  - Open-source
- GenAI interface



Skill Identification

Issue: Identify Skills in Data

Example Job Posting

Role

All customer service representatives must provide friendly, confident, and efficient service to all customers.

- Answer telephone calls and emails promptly in Spanish and English (may also be required to answer inquiries via social media on occasion)
- Maintain and extend client base through positive customer interaction
- Collaborate with the customer service team to provide service to a variety of customers from across the globe
- Provide efficient and accurate solutions to customer inquiries and concerns according to company policy
- Update company policy documentation in Spanish and English as needed
- Document all client interactions in Spanish and English into the company database
- If the answer is not within your department, inquire about solutions from respective departments, and provide customers with a timely and effective response

Benefits

- $38,124–41,975 annual salary plus stock options
- Life, health, dental, and vision insurance
- 401(k) program
- Employee Assistance Program
- Paid time off, holiday pay, and paid family leave
- Discounts on airline tickets, merchandise, and other benefits

Qualifications

- Bachelor's degree in PR, Communication and Leadership, Journalism, or equivalent field
- Minimum one year of customer service experience
- Fluent in Spanish and English
- Excellent verbal and written skills in both Spanish and English
- Excellent interpersonal skills
- Strong computer skills
- Strong organizational and analytical skills
- Flexible—must be able to work independently AND with a team daily
- Experience working for airline companies preferred but not required

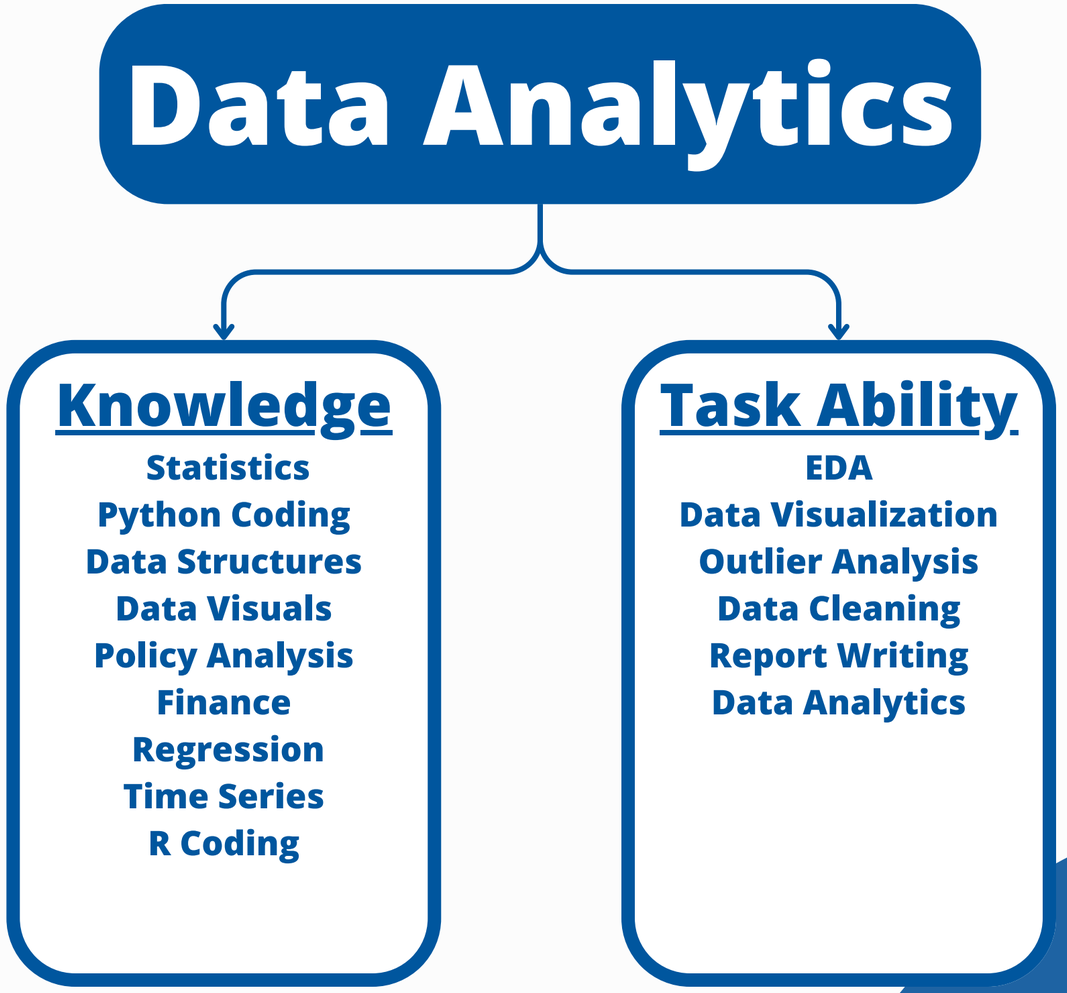
- **Goal:** Label KStAs in human readable documents
- Available data
  - Job Descriptions
  - Syllabi
  - Textbooks (AI summary)
  - Industry Documents
- Expected outcome: Semi-structured data

Where to Start? A Definition of Skill

Definition: Skills are the operationalization of knowledge to perform tasks.

Knowledge: Information learned from experience or transmission
Skills: Synthesis of learning
Task Abilities: Smallest unit of work involved in an economic process

- Key Insights:
  - Skills are cognitive processes
  - Skills are unique to individuals
  - We can't measure skill, only knowledge/task abilities
  - Measurements are point in time only



Education and Labor Market Datasets

Education Data: Open Syllabus

- 27.6 million records - many but not all U.S.
- Syllabus abstracts - not all data captured
- Coverage is inconsistent - combination of web scraping and institutional repositories
- No minimum data quality standards

Job Posting Data: NLx

- Up to 3 million records daily, tens of millions of archived job postings
- Jobs sourced from employer job banks, public workforce agencies
- Good coverage in all U.S. states and territories
- Sometimes described as an “open source Lightcast”

Issue: Identify Skills in Data

Role

All customer service representatives must provide friendly, confident, and efficient service to all customers.

- Answer telephone calls and emails promptly in Spanish and English (may also be required to answer inquiries via social media on occasion)
- Maintain and extend client base through positive customer interaction
- Collaborate with the customer service team to provide service to a variety of customers from across the globe
- Provide efficient and accurate solutions to customer inquiries and concerns according to company policy
- Update company policy documentation in Spanish and English as needed
- Document all client interactions in Spanish and English into the company database
- If the answer is not within your department, inquire about solutions from respective departments, and provide customers with a timely and effective response

Benefits

- $38,124–41,975 annual salary plus stock options
- Life, health, dental, and vision insurance
- 401(k) program
- Employee Assistance Program
- Paid time off, holiday pay, and paid family leave
- Discounts on airline tickets, merchandise, and other benefits

Qualifications

- Bachelor's degree in PR, Communication and Leadership, Journalism, or equivalent field
- Minimum one year of customer service experience
- Fluent in Spanish and English
- Excellent verbal and written skills in both Spanish and English
- Excellent interpersonal skills
- Strong computer skills
- Strong organizational and analytical skills
- Flexible—must be able to work independently AND with a team daily
- Experience working for airline companies preferred but not required

- **Goal:** To create functional definitions for Knowledge, Skills, Task Abilities
- Sample data
  - Job Descriptions
  - Syllabi
  - Textbooks (AI summary)
  - ~100 examples from each
- Hand label a training dataset
  - Multiple coders
  - Intercoder reliability tests
- Train AI to replicate our identifications

Skill Extraction Tool

- Python Package
- Uses trained AI to identify KStAs in text documents
  - Taxonomies
- Open Source
  - Free to use/modify
- AI model agnostic
  - Can be run without AI!
- Can be deployed locally

LAiSER

Leveraging AI for Skills Extraction & Research

GitHub Repository



Linked Data

Issue: Identify Hidden KStAs in Data

Role

All customer service representatives must provide friendly, confident, and efficient service to all customers.

