CSCI 181DV Final Project
Project Proposal: Impact of AI on Digital Media (2020–2025)

Aiko Kato

1 - Project Definition & Scope

Problem Statement and Motivation

AI-generated content is rapidly transforming the digital media landscape, with profound effects on how industries operate and how people consume information. As a Computer Science major, I was particularly concerned by how quickly these AI tools are affecting jobs, public trust, and productivity around the world. The rise of generative models raises urgent questions about the future of labor, creative industries, and governance.

This project builds an interactive data visualization system to analyze the impact of AI-generated content on digital media industries across different countries between 2020 and 2025. By integrating data on AI adoption rates, job displacement, revenue growth, human-AI collaboration, regulation status, and consumer trust, the dashboard allows users to explore the complex and evolving relationship between AI and society.

The core motivation is to better understand this transformation so that policymakers, business strategists, and researchers can make informed decisions. As AI tools become more accessible and powerful, it's important to analyze where and how they are reshaping digital labor, especially across sectors like journalism, social media, marketing, and entertainment.

Target Users and Use Cases

- Policymakers: Assess AI's impact to inform future regulations.
- Business strategists: Understand industry-specific adoption and its correlation with revenue and trust.
- Researchers: Analyze temporal and regional trends, correlations, and anomalies in global AI adoption.
- Educators/Students: Use the dashboard as a learning tool to understand how AI intersects with the digital economy.

Expected Insights and Outcomes

- Identify countries with the fastest-growing AI adoption over time.
- Compare AI adoption and outcomes across industries.
- Detect correlations between adoption rate and job loss, revenue increase, and collaboration.
- Visualize how regulation and public trust interact globally.
- Discover which AI tools dominate in each industry.
- Reveal disparities between countries and sectors in AI impact.

Project Boundaries and Constraints

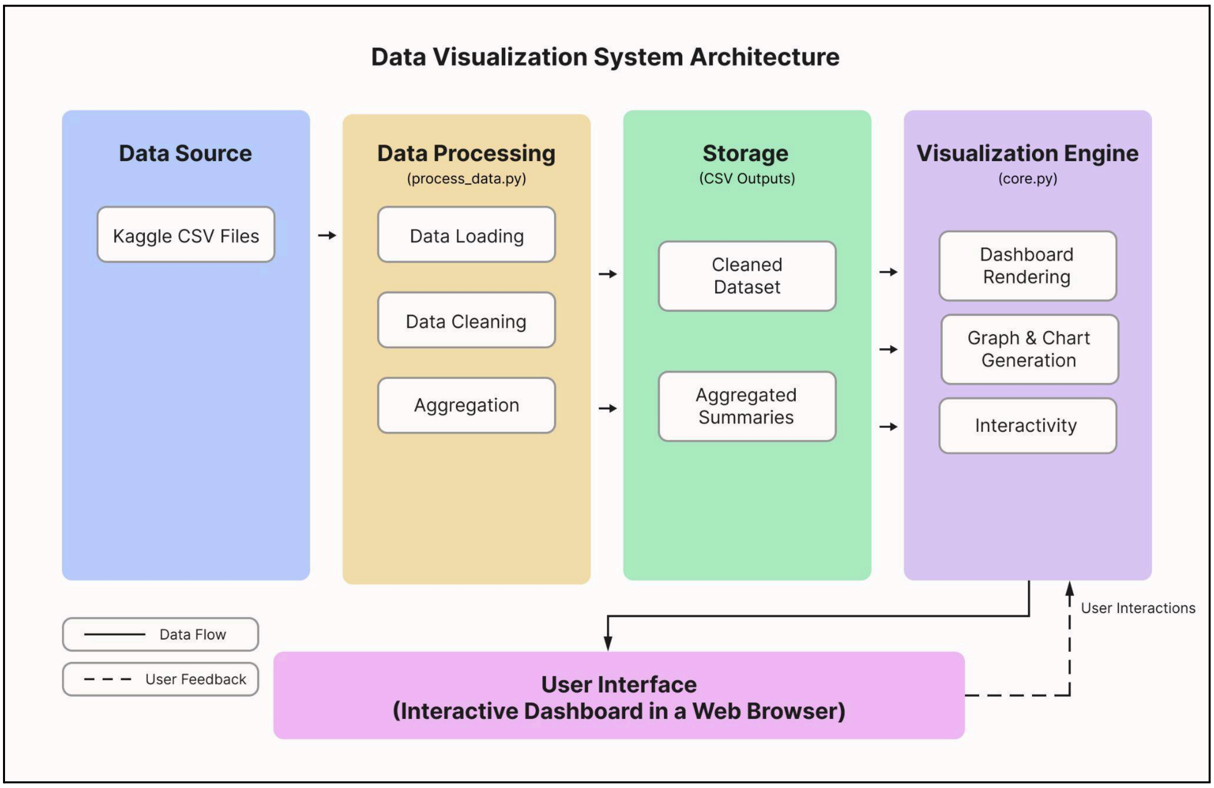
- Focus is limited to the 2020–2025 timeframe.
- Limited diversity in countries and industries.
- Dataset is relatively small (~200 rows) and missing some values, including Education in 2024.
- No real-time updates or API integration assumed.
- Potential data bias due to unclear source provenance and overlapping entries.
- Interface optimized for desktop, with limited mobile accessibility.

