Surviving and Professionally Thriving in the Face of AI and Machine Learning

2024 SCASA Meeting

Chong Ho Alex Yu, Ph.D., D. Phil.
Professor and Program Director of Data Science

HAWAIʻI
PACIFIC
UNIVERSITY

Neil Postman, American author and educator

“The effects of technology are always unpredictable. But they are not always inevitable.”

Introduction

- The rapid advancement of AI presents humanity with a complex landscape of opportunities and challenges, characterized by profound uncertainty.
- As we stand on the cusp of what many consider a technological revolution, it's crucial to approach AI development with both **optimism** and **caution**.

Job Displacement

“This process of Creative Destruction is the essential fact about capitalism.”

- Joseph Schumpeter. (1942) *Capitalism, Socialism and Democracy*.

WORLD ECONOMIC FORUM

Future of Jobs Report 2023

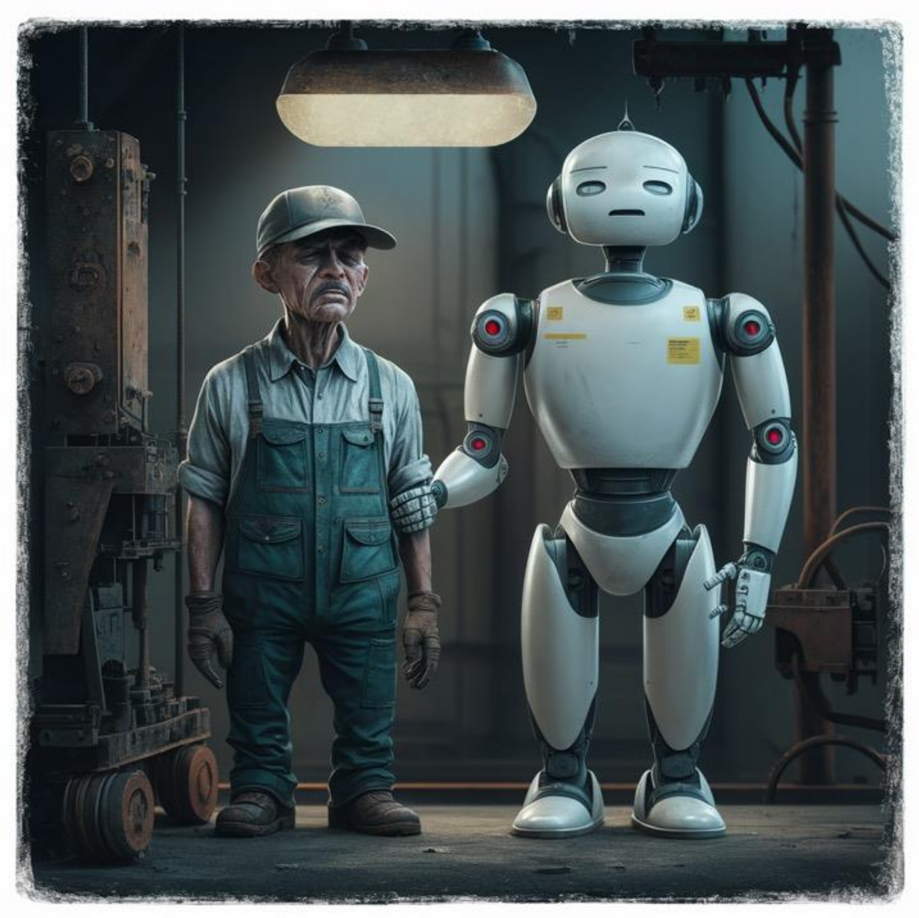
INSIGHT REPORT

MAY 2023

Job Creation and Jobs at Risk

- According to the **World Economic Forum's Future of Jobs Report** 2023, artificial intelligence (AI) is expected to have a significant impact on jobs in the coming years.
- 69 million new jobs are expected to be created in the next five years, driven by new technologies and the green transition.
- This will be offset by 83 million jobs being put at risk due to economic pressures and automation.

Job Creation and Jobs at Risk



- According to Worldmetrics.org, up to 800 million global workers could be replaced by AI by 2030, with AI potentially eliminating around 15% of current jobs worldwide by 2025.
- While AI may displace 75 million jobs by 2030, it is also expected to create 133 million new ones, indicating a net positive effect on job creation in some scenarios.

Job Creation and Jobs at Risk

- **Job loss by 2030: Regional variations**
  - **United States**: Approximately 73 million jobs could be at risk of automation by 2030, with 44% of U.S. jobs considered at high risk of automation.
  - **China**: Up to 236 million jobs could be displaced by automation by 2030.
  - **European Union**: 34% of all jobs.
  - **India**: 70% of all jobs.