- Answer telephone calls and emails promptly in Spanish and English (may also be required to answer inquiries via social media on occasion)
- Maintain and extend client base through positive customer interaction
- Collaborate with the customer service team to provide service to a variety of customers from across the globe
- Provide efficient and accurate solutions to customer inquiries and concerns according to company policy
- Update company policy documentation in Spanish and English as needed
- Document all client interactions in Spanish and English into the company database
- If the answer is not within your department, inquire about solutions from respective departments, and provide customers with a timely and effective response

Benefits

- $38,124–41,975 annual salary plus stock options
- Life, health, dental, and vision insurance
- 401(k) program
- Employee Assistance Program
- Paid time off, holiday pay, and paid family leave
- Discounts on airline tickets, merchandise, and other benefits

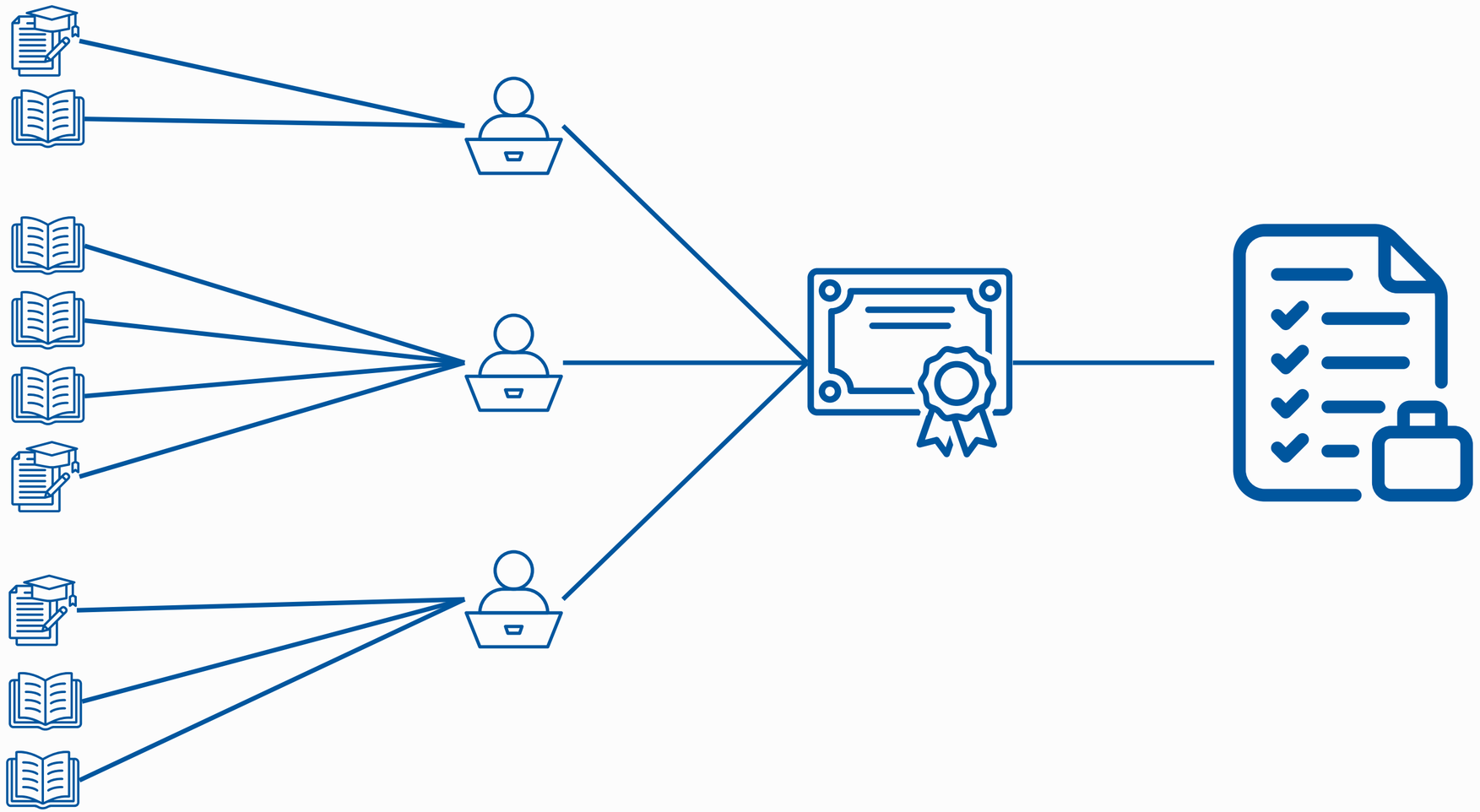
Qualifications

- Bachelor's degree in PR, Communication and Leadership, Journalism, or equivalent field
- Minimum one year of customer service experience
- Fluent in Spanish and English
- Excellent verbal and written skills in both Spanish and English
- Excellent interpersonal skills
- Strong computer skills
- Strong organizational and analytical skills
- Flexible—must be able to work independently AND with a team daily
- Experience working for airline companies preferred but not required

- **Goal:** Identify indirectly referenced KStAs
  - Summary data (job postings etc.) doesn't always list all KStAs
  - Experience and Education are references to larger sets of KStAs
  - Sometimes represented broadly
- **Solution**:
  - Infer missing KStAs using linked data

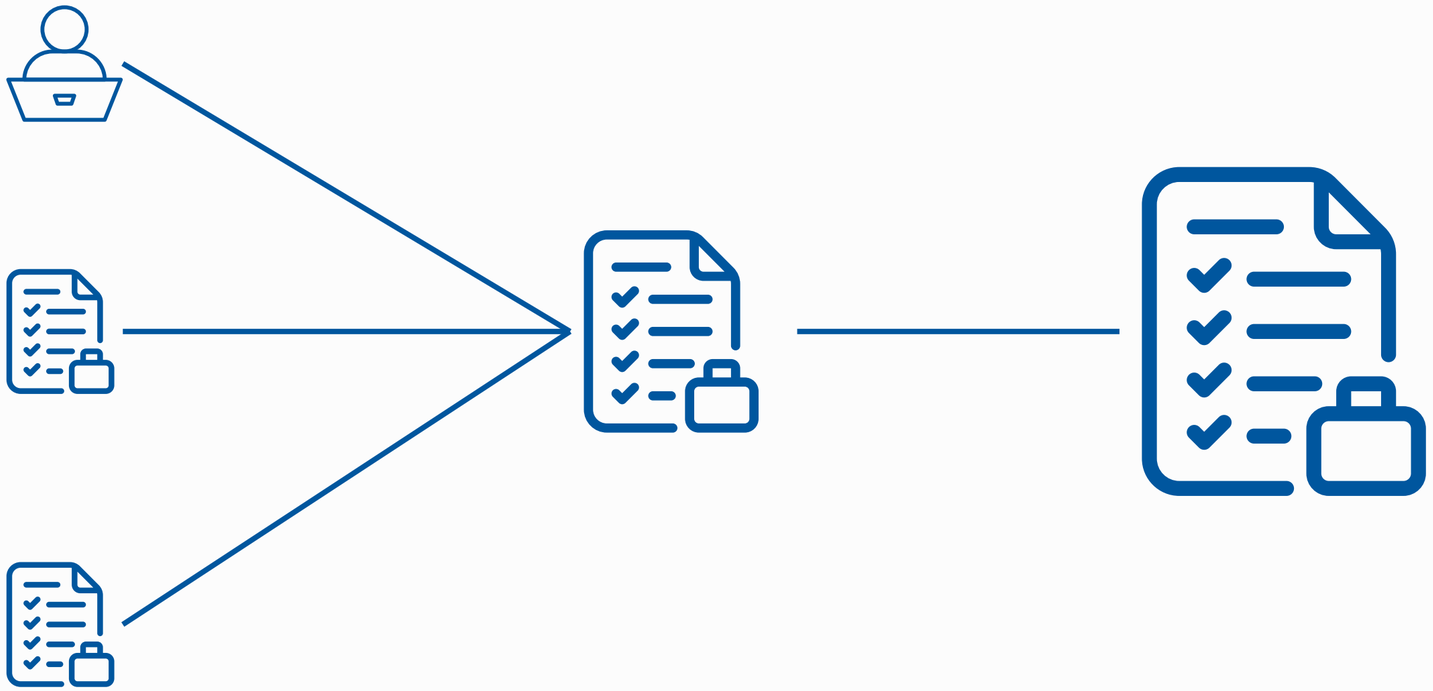
Improving Job Data (Degrees)

Textbooks/ Articles → Courses → Degrees → Job Posting



Improving Job Data (Experience)

Degrees/ Jobs → Experience → Job Posting



Solution: Represent KStAs in Data

Role

All customer service representatives must provide friendly, confident, and efficient service to all customers.