Potential Future Improvements

- Expand the dataset to include more countries, industries, and years before 2020 and beyond 2025.
- Integrate live data sources via APIs for real-time monitoring and auto-refresh dashboards for continuous updates.
- Enhance UI/UX design for better accessibility, mobile responsiveness, and layout optimization.
- Support multilingual and accessibility features for broader public use.

2 - Technical Design

System Architecture Diagram



Technology Stack Selection and Justification

1. Visualization Libraries
    - Plotly/Dash
        - Used to build the main interactive visualizations including choropleth maps, multi-line charts, scatter plots, network graphs, and heatmaps. These libraries support zooming, filtering, and linked views, and form the core of the dashboard interface.
    - Streamlit (alternative option)
        - Considered during the prototyping phase for its simplicity and ease of setup. Streamlit was suitable for fast development but not used in the final product due to limited support for complex callbacks and multi-view interactions.
2. Frameworks
    - Dash
        - A Python-based framework used to create the web-based dashboard. Dash supports layout rendering, multi-tab navigation, user input components, and the reactive callback system that updates visualizations dynamically in response to selections and filters.
3. Data Processing
    - Python (Pandas)
        - Used to load the Kaggle dataset, handle missing or inconsistent data, standardize country and industry names, normalize numeric fields such as AI adoption percentages, and aggregate results by country, industry, and year. Additional metrics such as collaboration rate, job loss ratio, and revenue increase rate were computed for visualization.

4. Storage
    - File-based storage (CSV)
        - Cleaned datasets and precomputed summaries are saved to local files to reduce dashboard loading time and avoid repeating data processing.
    - In-memory storage (Pandas DataFrames)
        - DataFrames are used at runtime to support fast filtering, searching, and dynamic visualization updates without re-reading raw files from disk.

Data Flow and Processing Pipeline Design

- Load data
    - The Kaggle dataset titled “Impact of AI on Digital Media 2020–2025” is loaded into memory using Pandas.
- Data cleaning
    - Missing values are handled and text fields such as country and industry names are standardized. Numeric fields such as adoption rate, job loss, and trust percentage are converted and normalized.
- Data transformation and aggregation
    - Metrics are grouped by country, industry, and year. Averages, totals, and proportions are computed to prepare summary views.
- Caching and storage
    - Cleaned and aggregated data is saved to local CSV files to ensure faster dashboard startup and reduce unnecessary computation during user interaction.
- Visualization input
    - The cached datasets are loaded into the visualization components. They serve as inputs for choropleth maps, line charts, scatter plots, network graphs, bar charts, and heatmaps with support for zooming, filtering, and linked selections.

