HOW TO ELIMINATE 27 MILLION PROFESSIONALS FROM THE WORKFORCE

DATA COLLECTION-TO-HIRING DECISION PIPELINE

DATA SCRAPING	AI MODEL TRAINING	SHADOW PROFILE GENERATIION	ATS INTEGRATION	ALGORITHMIC RANKING AND FILTERING	HIRING DECISION
From social media, job boards, public databases, online interactions	Data flows into large language models and predictive hiring algorithms	Inferred attributes llike skills, mental health, stability from unverified public data	AI systems embedded into hiring platforms like Workday, SAP Successs-Factors	Candidates scored and ranked automaticallly	AI outputs presented to HR teams and decision-makers

Illuminating the intricate web of data collection, algorithmic processing, and unconsented profiling in employment decisions conducted by private data brokers and automated hiring software vendors. This submission aims to provide clarity on how such systems harm due process, fairness, and the fundamental rights of job seekers, especially those unknowingly blacklisted, filtered out, or misrepresented through shadow profiling.

## I. INTRODUCTION: DATA HAS BECOME DESTINY

In modern employment practices, a significant and often opaque digital infrastructure now governs access to economic opportunity. At the heart of this infrastructure are systems powered by **shadow data**—massive volumes of personal, behavioral, and inferred data mined from the internet, commercial brokers, and passive surveillance.

Shadow data refers to **profiles constructed without the individual's direct input or consent**, aggregated from:
- Web scraping
- Social media APIs
- Academic and résumé databases
- Consumer spending patterns
- Public records and employment databases
- Surveillance technologies (e.g., keystroke tracking, browser fingerprinting)

When these fragmented datasets are synthesized using AI, they create **risk scores, cultural fit models, and predictive personas** that influence whether someone is deemed employable—even before a human ever sees their résumé.

II. THE PLAYERS: WORKDAY & LIGHTCAST (FORMERLY EMSI BURNING GLASS)

Workday's Partnerships for Integration

workday workforce

Lightcast

Workday partners with Lightcast to provide the most comprehensive view of skills in the labor market.

Global workforce data and insights — all in one place

300M+

Workforce ence that Powers

resumes and social networking profiles

Lightcast Data is a combination of labor market information, job postings, global data, skills data, compensation data, + online profiles.

We give data the Lightcast treatment

Our team and tech clean, categorize, and draw meaning from the data. We turn millions of relevant inputs into actionable insights just for you.

Lightcast Manual   Certified Integration workday   HireVue

Certified Integration

The Certified Integration Badge signifies the highest technical validation beyond Workday's core products. It meets customer expectations for top-tier performance, scalability, and reliability, ensuring cost-effective deployment with market-tested assurance.

Customer Testimonial

They've now moved away from using CVs to assess and screen applications. A Workday profile is created which triggers the HireVue chatbot to send out the screening questions. Candidates are then sent a HireVue online assessment and self-schedule an in-person group assessment - again using the HireVue chatbot.

HireVue

HireVue for Workday

By HireVue, Inc.

A leading Workday Certified Talent Experience Platform with over 250 integrations delivering streamlined hiring

Maximize moments that matter with data-driven, timely notifications and actionability across requisitions, driving process alignment and helping recruiters focus on high-value tasks. You can also elevate candidate evaluation through AI-enhanced candidate profiles which provide a single view into key insights for faster, smarter hiring decisions.

Key Features
- Candidate grading
- Smart candidate profiles
- Candidate rediscovery
- Candidate review dashboard
- Global talent search
- Real-time notifications & guided experiences
- Hiring manager experience via MS Teams
- Hiring manager dashboard
- Diversity insights
- Admin insights & reporting

State Sources
- In addition to our federal sources, we use state-level industry projections produced by all 50 states.

Job Postings
- We scrape publicly available data from over 65,000 sources worldwide from job boards and company websites. All of these are from sources that are fully vetted by our 100 person strong team dedicated to data scraping. In 2022 alone we scraped over 348 million job posting advertisements.