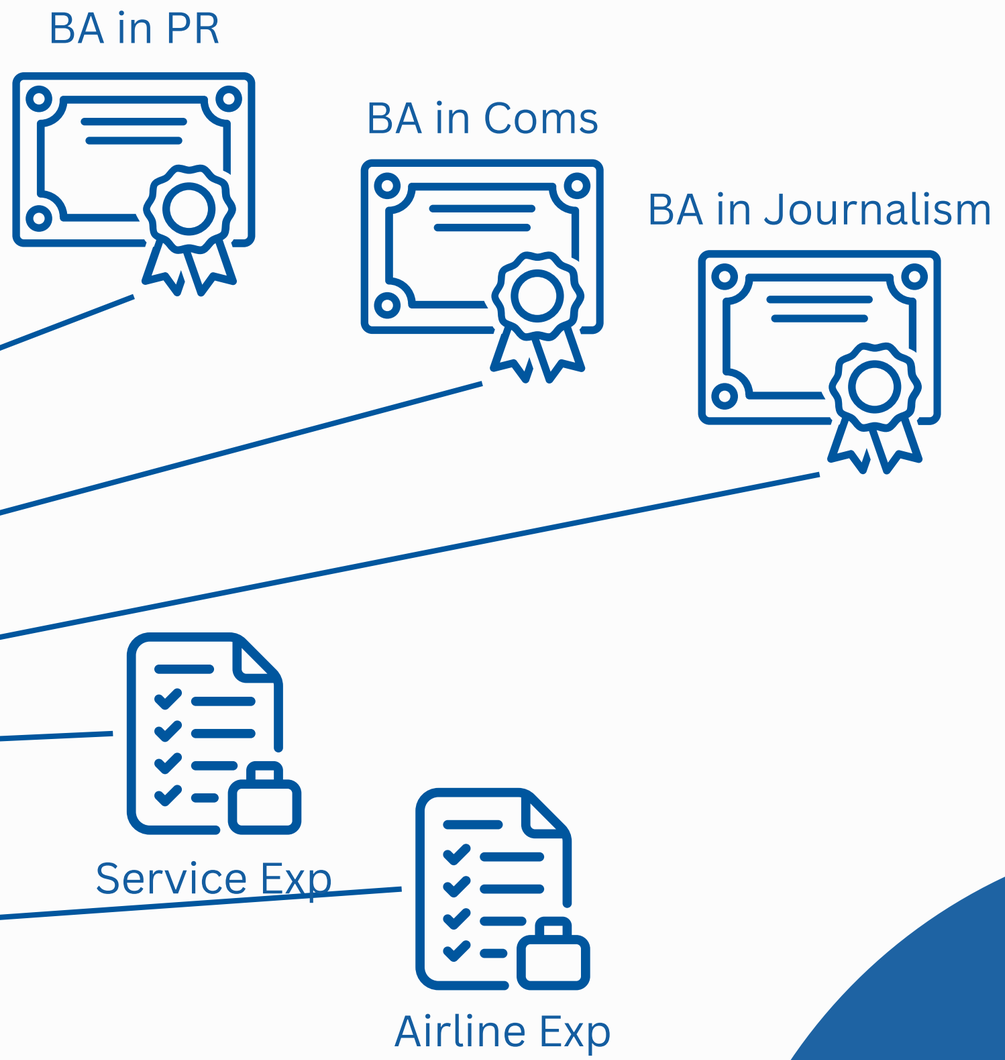
- Answer telephone calls and emails promptly in Spanish and English (may also be required to answer inquiries via social media on occasion)
- Maintain and extend client base through positive customer interaction
- Collaborate with the customer service team to provide service to a variety of customers from across the globe
- Provide efficient and accurate solutions to customer inquiries and concerns according to company policy
- Update company policy documentation in Spanish and English as needed
- Document all client interactions in Spanish and English into the company database
- If the answer is not within your department, inquire about solutions from respective departments, and provide customers with a timely and effective response

Benefits

- $38,124–41,975 annual salary plus stock options
- Life, health, dental, and vision insurance
- 401(k) program
- Employee Assistance Program
- Paid time off, holiday pay, and paid family leave
- Discounts on airline tickets, merchandise, and other benefits

Qualifications

- Bachelor's degree in PR, Communication and Leadership, Journalism, or equivalent field
- Minimum one year of customer service experience
- Fluent in Spanish and English
- Excellent verbal and written skills in both Spanish and English
- Excellent interpersonal skills
- Strong computer skills
- Strong organizational and analytical skills
- Flexible—must be able to work independently AND with a team daily
- Experience working for airline companies preferred but not required



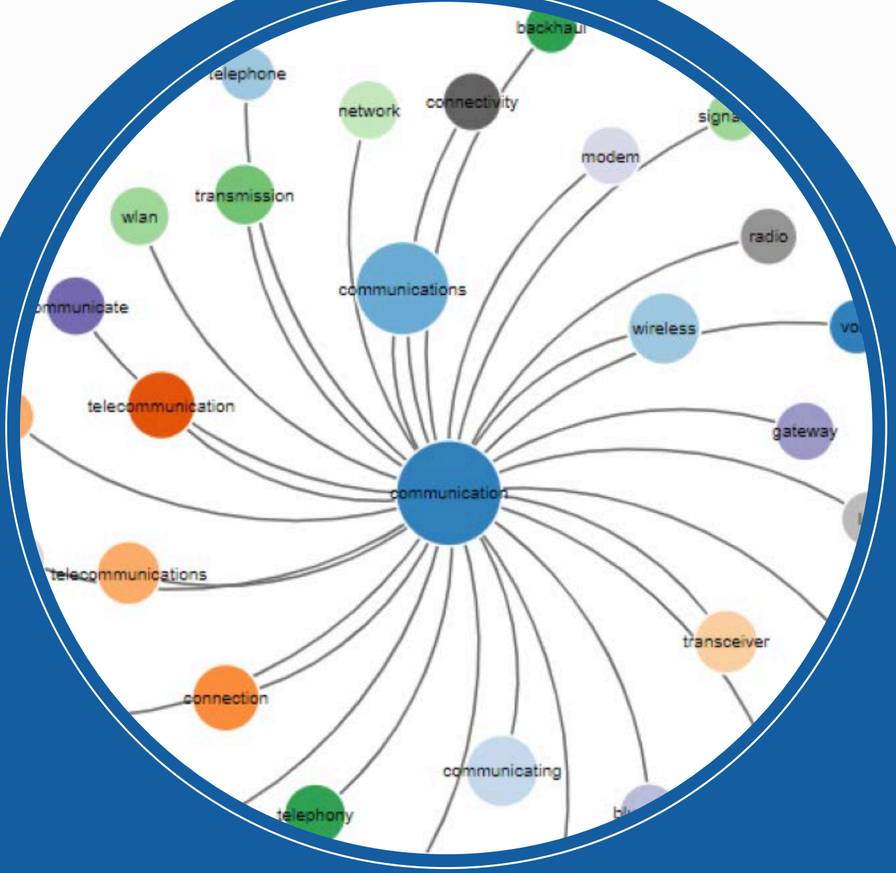


SLM Skill Model

Issue: Lots of Skill Data

- **Solution**: Build a Small Language Model (SLM)
  - Layer different data sources to enhance understanding
    - Condense data using *knowledge graphs*
  - Limited AI built for purpose
  - Supports logical reasoning models
    - Agentic AI
  - Confidence in source data = confidence in inference
  - Open Source