Performance Considerations and Strategies

- Precompute aggregates
    - Data is grouped and summarized ahead of time by country, industry, and year to prevent heavy processing during dashboard use.
- Caching intermediate results
    - All cleaned and aggregated datasets are stored locally to allow immediate access during dashboard initialization.
- Lazy loading
    - Visualizations that require larger datasets, such as network graphs, are only loaded when the user navigates to their respective tabs.
- Optimized filtering
    - In-memory DataFrames are used to enable fast filtering and search operations without reading from disk each time a filter is applied.
- Efficient data access
    - Key columns such as country, industry, and year are grouped and filtered during interaction.
- Minimized redraws
    - Only visual components affected by user actions are updated. This reduces rendering load and ensures smooth interaction.

3 - Data Strategy

Dataset Selection and Justification

Dataset: [Impact of AI on Digital Media (2020-2025)] from Kaggle

This dataset covers multiple metrics (adoption rate, job loss, revenue impact, etc.) across several countries and industries. It is comprehensive and well-suited for multi-dimensional analysis and interactive storytelling.

Data Cleaning and Preprocessing Approach

- Remove rows with missing key fields (e.g., year, country, AI adoption rate).
- Standardize text fields (e.g., industry and country names).
- Convert numeric fields to ensure compatibility and drop rows with invalid values.

Storage and Retrieval Strategy

- Use pandas DataFrames during processing.
- Export cleaned/aggregated data to CSV or pickle for quick reload.
- Load preprocessed data once at app startup to reduce redundant reads.

Data Update and Maintenance Plan

- Since this is a static dataset, no automatic updates are assumed.
- Maintain versioning for all transformed data files.
- Include scripts for re-cleaning and re-generating visualizations from raw data.

4 - Visualization Design

Mockups of Key Visualizations

- Choropleth Map
  - Displays AI adoption rate or content volume by country for a selected year. Users can choose the metric and year using a radio toggle and slider.
  - Essential for policymakers and researchers to visually compare adoption levels or content generation across countries in a single time slice. Enables quick identification of geographic disparities and high-impact regions.

- Choropleth Map (Animated)
  - Shows how AI adoption evolves globally from 2020 to 2025. A built-in animation cycles through years automatically for temporal comparison.
  - Highlights temporal dynamics for policymakers and students, helping them track how countries accelerate or slow down in adoption. Useful for analyzing policy effects over time.

- Line Chart
  - Compares AI adoption trends across multiple industries over time. Users can select industries to highlight.
  - Crucial for business strategists to monitor long-term trends within specific sectors and identify rising or declining industries. Helps users understand sector-specific digital transformations.

- Bar Chart (AI-Generated Content Volume by Industry)
  - Visualizes the total AI-generated content (in terabytes) per industry across the full time span. Bars include percentage labels for comparison.
  - Offers a direct way to compare content production volume between industries. Enables business analysts to assess where AI is driving the most output and compare market size.

- Bar Chart (Consumer Trust in AI by Regulation Status)
  - Compares average trust levels across Lenient, Moderate, and Strict regulatory environments.
  - Designed for policymakers and researchers to assess how regulation levels correlate with public trust. Helps support arguments for or against certain policy models.

- Scatter Plot
  - Each point represents a country-industry-year. The X-axis shows AI adoption rate, and the Y-axis shows either job loss or revenue increase (toggleable). Color-coded by industry.
  - Facilitates exploration of potential trade-offs or gains—such as whether higher adoption correlates with job displacement or financial growth. Supports strategic business and labor market analysis.

- Scatter Plot (Linked View)
  - Users can select points in the scatter plot to filter the line chart, which then shows AI adoption trends over time for the selected industries.
  - Provides interactivity for advanced users (researchers, analysts) to drill down into specific subsets of data. Helps discover deeper patterns and outliers across time and industry.

- Bar Chart (AI Tools Used Across Industries)
  - Ranks AI tools by usage frequency across all industries. Bars are labeled with a percentage share of total tool mentions.
  - Helps all user groups identify dominant technologies in the digital media space. Useful for evaluating industry reliance on specific tools and anticipating future developments.