Jobs at Risk

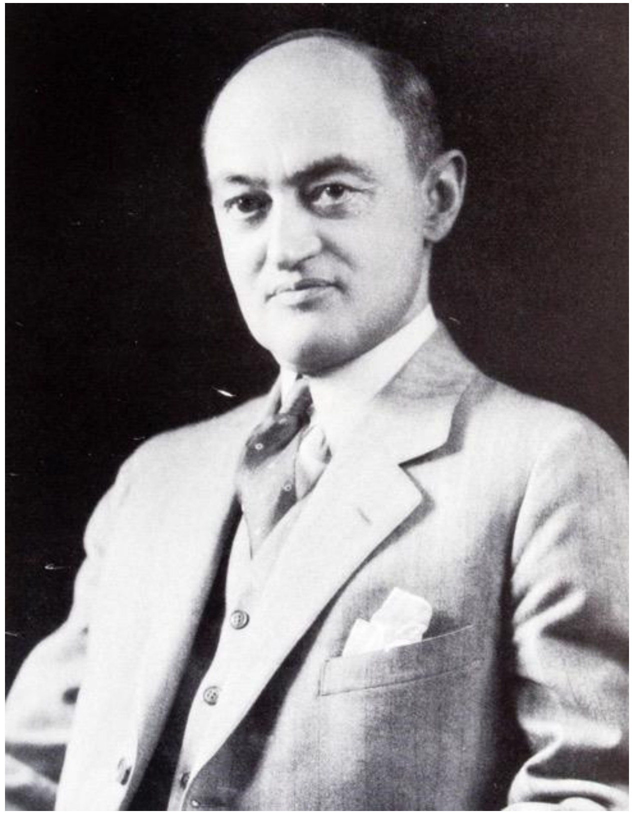
- **Entry-Level Programming**: While complex programming still requires human creativity, AI tools like ChatGPT can already write code for more straightforward tasks.
- **Research and Analysis Positions:** AI's ability to process large volumes of data, detect patterns, and organize findings makes it well-suited for research-centric roles. Market research analysts and financial analysts may see their jobs partially automated.
- **Video Production**: The demand for conventional video production and editing will be greatly reduced while text-to-video tools can make video faster and cheaper.

Jobs at Risk

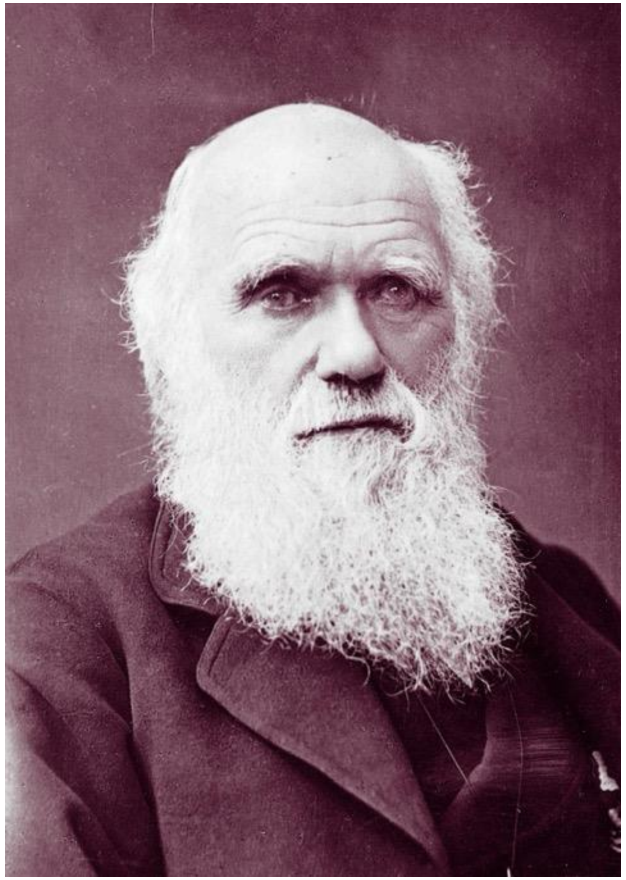
- Contrary to previous waves of automation that primarily affected blue-collar work, AI is expected to have a significant impact on white-collar and higher-paying jobs. This is because AI is designed to mimic cognitive functions, potentially affecting jobs that require analysis, decision-making, and other intellectual tasks.
- While these jobs may be reduced or replaced, AI is also expected to create new jobs and enhance productivity in many fields. The overall impact will likely involve a shift in the types of skills and roles that are in demand rather than a simple reduction in total employment.

Creative Destruction

- Introduced by Austrian economist Joseph Schumpeter.
- It describes the process by which innovation and technological advancement drive economic growth, but in doing so, they disrupt and dismantle existing industries, products, or business models.
- This process is essential for **long-term** economic progress, as it clears the way for new ideas, businesses, and opportunities to flourish, even though it can cause short-term disruption and loss.



Creative Destruction



- Schumpeter drew parallels between his concept of creative destruction and evolutionary theory, particularly the ideas associated with Charles Darwin.
- Innovation acts as a driving force for change, much like natural selection in biological evolution. Just as in Darwinian evolution, where only the fittest organisms survive and thrive.
- So, what kind of jobs will not survive?

AI-powered Web Scraping

HAWAIʻI PACIFIC UNIVERSITY

What is Web Scraping?

- Web scraping is an automated process of extracting data from websites.
- Instead of manually copying and pasting information, web scraping uses software (like Python’s BeautifulSoup, Scrapy, or Selenium) to navigate web pages, identify specific data, and gather it systematically.
- Web scraping mimics human browsing behavior, but it automates the data collection to retrieve large amounts of data more efficiently.