It's the foundation of Agentic Agents



Build Context

Advantages of a Small Language Models



Faster to Develop

Requires a relatively low number of GPUs to develop



More Energy Efficient

Requires fewer computational recourses and thus less energy



Easier to Customize

Simple models can be customized to context on a single GPU



Accessible & Affordable

Can be run on limited computing resources

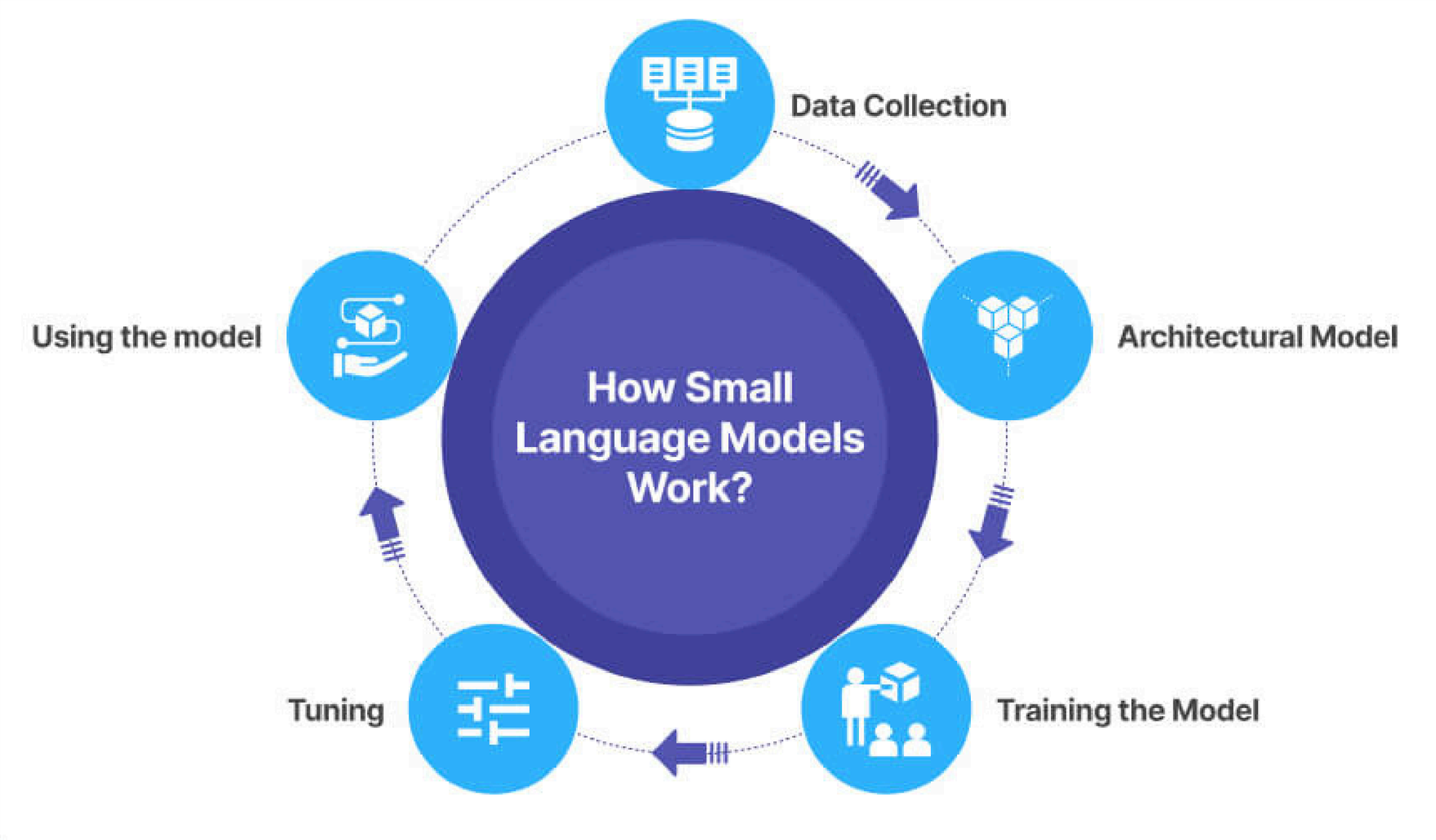


Easier to Understand

More manageable process for tweaking and explaining

V 3.0

Building a SLM



Architectural Model

- KStAs are often described by taxonomies.
- There are many taxonomies, few can be considered complete
- Skills Taxonomies (ESCO, O*Net, etc.)
- Knowledge Taxonomy (Library of Congress Classification)
- Tasks (O*Net task statements)
- We can stack taxonomies to show overlap between definitions

LAiSER

Leveraging AI for Skills Extraction & Research

Taxonomy Landscape Scan

Model Training: Labeled Data

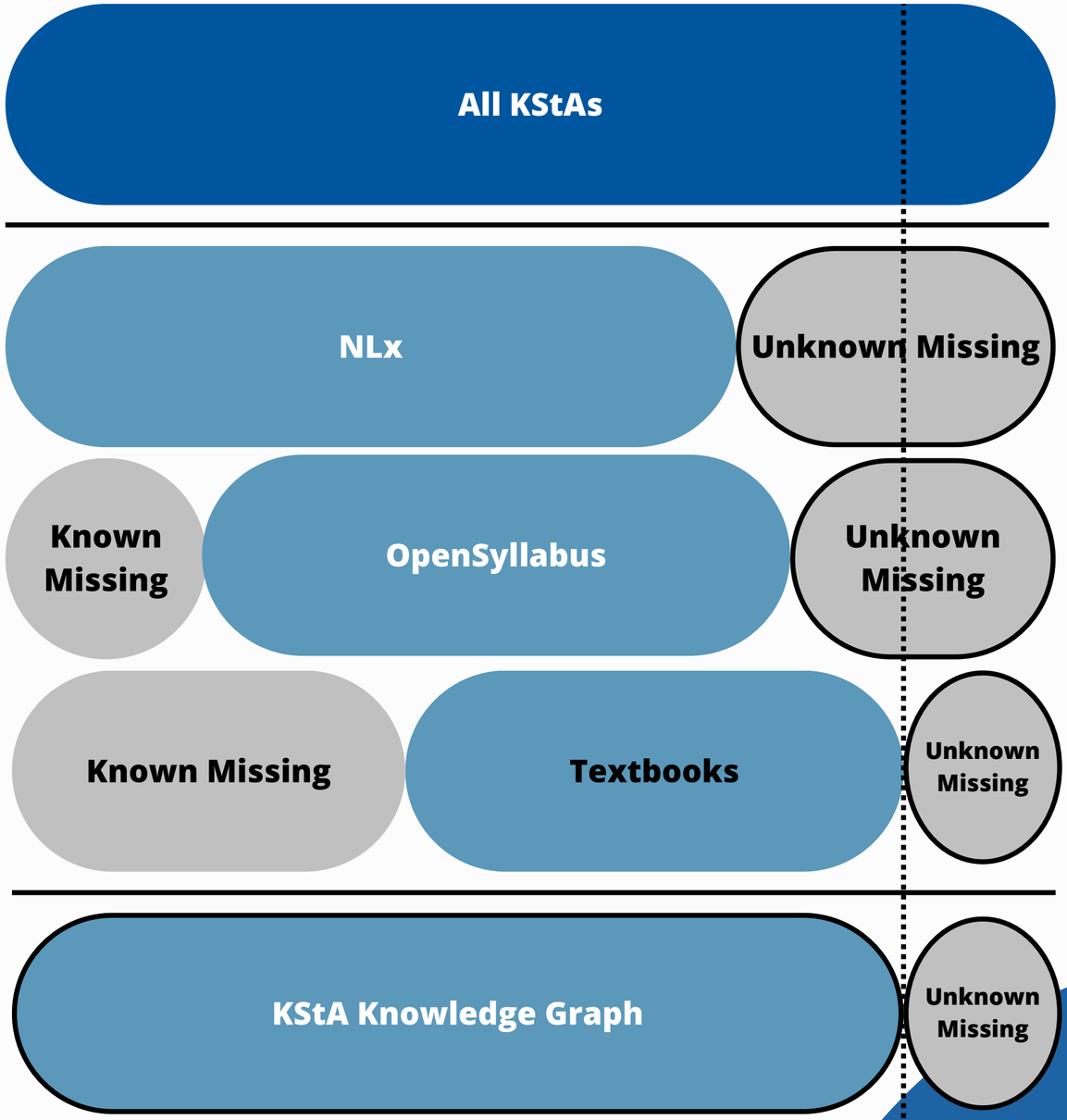
- **Label KStAs in real documents**
  - **NLx (Job Postings)**
  - **OpenSyllabus (Syllabi)**
  - **Textbooks/Journal Articles**
  - **Industry Documents**
- **Compress data into knowledge graphs**
- **Compress knowledge graphs into a layer**