Social Profiles
- Profiles are scraped from publicly available information from the web, third-party resume databases and job boards, the recruiting industry, opt-in data from employers and applicant tracking systems, sales and marketing CRM databases, and various consumer/identity databases.

Wage and Hour Division

U.S. Department of Education

National Center for Education Statistics

U.S. Department of Health and Human Services

National Center for Health Statistics

U.S. Postal Service

U.S. Railroad Retirement Board

Oak Ridge National Laboratory

Private sources

State Sources

Job Postings

Social Profiles

[Complete] API - Classification, Postings, Profiles, Some Models - LOT v7 Becomes the Default

The maintenance is now complete. Thanks for your patience.

Schedule Information

Ended on 12 April 2025 at 00:59 EDT, and took 23 hours, 58 minutes, and 56 seconds.

Subscribed Components Affected
- API
- Aggregate Profiles API
- Global Profiles API
- US Profiles API
- Career Pathways API
- Classification API
- DDN Skills API
- Job Postings API
- Canada Job Postings API
- Global Job Postings API

1 / 2

- UK Job Postings API
- US Job Postings API
- Salary Boosting Skills API
- Similarity API

View the full notice →

Innovation Partner workday   Certified

Workday partners with Lightcast to provide the most comprehensive view of skills in the labor market.

As organizations attempt to figure out the best next step in their skills efforts, we're thrilled to offer all Workday Customers that next step, by giving Free Limited Access to Talent Transform. With this offer, Workday Customers will receive:

- Skill profiles and insights into their top roles
- An analysis of where there are gaps and opportunities to improve

---

What's the Complete List of Sources Lightcast Uses in US data?

Updated over a year ago

U.S. Department of Commerce

Bureau of Economic Analysis
- State Personal Income and Employment (SPI)
- Local Area Personal Income and Employment (LPI)
- National Income and Product Accounts (NIPA)
- Annual Input-Output (I-O) Accounts
- Benchmark Input-Output (I-O) Accounts
- GDP by State

U.S. Census Bureau
- American Community Survey (ACS)
- County Business Patterns (CBP)
- ZIP Code Business Patterns (ZBP)
- Nonemployer Statistics (NES)
- Quarterly Workforce Indicators (QWI)
- OnTheMap (OTM)
- TIGER/Line File (with additions by DM Solutions Group)
- Population Estimates
- U.S. National and State Population Projections
- Census 2000 & 2010 Summary Files
- Census of Government – State and Local Government Finances by State
- Journey-to-Work (JtW)
- Consumer Expenditures Survey (CEX)
- Current Population Survey (CPS)
- Commodity Flow Survey (CFS)

A. Lightcast.io's Role in Data Brokerage

Lightcast, a leading labor market analytics firm, boasts access to **over 1 billion job postings** and mining of **thousands of datasets** related to workers' job history, skills, education, and inferred career trajectories. Its tools integrate with HR tech like **Workday**, **SAP SuccessFactors**, and **LinkedIn Recruiter**, sharing labor market signals and even individual behavioral metadata.

According to Lightcast's own website:
> *"We merge billions of historical job postings, online profiles, census, and education data into a unified labor market signal to inform talent acquisition."*

Much of this data originates from **scraping sources like LinkedIn, GitHub, online résumés, and job boards**—regardless of user consent. As described by the company, its **real-time APIs** push these insights directly into platforms like **Workday Skills Cloud** and **HiredScore**, creating dynamic profiles and predictive hiring filters.

How it works

Job Posting Analytics → Enrichment / Meaning / Intelligence → **ACCESS IN ANY FORMAT** — API, SOFTWARE, CONSULTING

Government Sources

Internal Customer Sources

Online Profiles & Resumes

1 Comprehensive Data Collection
We aggregate data from 160,000+ sources, including job postings, government reports, skills databases, compensation records, and professional profiles.

2 Intelligent Data Processing
Our expert team and AI tools clean, categorize, and extract meaningful patterns from these billions of data points, turning raw information into precise, actionable intelligence.

3 Insight Where You Want It, When You Need It
Access our data through software, consulting, APIs, cloud storage, or direct warehouse integration—with ongoing support to help you maximize every insight.