Why Do We Need Web Scraping?

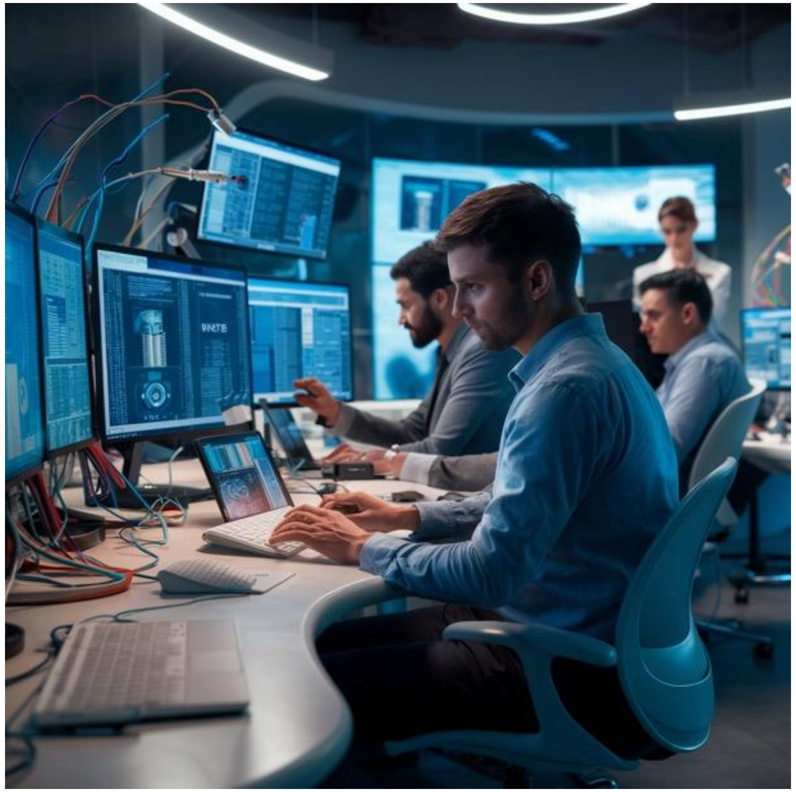
- **Access to Unstructured Data**: Much of the web's information is in an unstructured format, such as HTML or plain text, making it hard to analyze without a structured data format (like CSV or a database).
- **Data Collection Efficiency**: Web scraping automates what would otherwise be a slow and tedious process, enabling quick gathering of large datasets.

Why Do We Need Web Scraping?

- **Dynamic Data**: Some websites don’t offer direct data downloads or APIs (Application Programming Interfaces), making scraping the only viable way to extract updated or real-time data.
- **Cost-Effective**: For companies or researchers, scraping is a cost-effective way to collect large volumes of data without needing an official data provider.

Who Benefits from Web Scraping?

- **Businesses and E-commerce**: Scraping competitor prices, customer reviews, and product descriptions allows companies to track market trends, optimize pricing strategies, and improve product offerings. For instance, e-commerce sites can scrape competitor websites to adjust their own prices and promotions accordingly.



Who Benefits from Web Scraping?

- **Researchers and Academics**: Scraping data from news articles, social media, or government websites helps researchers collect information on public opinion, economic indicators, and scientific research trends. For example, a social scientist could use scraping to analyze sentiment in tweets about a political event.

Who Benefits from Web Scraping?

- **Data Journalists**: Journalists often need large amounts of real-time or historical data to tell data-driven stories. Web scraping can help them gather data for investigative pieces. For example, a journalist might scrape court records to investigate patterns in legal cases.
- **Marketers**: For digital marketing, scraping allows for trend analysis, brand monitoring, and sentiment analysis on social media platforms. Marketers can also analyze keywords and other metadata from competitors’ websites to optimize their SEO strategies.

Who Benefits from Web Scraping?

- **Real Estate Analysts**: Real estate companies can use web scraping to track property prices, rental listings, and housing trends. By analyzing this data, they can make more informed investment decisions and offer valuable insights to their clients.
- **Travel and Hospitality**: Travel agencies and booking platforms can scrape data from airline websites, hotel listings, and travel reviews to monitor prices and trends. This enables them to offer competitive pricing, exclusive deals, or better-informed recommendations to travelers.