14+

Major Taxonomies Globally

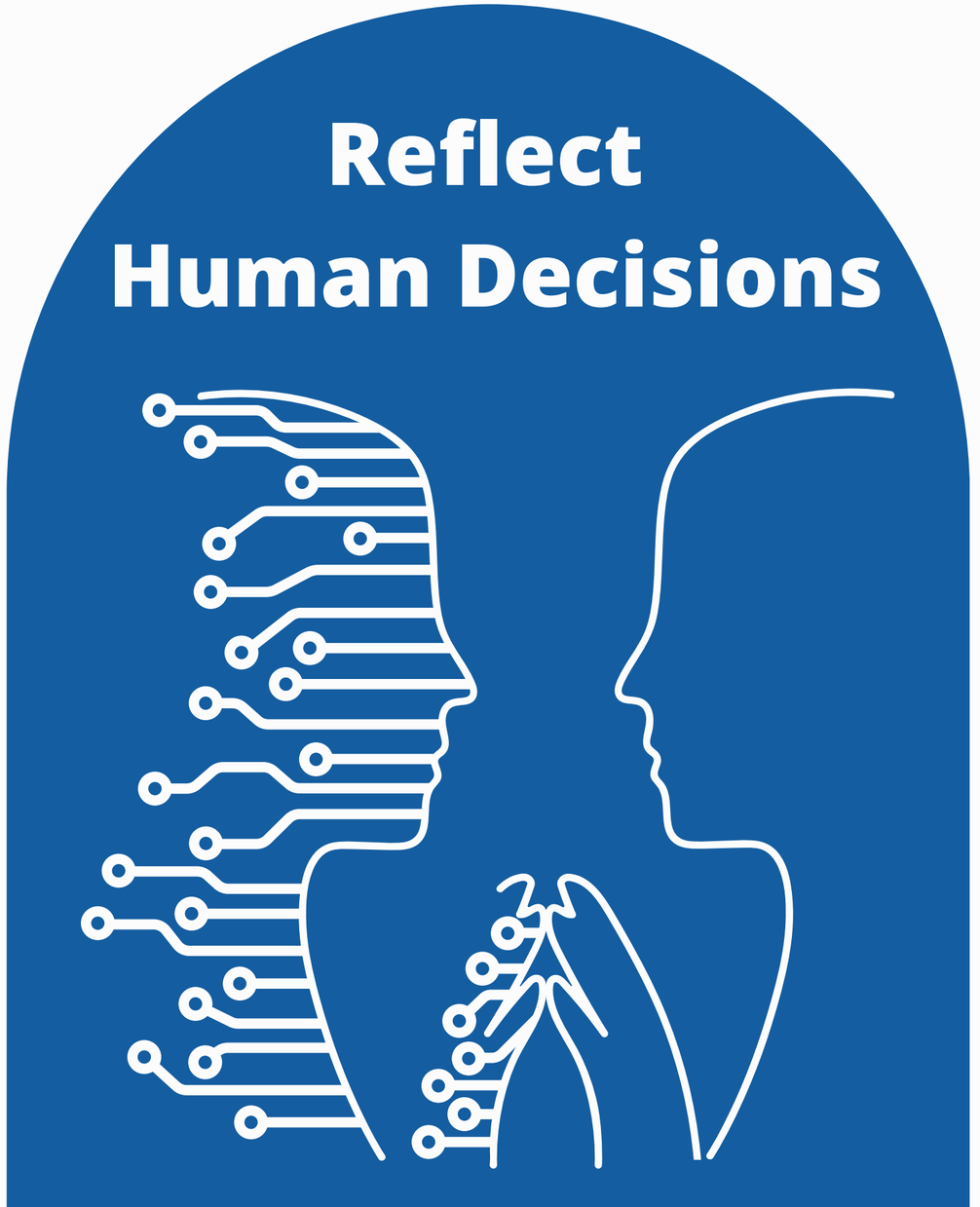
30,000+

Skills Represented Globally



Fine-tuning: Context

- **Issue:** Need to represent human choices
- **Solution**: Observed choices + human in the loop
  - Data Sources
    - Resumes
    - Transcripts
    - Wallet information
  - Partners
    - AACRAO
    - DCC
  - RAG/Feedback loop to adjust



Solution: Hybrid Approach

Role

All customer service representatives must provide friendly, confident, and efficient service to all customers.

- Answer telephone calls and emails promptly in Spanish and English (may also be required to answer inquiries via social media on occasion)
- Maintain and extend client base through positive customer interaction
- Collaborate with the customer service team to provide service to a variety of customers from across the globe
- Provide efficient and accurate solutions to customer inquiries and concerns according to company policy
- Update company policy documentation in Spanish and English as needed
- Document all client interactions in Spanish and English into the company database
- If the answer is not within your department, inquire about solutions from respective departments, and provide customers with a timely and effective response

Benefits

- $38,124–41,975 annual salary plus stock options
- Life, health, dental, and vision insurance
- 401(k) program
- Employee Assistance Program
- Paid time off, holiday pay, and paid family leave
- Discounts on airline tickets, merchandise, and other benefits

Qualifications

- Bachelor's degree in PR, Communication and Leadership, Journalism, or equivalent field
- Minimum one year of customer service experience
- Fluent in Spanish and English
- Excellent verbal and written skills in both Spanish and English
- Excellent interpersonal skills
- Strong computer skills
- Strong organizational and analytical skills
- Flexible—must be able to work independently AND with a team daily
- Experience working for airline companies preferred but not required

- Directly referenced KStAs are identified by the extraction tool
- Indirectly referenced learning and experience are added
  - Done by the SLM based on context
- The combined KStA list is aligned to best represent skills
  - Inferred missing
  - Outlier detection
  - Alignment with taxonomy



Skill Exploration

(Using the Data)

Using the SLM

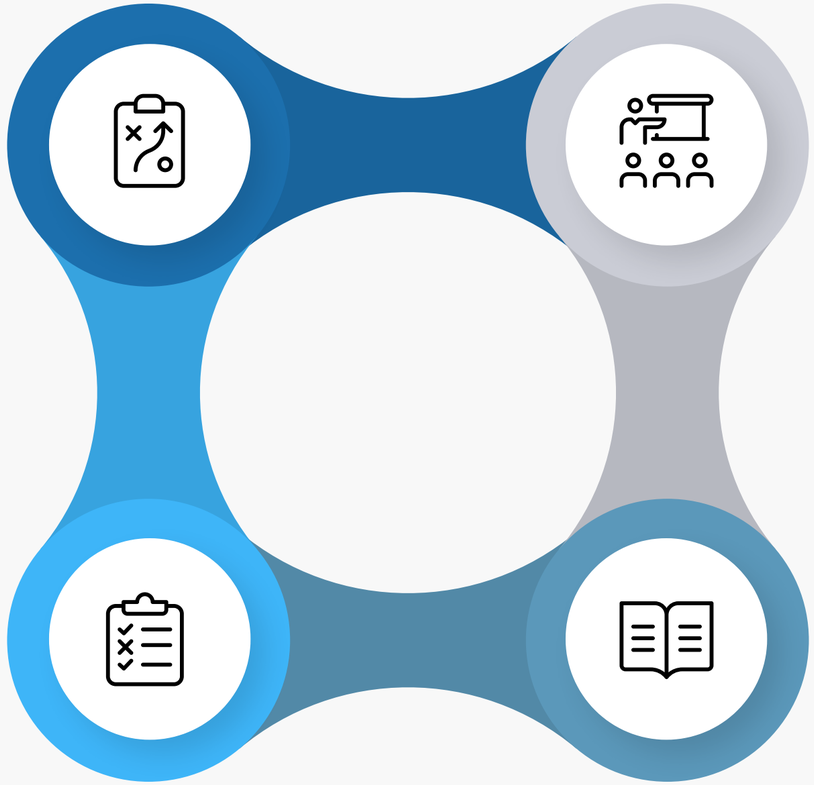
Lots of open-source tools use CLR 2.0/Badge 3.0