B. Workday's AI Stack and Integration with Brokered Data

Workday's AI-driven hiring tools—including **Skills Cloud**, **HiredScore**, and **Persona Synthesis**—incorporate Lightcast and similar third-party datasets. These platforms advertise their ability to **"analyze over 625 billion data points"**, which include job seeker behavioral metadata, **inferred capabilities**, and **"adjacent" skills not explicitly listed on an applicant's résumé.**

"The system learns from millions of successful employees' careers and maps those trajectories onto new applicants to suggest their future performance." – Workday AI Product Page

The outcome is a **model-based exclusion** system:

Candidates are algorithmically removed from consideration due to perceived skill gaps, inferred instability, or statistical misalignment with corporate culture, even when they've never interacted with the hiring employer directly.

Data Drives Skills-based Talent Optimization

Talent Agencies rely on inaccurate AI-generated profiles to screen candidates. A single misinterpreted data point can lead to disqualification, locking people out of the workforce indefinitely.
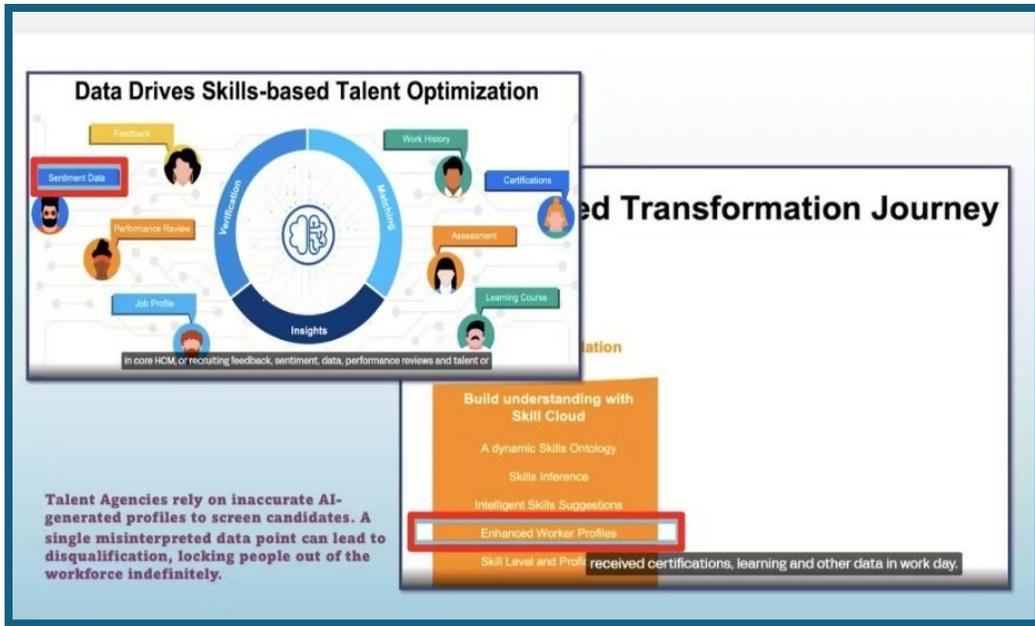
III. RESEARCH: SHADOW DATA'S CONSEQUENCES FOR JOB SEEKERS

1. **The Scale of Harm**
   According to a 2022 *Harvard Business Review* study, more than **27 million qualified job seekers in the U.S. alone** are filtered out by algorithmic hiring systems that rely on automated "fit" scoring. (Bessen et al., 2022). Many are rejected due to **inaccurate or missing data in their shadow profiles**, not based on actual applications or interviews.

2. **Disproportionate Impact on Vulnerable Populations**
   A 2023 paper by the *Center for Democracy & Technology* found that **Black, disabled, formerly incarcerated, and neurodivergent candidates** are disproportionately harmed by automated hiring tools that synthesize data through biased training sets and incomplete profiles.

3. **Absence of Transparency**
   Research from the *AI Now Institute* underscores that job seekers cannot meaningfully access or dispute these profiles. The data—often incorrect or outdated—becomes a "digital rumor" masquerading as a résumé, **barring equal access to work**.