Programming vs. Web Scraping Companies

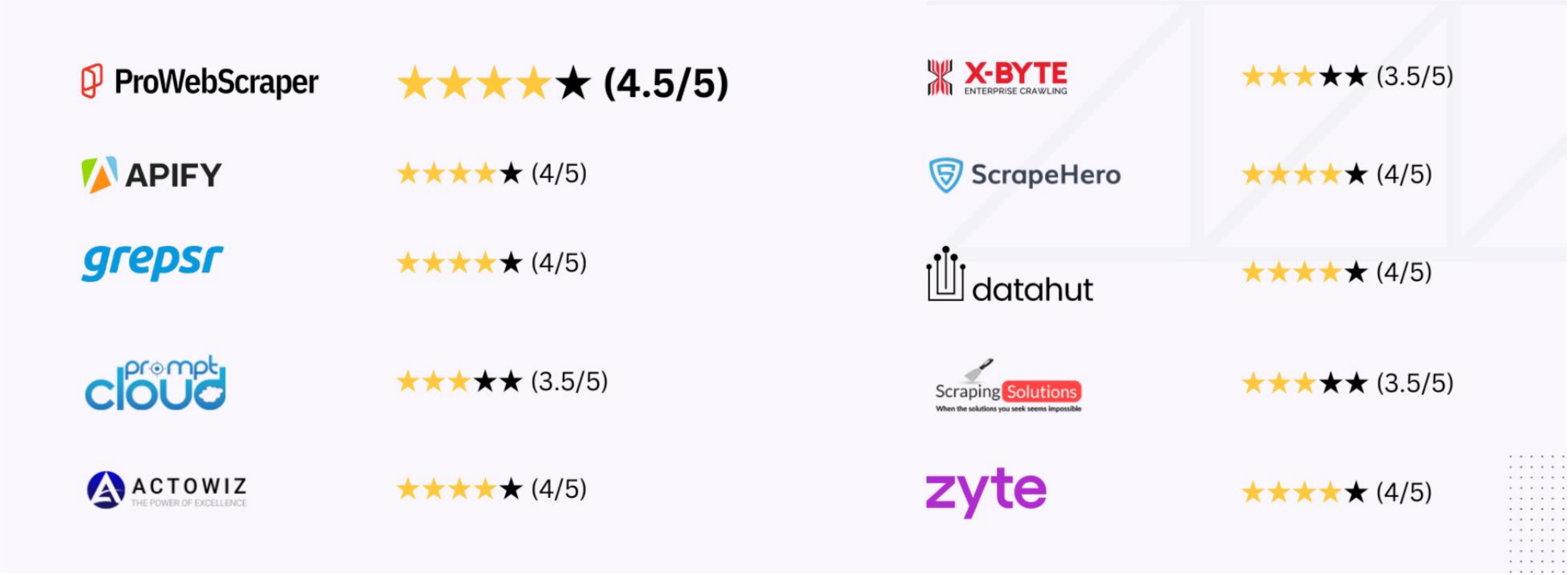
Libraries in Python

Web scraping companies

AI- based Web Scraping

- **Ease of use**: Often provide no-code or low-code solutions.
- **Scalability**: Built to handle large-scale scraping projects efficiently.
- **Maintenance-free**: Companies handle updates and changes to target websites.
- **Advanced features**: Often include proxy management, CAPTCHA solving, and data cleaning.