Wallets

Digital Credentials Consortium (DCC) Learner Credential Wallet
LearnCard

Course Tools

GW Course Comparison Tool



Job Tools

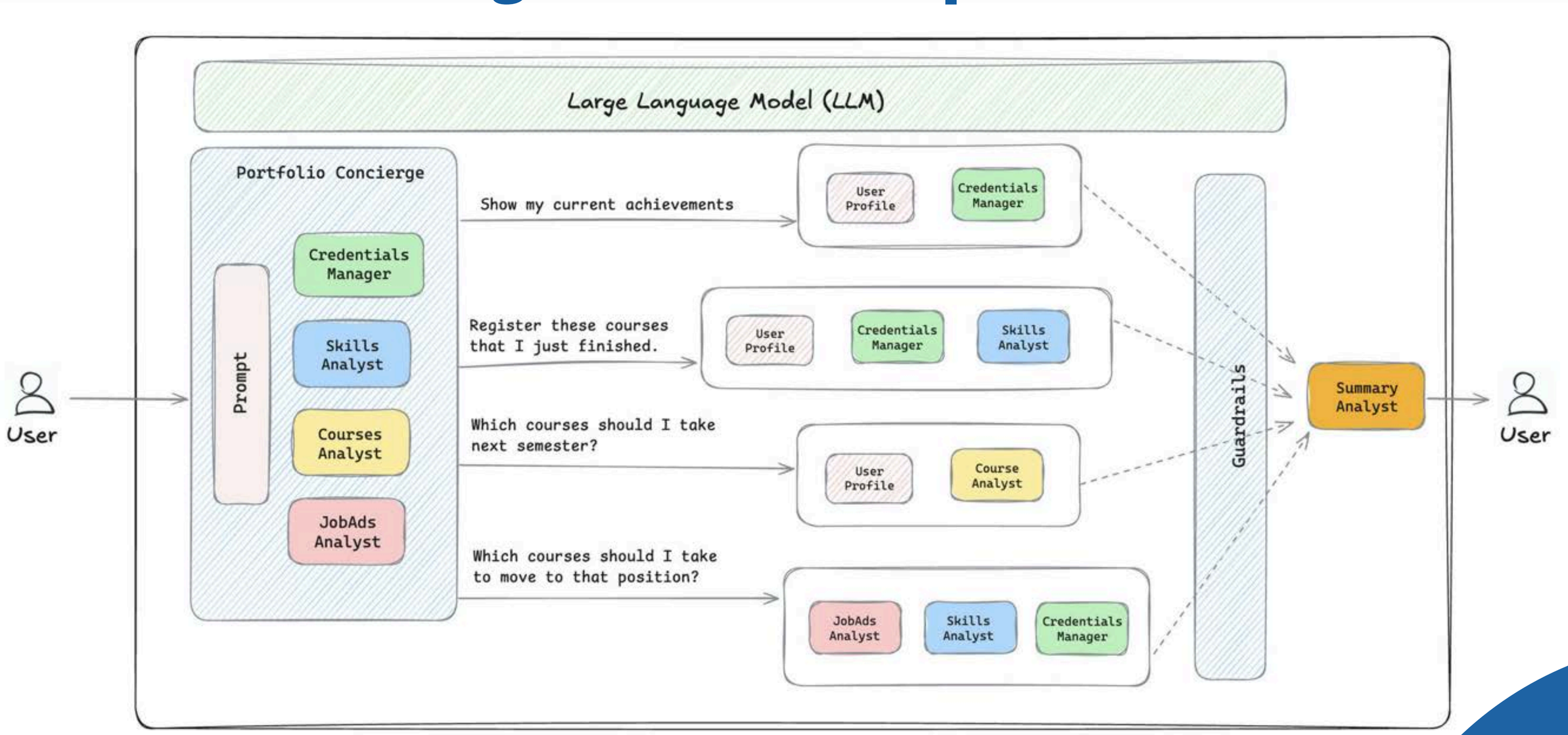
Jobside*

Content Delivery

OpenEdX
Center for Open Education

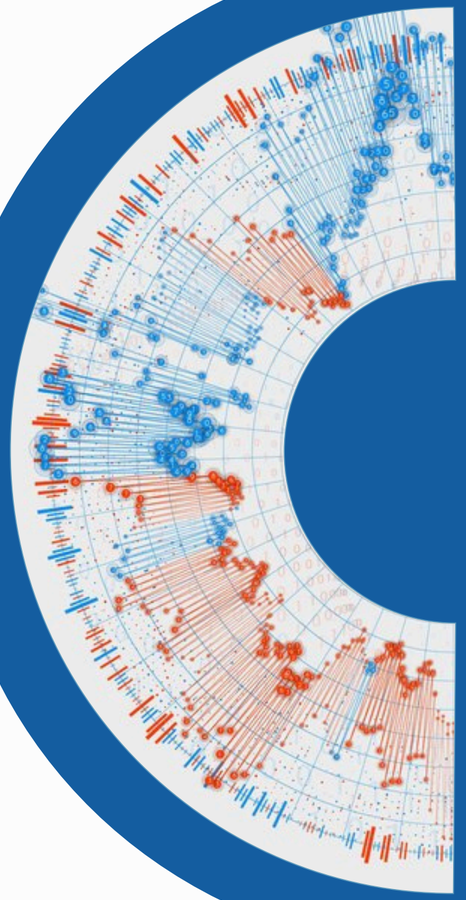
*CLR = Comprehensive Learner Record

Agentic Skills Explorer



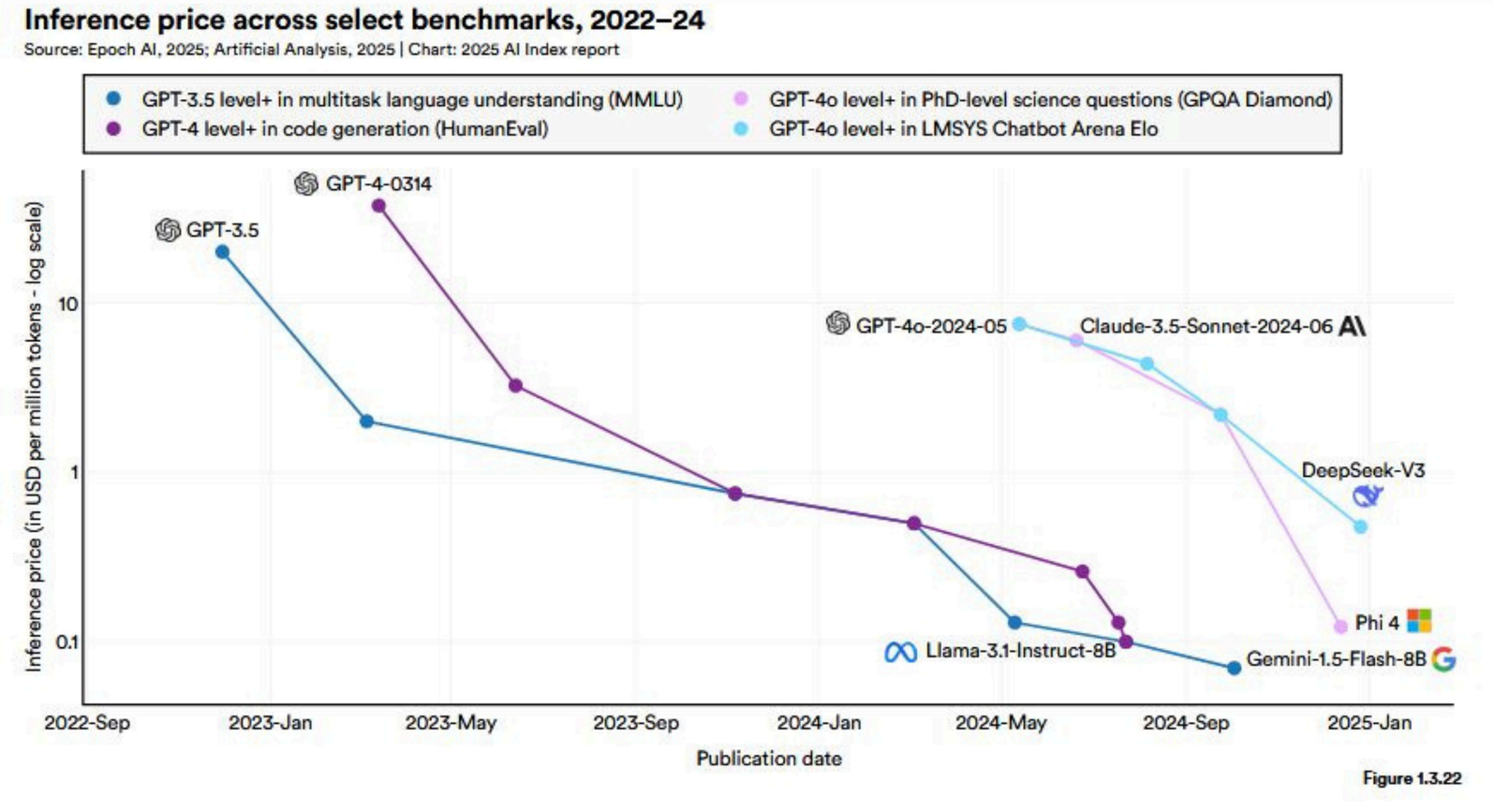
What does AI-driven skill linkage mean for credential evaluation (and evaluators)?