4. **Legal and Ethical Failures**
   Few mechanisms exist to opt out of such profiling. Even under **GDPR or CCPA**, the right to explanation, correction, or deletion is **rarely enforced or understood** by users—especially when employers claim not to "control" the data but only license it from intermediaries.

IV. SOURCES OF SHADOW DATA: THE ENTIRE ECOSYSTEM

Shadow profiling relies on an expansive supply chain of data origins:

Source Type	Examples
Social Media & Résumé Sites	LinkedIn, GitHub, StackOverflow, Facebook, X

Education Records	College registries, MOOC enrollments, certificate APIs
Browser/User Tracking	Cookie tracking, heatmaps, behavioral marketing datasets
Financial Behavior	Credit indicators, purchase patterns, rent/payment histories
Data Brokers	Lightcast, People Data Labs, Acxiom, Oracle BlueKai
Employer Feedback Loops	Internal ATS rejections, resume scoring, turnover tracking

These sources **merge into centralized profiles**, even across employers, leaving candidates unable to reset or contest their digital identity. The **Lack of Consent** is systematic, not incidental.

V. DANGERS: UNVERIFIED DATA & BLACKLISTING
- **False Negatives**: Qualified candidates labeled as "low potential" or "non-strategic."
- **Reputation Traps**: Once labeled as a "job hopper" or "flight risk," future employers are algorithmically discouraged from engaging with the applicant.
- **Lack of Human Review**: Automated sorting means hiring decisions are made long before a recruiter sees the name, increasing the risk of bias and exclusion.

VI. CONCLUSION: CALL FOR TRANSPARENCY AND DATA MINIMIZATION
The entanglement between AI, shadow data, and employment access presents a clear and present danger to the fundamental right to work. The Amicus Curiae urges the Court to recognize:
- The **lack of meaningful consent** in the current system;
- The **discriminatory effect of unverified profiling**;
- The need for **auditable, explainable systems** in employment tech;
- The urgent necessity of **discovery access** for applicants challenging unfair digital evaluations.

Only through judicial scrutiny and regulatory transparency can we prevent a future where workers are nothing more than datasets in a black box.

V-A. THE FULL SCOPE OF SHADOW DATA USED IN EMPLOYMENT SCREENING
Contrary to common assumptions, shadow profiles used in employment contexts draw from a **broad spectrum of highly sensitive personal data**—far beyond traditional résumés or social media accounts.

These datasets include:

Data Category	Examples and Risks
Demographic Data	Race, ethnicity, gender, disability status, sexual orientation, zip code–derived proxy variables
Health-Related Data	Mental health indicators, inferred disabilities, behavioral patterns from wearable devices or web searches
Education & Credentials	MOOC records, professional certifications, degree verification, academic citations
Employment History	Prior job titles, gaps in employment, inferred performance (even from former ATS systems)
Psychographics	Personality profiling, inferred attitudes, emotional tone (from emails, social posts, behavioral data)
Web Behavior & Metadata	Time spent on job sites, online application abandonment, IP/geolocation metadata
Financial Data	Credit risk proxies, purchasing habits, transactional histories, inferred socioeconomic class
Reputation Signals	Ratings on gig economy platforms, community engagement, blacklist status from prior employer systems

Much of this data is **aggregated without consent** and fed into hiring filters that evaluate a candidate's supposed **"fit,"** **"stability," or "risk" based on experimental probabilistic models**—not human judgment.