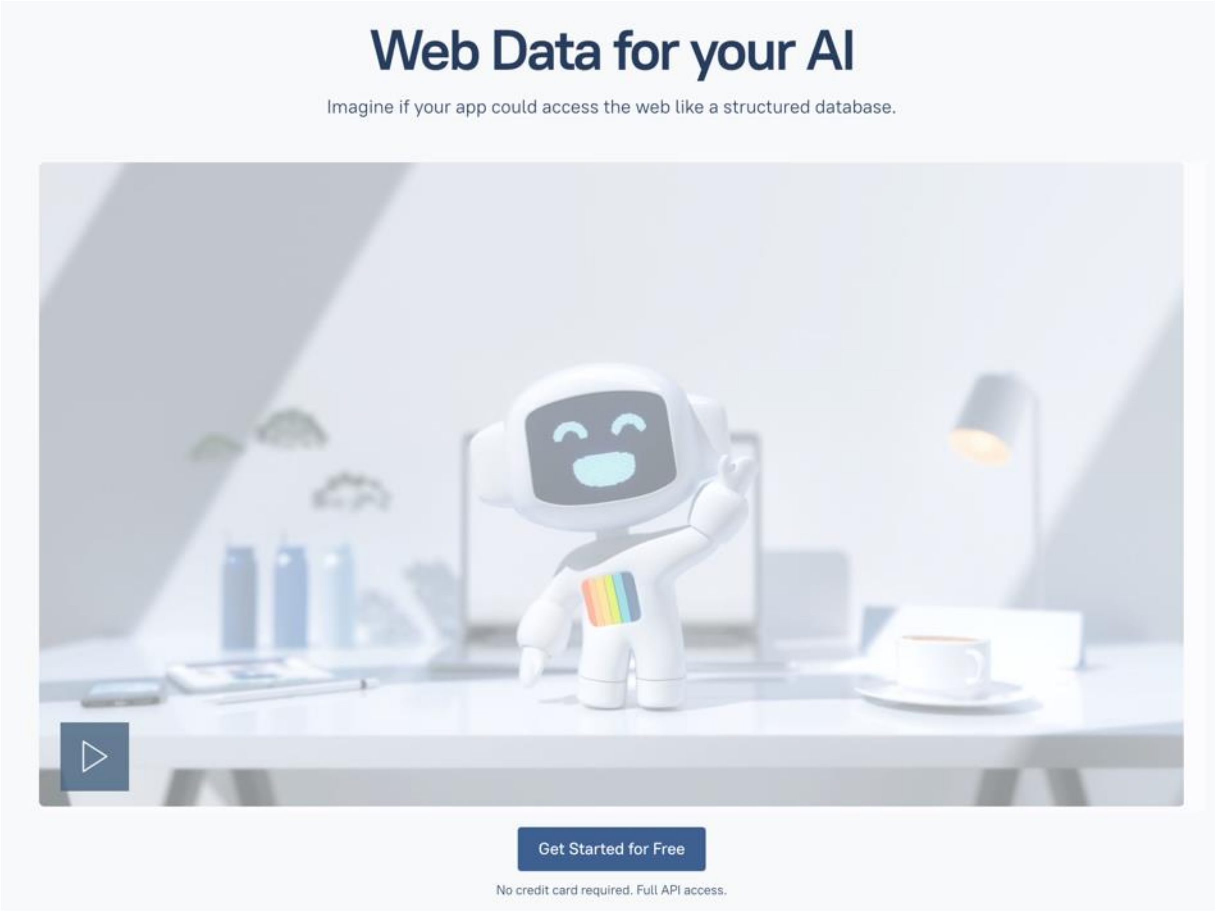
Web Scraping Companies



ProWebScraper	★★★★★ (4.5/5)	X-BYTE ENTERPRISE CRAWLING	★★★★★ (3.5/5)
APIFY	★★★★★ (4/5)	ScrapeHero	★★★★★ (4/5)
grepsr	★★★★★ (4/5)	datahut	★★★★★ (4/5)
prompt cloud	★★★★★ (3.5/5)	Scraping Solutions When the solutions you seek seems impossible	★★★★★ (3.5/5)
ACTOWIZ THE POWER OF EXCELLENCE	★★★★★ (4/5)	zyte	★★★★★ (4/5)

Example

- There are many AI-enabled Web scraping tools in the market. Diffbot is one of the best.
- Diffbot offers different types of subscription. You can sign up for a free account at: https://www.diffbot.com/

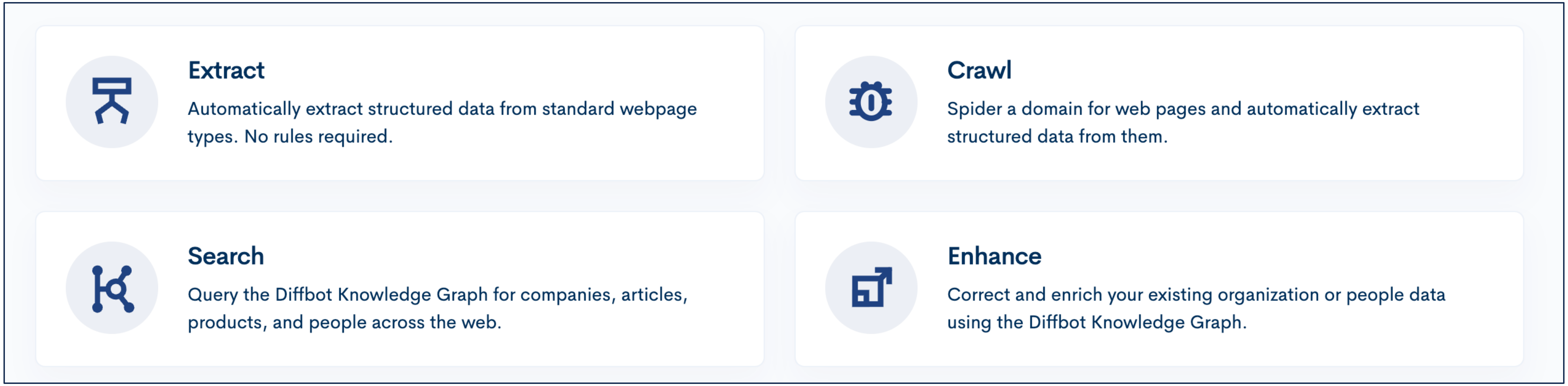


Web Data for your AI

Imagine if your app could access the web like a structured database.

Example

- To get started, press **Extract** in the home page



Extract

Automatically extract structured data from standard webpage types. No rules required.

Crawl

Spider a domain for web pages and automatically extract structured data from them.

Search

Query the Diffbot Knowledge Graph for companies, articles, products, and people across the web.

Enhance

Correct and enrich your existing organization or people data using the Diffbot Knowledge Graph.

The rise of AI-enabled Web Scraping

- The rise of AI is challenging both Python programming and conventional Web scraping methods.
- AI-enabled scraping can automatically identify, learn, and adapt to complex data patterns. For instance, instead of manually writing code to extract specific data fields, an AI model could learn how to identify these fields based on past patterns. This can be especially useful in cases where page structures vary significantly.
- **The only constant in the world is change.**