- Network Graph
    - Displays a network where nodes represent AI tools and industries. Edges indicate usage relationships, with thickness based on frequency. Hover reveals breakdowns per node.
    - Ideal for researchers and students to visualize complex relationships between AI tools and industries. Helps uncover clusters, central technologies, and potential cross-industry influences.

- Heatmap
    - A matrix-style visualization showing content volume per industry per year. Color intensity reflects terabytes generated.
    - Useful for spotting patterns over time and across sectors. Supports educators and analysts in identifying which industries scaled AI usage consistently and when major shifts occurred.

- Dashboard
    - Mockup visualization for dashboards are attached below.
    - The dashboard brings all visualizations together in one place, making the data easy to explore through interactive filters and charts. It helps users compare trends, find insights, and make informed decisions quickly.

Interaction Design Specifications

- Hover tooltips display metric values along with metadata such as country, year, industry, regulation status, and usage breakdowns depending on the chart.
- Interactive filters include radio buttons (for metric and regulation), a year slider (for choropleth and heatmap), and an industry dropdown (for the line chart).
- Time-series charts support zooming and panning for closer inspection of trends.
- Selecting points in the scatter plot updates a linked line chart that shows adoption trends for selected industries.
- These interaction features are designed to allow users to dynamically explore trends, focus on particular regions or industries, and uncover detailed patterns through flexible, user-driven exploration.

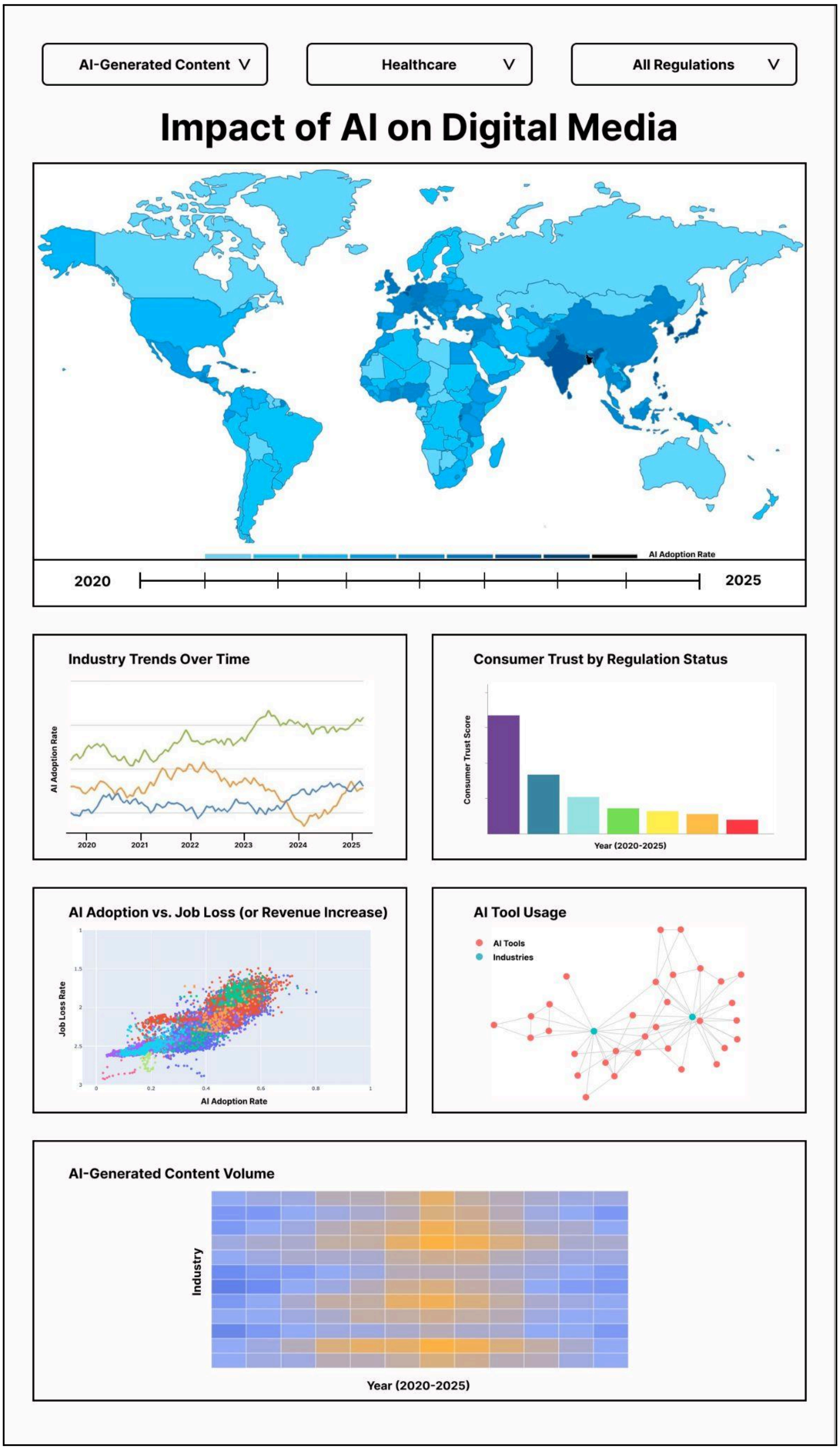
Color Schemes and Visual Encoding Choices