This data is routinely shared between brokers like **Lightcast.io**, **People Data Labs**, **Acxiom**, and platforms like **Workday**, **SAP SuccessFactors**, and **Oracle HCM Cloud**, where it is **used to construct AI-based hiring decisions without notification or transparency** to the job seeker.
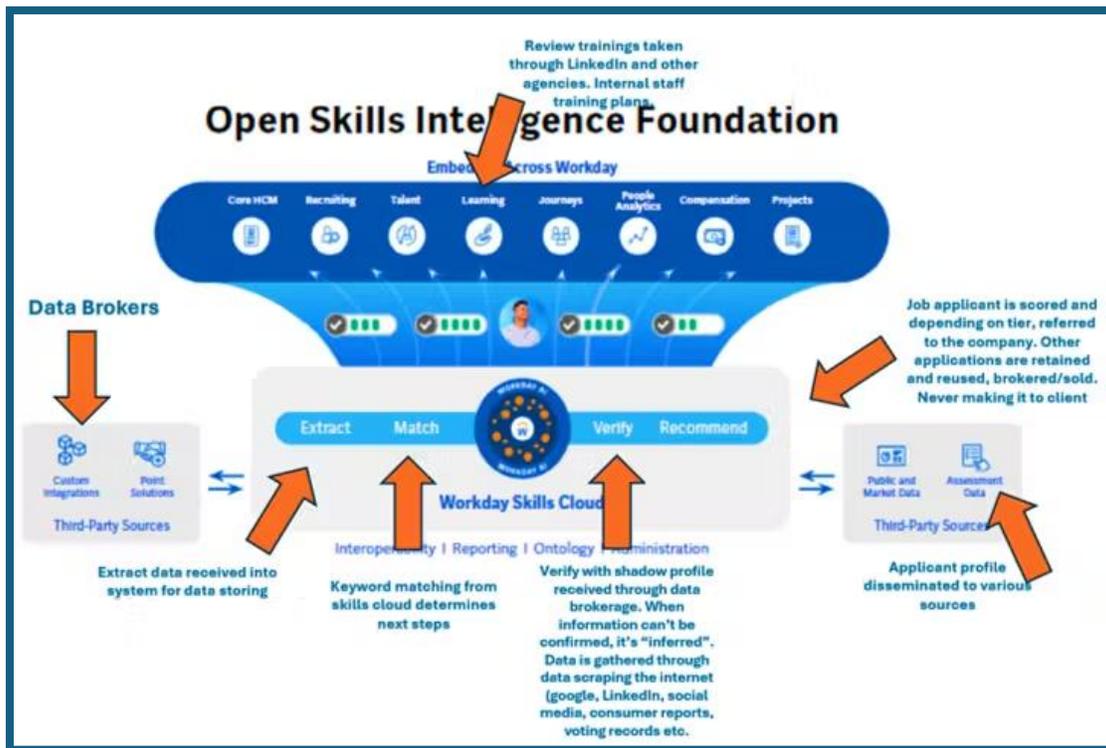
V-B. HOW LARGE LANGUAGE MODELS ARE EMBEDDED INTO ENTERPRISE HR SYSTEMS

Large language models (LLMs), including those developed using **transformer-based architectures like GPT, BERT, and T5**, are now deeply embedded within recruitment, HR, and hiring ecosystems. They are not standalone tools; they are **subsumed into platforms** via APIs and SDKs, then **licensed and distributed at scale**.

1. Embedded LLMs in Platforms like Workday

Workday integrates **LLMs to enhance resume parsing, job description generation, candidate Q&A, and sentiment analysis**. These models claim:
- Trained on **millions of historical résumés, job postings, and interview transcriptions**
- Fine-tuned with **employer-specific talent data** (including rejected applicants, promotion histories, etc.)
- Capable of **synthesizing candidate profiles** using inferred data and embedding soft indicators like tone, personality, or cultural fit

LLMs are also used to:

- Match résumés to job descriptions using **semantic similarity scores**
- Identify "adjacent" skills not explicitly listed
- Predict potential employee trajectories and attrition risks
- "Explain" why a candidate might be a poor fit—based on linguistic patterns and metadata

2. Distribution Across Employers

Workday markets these capabilities through its **Skills Cloud and HiredScore integrations**, which are built atop LLM infrastructure and available to:

- **Over 10,000 enterprise clients globally**
- Including **Fortune 500 companies**, government contractors, healthcare systems, and educational institutions
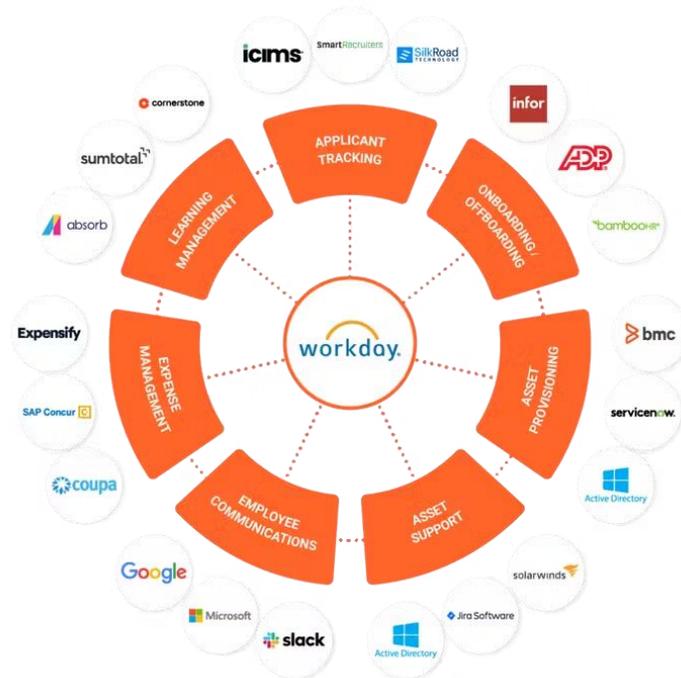
Clients may not be aware they are leveraging a **centralized profiling infrastructure**, as the AI features are abstracted into user-friendly dashboards and dashboards for recruiters. However, what they're using are **multi-layered LLM models trained on billions of datapoints**, many of which come from public surveillance, scraped platforms, or third-party data brokers.

Notably, a Workday presentation in 2023 claimed it had **access to and analyzed over 625 billion data points**, used to "build contextual understanding" of job seekers beyond their explicit profiles.

**V-C. LEGAL AND ETHICAL FAILURES IN TRANSPARENCY AND DATA PROVENANCE**

For legal matters, it is advisable to **consult a qualified lawyer** who can provide specific guidance relevant to individual circumstances. Lawyers can assist in various areas such as criminal law, civil rights, family law, corporate law, and intellectual property, among others. It is important to choose a lawyer who specializes in the relevant area of law to ensure the best possible advice and representation. When consulting a lawyer, individuals should prepare all relevant documents and questions in advance to make the most of the consultation.

V-D. IMMEDIATE RISKS TO ECONOMIC PARTICIPATION

The convergence of shadow data and LLM-powered AI in hiring is not a future risk—it's an ongoing **civil rights crisis**. Tens of millions of Americans may be **pre-emptively filtered from economic opportunity** due to misrepresented, outdated, or simply incorrect algorithmic interpretations of who they are.