AI-Powered Analytics

HAWAIʻI PACIFIC UNIVERSITY

Tableau Pulse: AI-Empowered Analytics

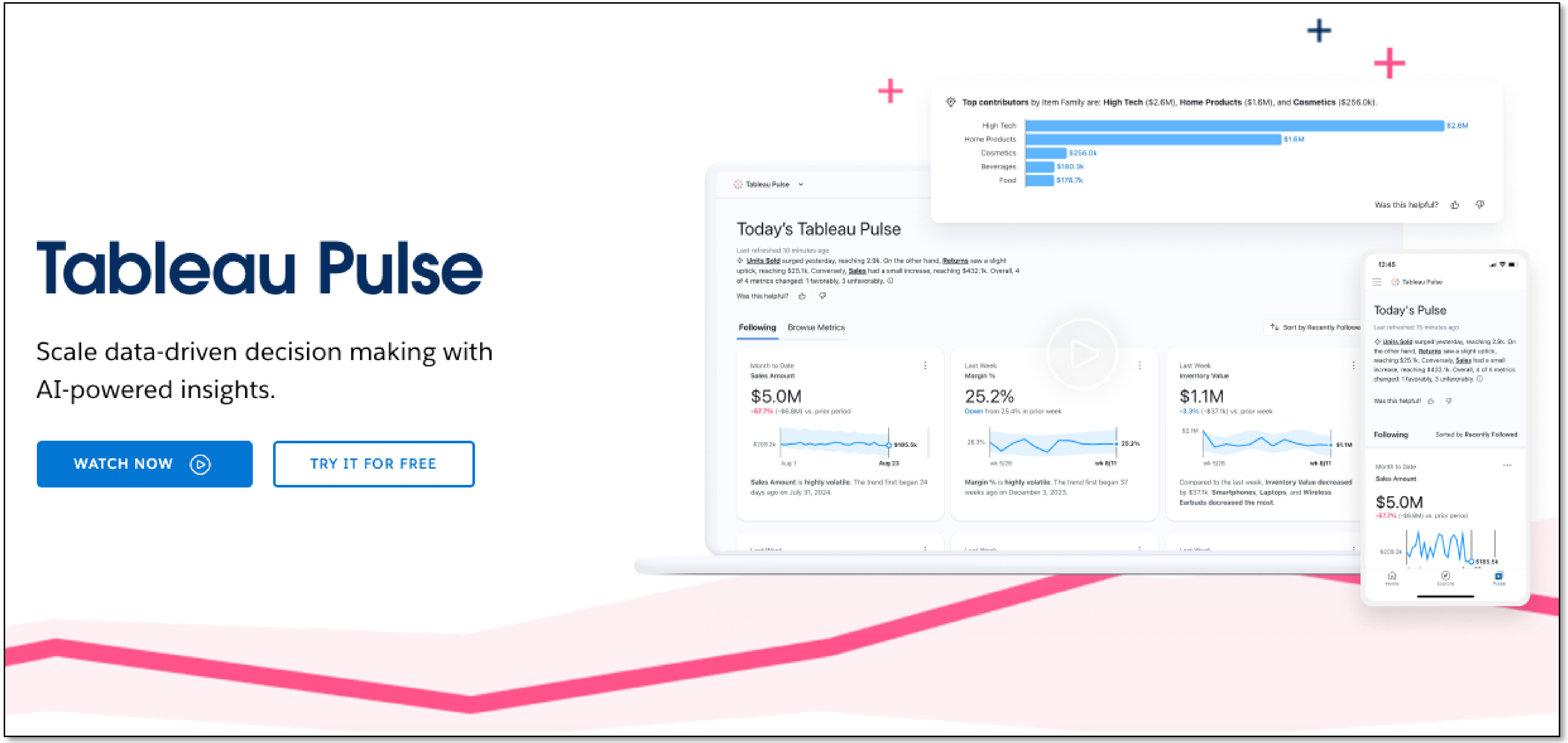


Tableau Pulse

Scale data-driven decision making with AI-powered insights.

WATCH NOW

TRY IT FOR FREE

Top contributors by Item Family are: High Tech ($2.6M), Home Products ($1.6M), and Cosmetics ($256.0k).

Today's Tableau Pulse

Last refreshed 10 minutes ago

Units Sold surged yesterday, reaching 2.9k. On the other hand, Returns saw a slight uptick, reaching $25.1k. Conversely, Sales had a small increase, reaching $432.1k. Overall, 4 of 4 metrics changed: 1 favorably, 3 unfavorably.

Was this helpful?

Following | Browse Metrics | Sort by Recently Followed

Month to Date
Sales Amount
$5.0M
-67.7% (-$6.8M) vs. prior period

Sales Amount is highly volatile. The trend first began 24 days ago on July 31, 2024.

Last Week
Margin %
25.2%
Down from 25.4% in prior week

Margin % is highly volatile. The trend first began 37 weeks ago on December 3, 2023.

Last Week
Inventory Value
$1.1M
-3.3% (-$37.1k) vs. prior week

Compared to the last week, Inventory Value decreased by $37.1k. Smartphones, Laptops, and Wireless Earbuds decreased the most.