Opportunities for Credential Evaluation



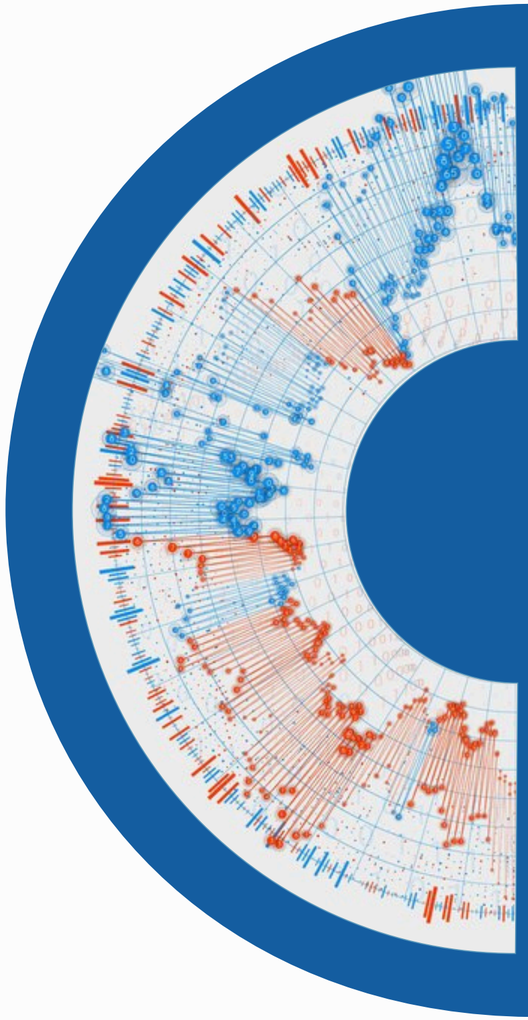
- **Help more learners faster and cheaper**
  - Provide instant insights to potential transferability of skills and credentials across many institutions
  - New ways of visualizing overlap in skill profiles and learning outcomes
  - Costs are likely lower than human evaluation today, and will fall dramatically in the coming years.

AI Cost Curve



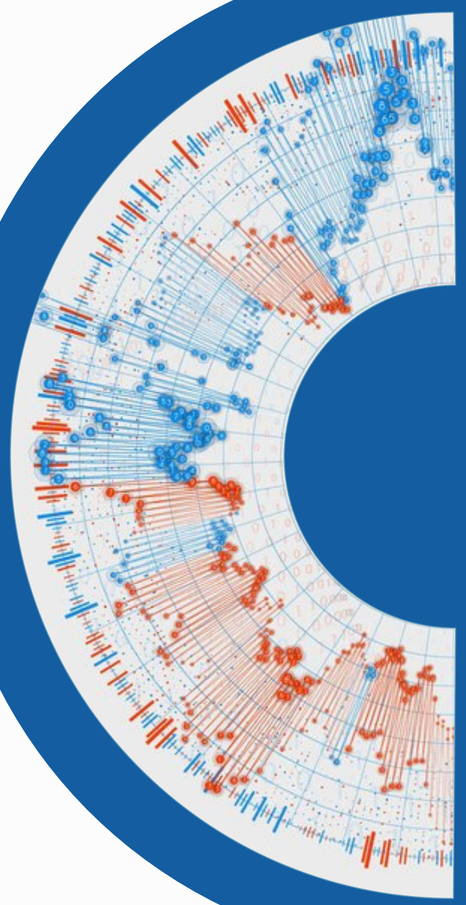
Source: Stanford University 2025 AI Index Report

The Transition to an AI-Guided Future



- Pain points lie ahead for credential evaluation firms
  - Competition from commercial and open source efforts (like ours!) could accelerate pressure to modify business models
    - Higher education could incubate and build trust in noncommercial AI projects
  - Learners could demand new, cheaper options
    - Learners and institutions alike may wonder why they can’t just feed transcripts into a chatbot