- False negatives can prevent marginalized populations from re-entering the workforce.
- Candidates with medical conditions may be excluded through inferred disability risk scoring.
- Blacklisted reputations—perhaps from a gig platform or social post—may follow them indefinitely.
- No legal system ensures redress at scale.
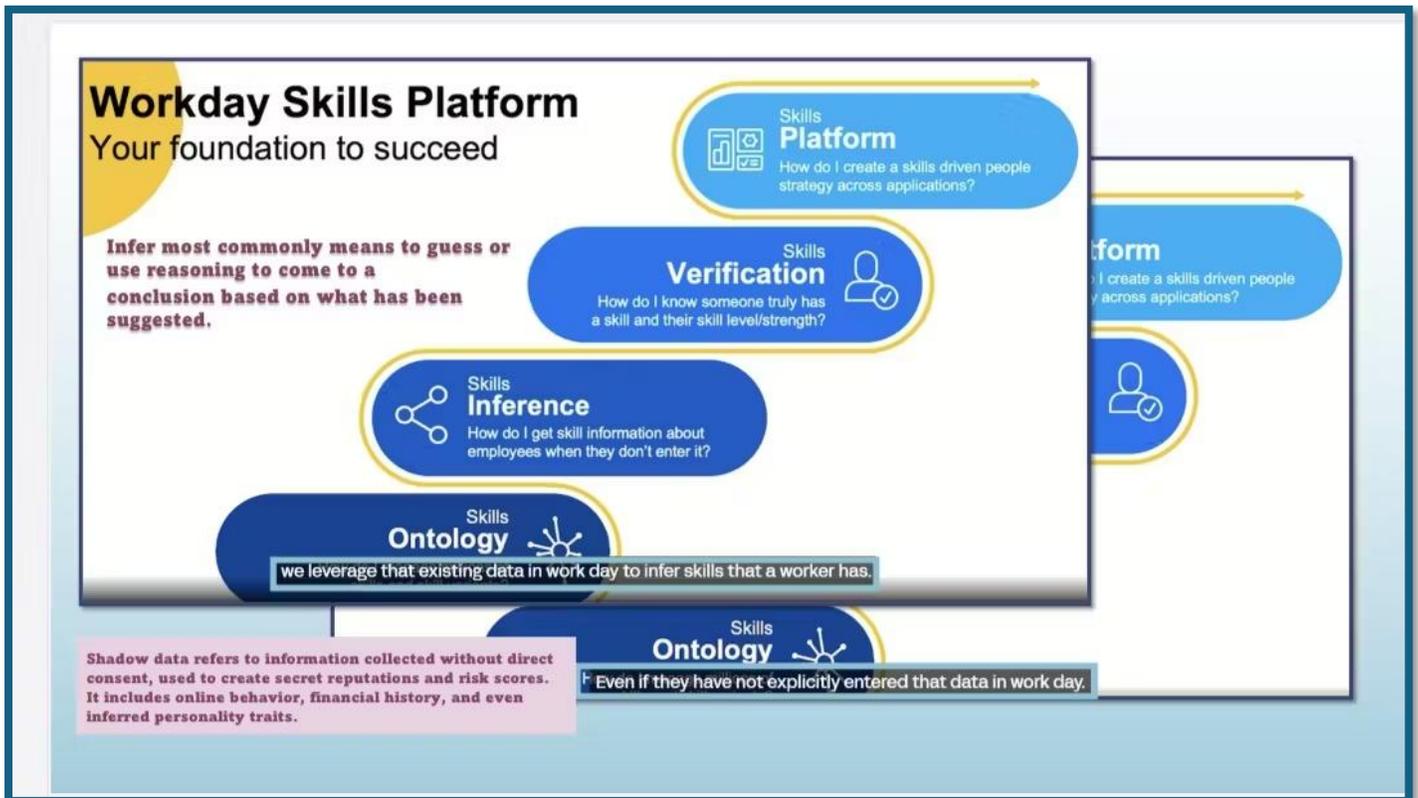
Adjacent Skills

Adjacent skills refer to **related or transferable skills** that are **not explicitly listed** on a resume, application, or profile but are **logically connected** or inferred based on a person's past experiences, job titles, industries, or training.

Definition of Adjacent Skills:

Adjacent skills are capabilities inferred through AI or logical association, based on an individual's existing or past skills, experiences, or job roles—even though those skills were not directly listed or claimed by the person.

Examples include:
- Inferring Excel or attention to detail from "Data Entry Clerk"
- Predicting CRM software familiarity for a "Customer Service Rep"
- Assigning project management skills based on "Team Lead" or "Developer" titles

How Adjacent Skills Are Used by AI Systems AI hiring platforms, such as Workday Skills Cloud, HiredScore, and Persona Synthesis, routinely:
- Use large language models (LLMs) and natural language processing to predict likely adjacent skills
- Generate "shadow profiles" from brokered and scraped data
- Evaluate job seekers based on both listed and inferred attributes
- Include predictive traits like "volatility," "gap risk," or "culture fit" using adjacent or behavioral signals

Legal and Ethical Risks
- Inaccuracy and Misrepresentation: Adjacent skills may not reflect true competencies.
- Lack of Consent: Job seekers rarely consent to the inference of skills they did not report.
- Disparate Impact: Inferred traits may replicate or worsen racial, gender, or age bias.
- FCRA Violation: Inferred skills used in hiring decisions may trigger FCRA obligations if treated as consumer reports.
- Title VII Discrimination: Disparate impact may occur if inferred data correlates with protected characteristics.

https://www.linkedin.com/pulse/how-eliminate-27-million-professionals-from-workforce-sheilah-i600c