Today's Pulse

Following

Month to Date
Sales Amount
$5.0M

Tableau Pulse: AI-Empowered Analytics

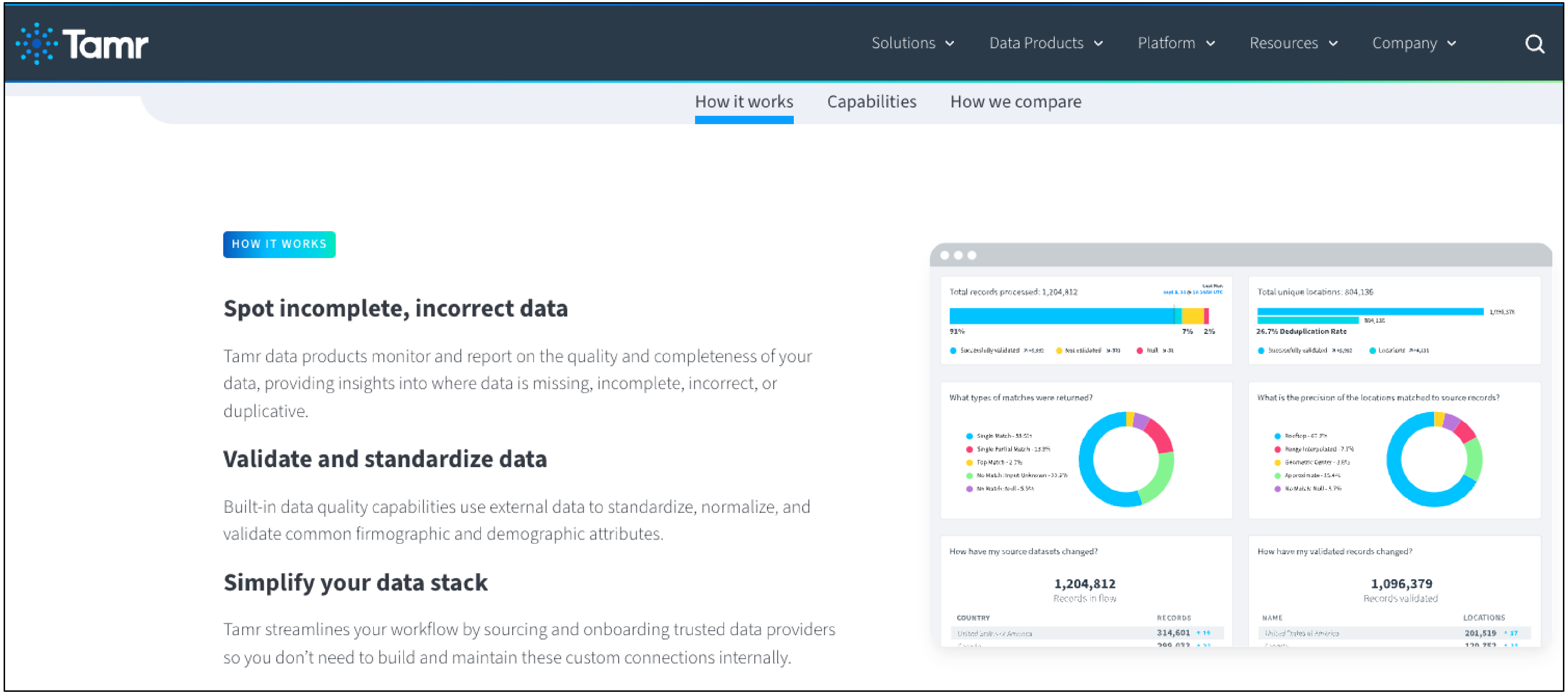
- Tableau Pulse is an advanced AI-driven analytics solution designed to integrate seamlessly within the Tableau platform, enhancing data analysis, accessibility, and decision-making processes.
- It introduces a **centralized Metrics Layer**, enabling organizations to define metrics and KPIs once and apply them consistently across the organization. This approach ensures data accuracy and reliability while saving analysts time by providing a single source of truth.

Tableau Pulse: AI-Empowered Analytics

- Tableau Pulse facilitates **real-time monitoring** of critical metrics, delivering proactive notifications and concise visual summaries. It alerts users to unexpected changes, emerging trends, key contributors, and outliers, empowering them to stay informed and make timely, data-driven decisions.
- Tableau Pulse allows users to interact through **natural language queries**, providing plain-language summaries and visual explanations. This makes insights accessible to everyone, including those with limited data analysis expertise, and encourages follow-up questions for deeper understanding.

AI-Powered Data Cleaning

Tamr



HOW IT WORKS

Spot incomplete, incorrect data

Tamr data products monitor and report on the quality and completeness of your data, providing insights into where data is missing, incomplete, incorrect, or duplicative.