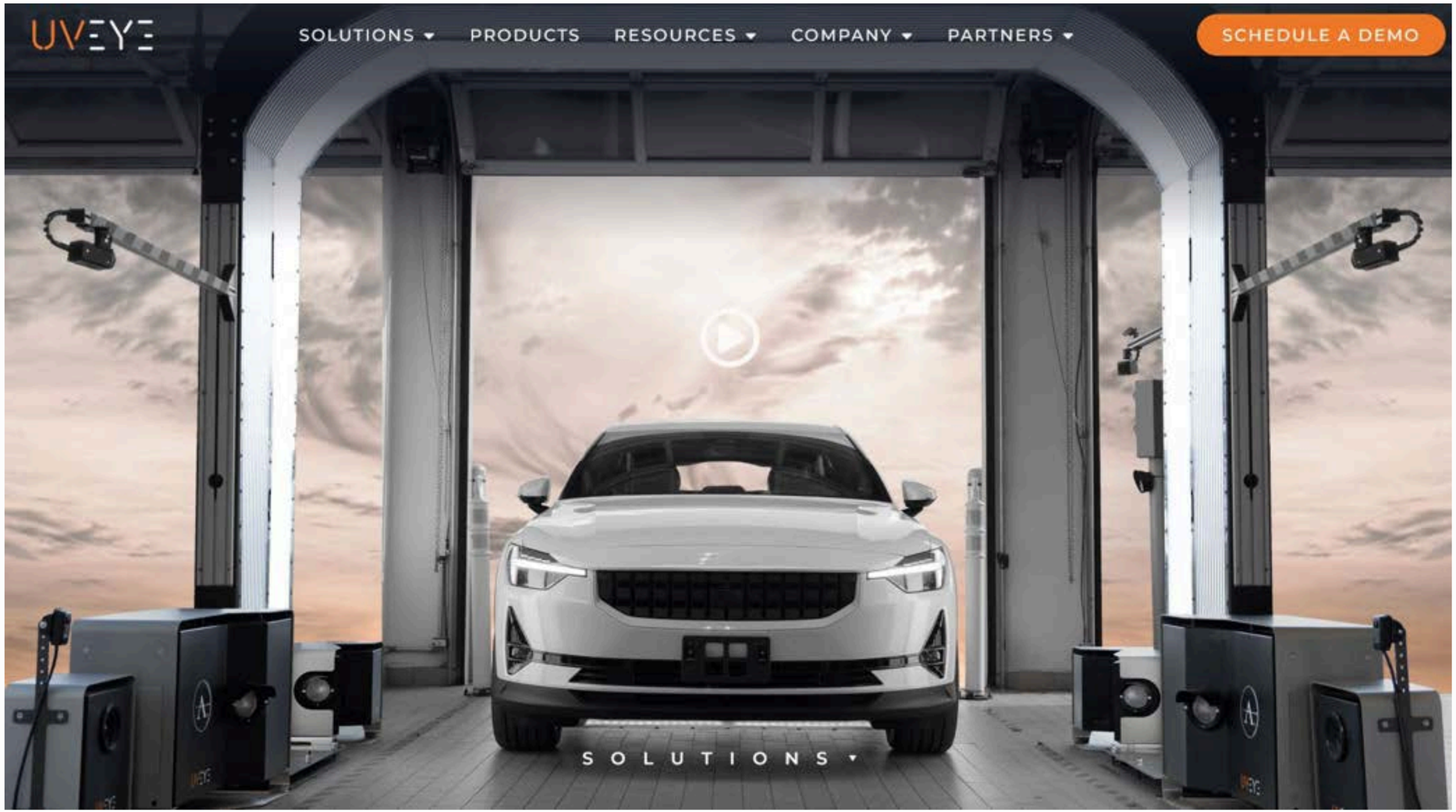
Human Evaluators in an AI World

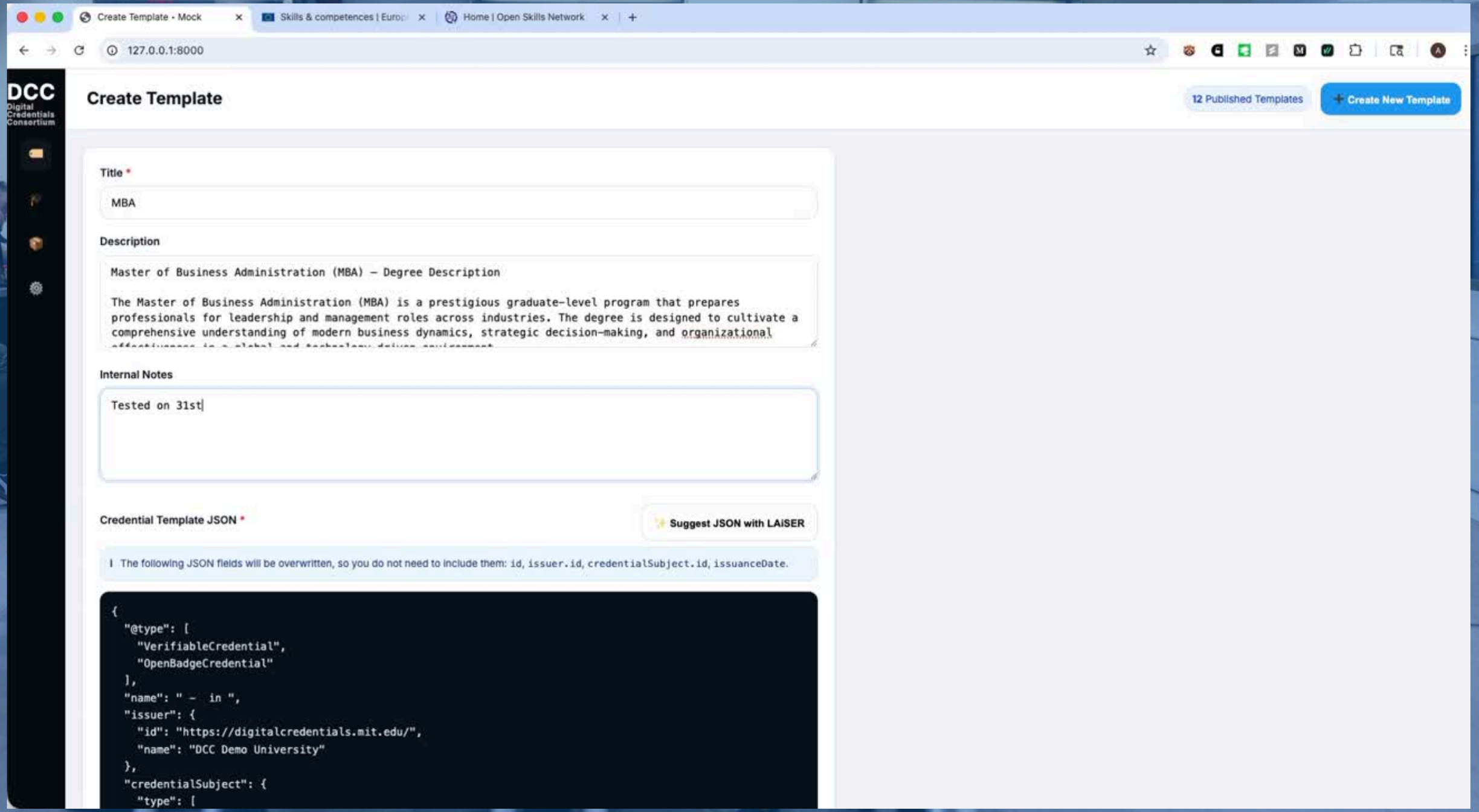


Many roles remain for human evaluators in a world where tools like LAiSER become commonplace!

- Validating AI's work
- More work on edge cases and student appeals
- More focus on credential/document fraud
- Value-added assessments of quality in different countries/systems/institutions
- Integrating with and populating credential wallets.
  - *Imagine a AI-aided, human-approved credential evaluation report that exports directly to a state workforce agency - and the state pays for it!*

More Hope for Humans




Create Template - Mock
Skills & competences | Europ
Home | Open Skills Network
127.0.0.1:8000
DCC
Digital Credentials Consortium
Create Template
12 Published Templates
+ Create New Template
Title *
MBA
Description
Master of Business Administration (MBA) – Degree Description
The Master of Business Administration (MBA) is a prestigious graduate-level program that prepares professionals for leadership and management roles across industries. The degree is designed to cultivate a comprehensive understanding of modern business dynamics, strategic decision-making, and organizational
Internal Notes
Tested on 31st
Credential Template JSON *
✨ Suggest JSON with LAiSER
The following JSON fields will be overwritten, so you do not need to include them: id, issuer.id, credentialSubject.id, issuanceDate.
{
"@type": [
"VerifiableCredential",
"OpenBadgeCredential"
],
"name": " - in ",
"issuer": {
"id": "https://digitalcredentials.mit.edu/",
"name": "DCC Demo University"
},
"credentialSubject": {
"type": [

Our Team

PSCWP
Program on Skills, Credentials & Workforce Policy

GW Institute of Public Policy
THE GEORGE WASHINGTON UNIVERSITY



Kyle Albert
Assoc. Research Professor



Tom Weko
Research Professor



Micah Sanders
Senior Research Associate



Divya Pandove
Senior Scholar



Luis Gonzalez
Post Doc

Our Team



Phanindra Kalaga



Shabnam Sawda



Bharat Khabdekwal



Anusha Umashankar

Anket Patil



Sameer Batra



HaeLee Kim

35+ Graduate Researchers

Research Partners

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

THE NATIONAL ACCREDITATION COMMISSION

NOVA
Northern Virginia Community College

NAU
NORTHERN ARIZONA UNIVERSITY

OPEN SYLLABUS

DC.gov

TEXAS WORKFORCE COMMISSION

Tecnológico de Monterrey

DCC DIGITAL CREDENTIALS CONSORTIUM

scqf
scottish credit and qualifications framework

AACRAO

NLx
National Labor Exchange

THE UNIVERSITY OF TEXAS SYSTEM

THANK YOU

Visit our website for more info

Github Repository for LAiSER

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Issue: Multiple Languages



Solution?:

- Use English build
  - AI to translate to English
    - Problems: AI is expensive, Poor results
- Rebuild in other language
  - Use LAiSER build process on native documents
    - Problems: Time Consuming, Data availability
- Sentence Embeddings (what we use in LAiSER)
  - Designed to be language agnostic
  - Captures context not syntax
  - Syntax can be applied based on language for interpretation
  - Removes much of the “reviewer bias” from loaded terms
  - Aligns with taxonomy of choice