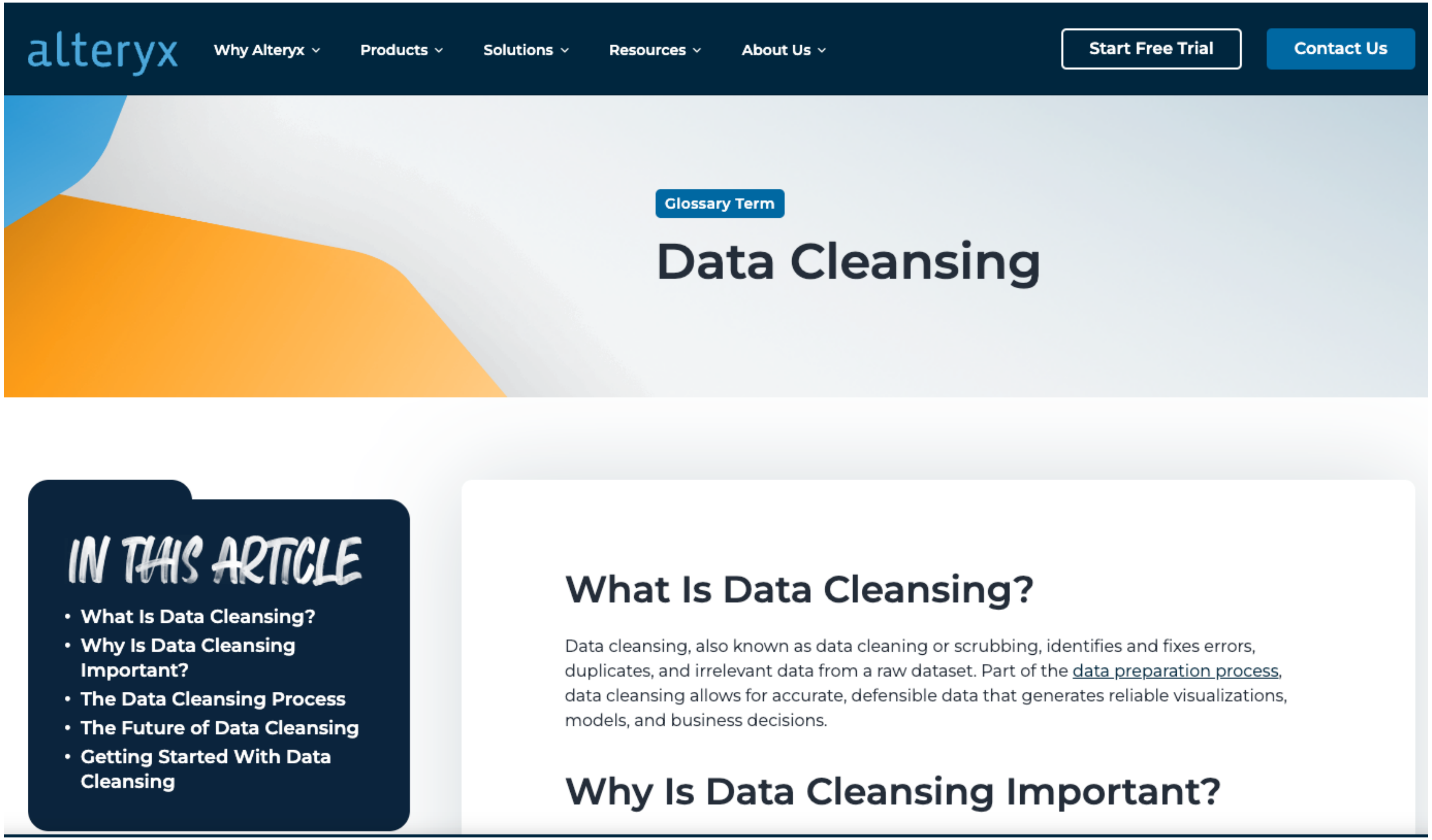
Validate and standardize data

Built-in data quality capabilities use external data to standardize, normalize, and validate common firmographic and demographic attributes.

Simplify your data stack

Tamr streamlines your workflow by sourcing and onboarding trusted data providers so you don't need to build and maintain these custom connections internally.

Alteryx



Data Cleansing

IN THIS ARTICLE

- What Is Data Cleansing?
- Why Is Data Cleansing Important?
- The Data Cleansing Process
- The Future of Data Cleansing
- Getting Started With Data Cleansing

What Is Data Cleansing?

Data cleansing, also known as data cleaning or scrubbing, identifies and fixes errors, duplicates, and irrelevant data from a raw dataset. Part of the data preparation process, data cleansing allows for accurate, defensible data that generates reliable visualizations, models, and business decisions.

Why Is Data Cleansing Important?

Alteryx

- **Pre-built transformations**: Alteryx offers a wide array of ready-to-use tools for data cleaning, such as handling null values, merging datasets, and correcting inconsistencies.
- **Workflow automation**: It enables users to build workflows for repetitive cleaning tasks, which can then be reused or scheduled for automated execution.
- **Visual interface**: Its intuitive interface allows users to perform complex data manipulations without needing to write code, making it accessible to analysts and business users.
- **AI-powered suggestions**: Some features provide recommendations for cleaning steps based on the dataset's structure and common issues.

Conclusion

Karim Lakhan, Harvard Business School Professor

“AI is not going to replace humans, but humans with AI are going to replace humans without AI.”

No code, low code, and prompts

- Today, many AI tools offer **no-code** or **low-code** solutions, with some even enabling users to interact through **natural language prompts**. This trend is reshaping the job landscape, leading to potential **displacement**. Those who understand how to leverage AI effectively will replace those who do not.
- In response to this shift, we must reconsider how we teach data analytics. Instead of focusing solely on programming and procedural skills, education should emphasize the ability to **conceptualize problems** and **formulate meaningful questions**. Developing these higher-order thinking skills will better prepare learners to work alongside AI and harness its capabilities to drive innovation and insight.

HAWAIʻI PACIFIC UNIVERSITY

Disclosure: AI tools are utilized for initial research and proofreading. The key ideas originates from the author.