- Custom high-contrast color maps are used consistently across charts for clarity and distinction.
- Adoption rate uses a blue gradient, while other metrics are mapped to consistent metric-specific scales.
- Categorical distinctions (industries, tools) via contrasting hues.
- These color choices ensure consistent interpretation across visualizations while maintaining accessibility and emphasizing key comparisons between numerical metrics and categories.

Accessibility Considerations

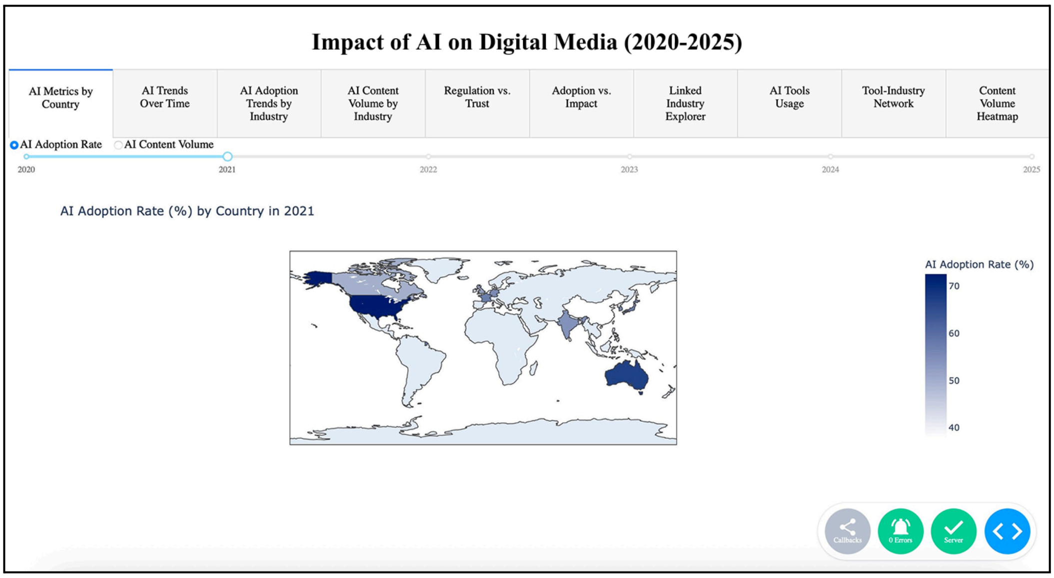
- Text contrast and readable fonts.
- Avoid reliance on color alone for encoding.
- These accessibility measures ensure that users of varying abilities can fully engage with the dashboard, promoting inclusivity and compliance with best practices for interactive design.

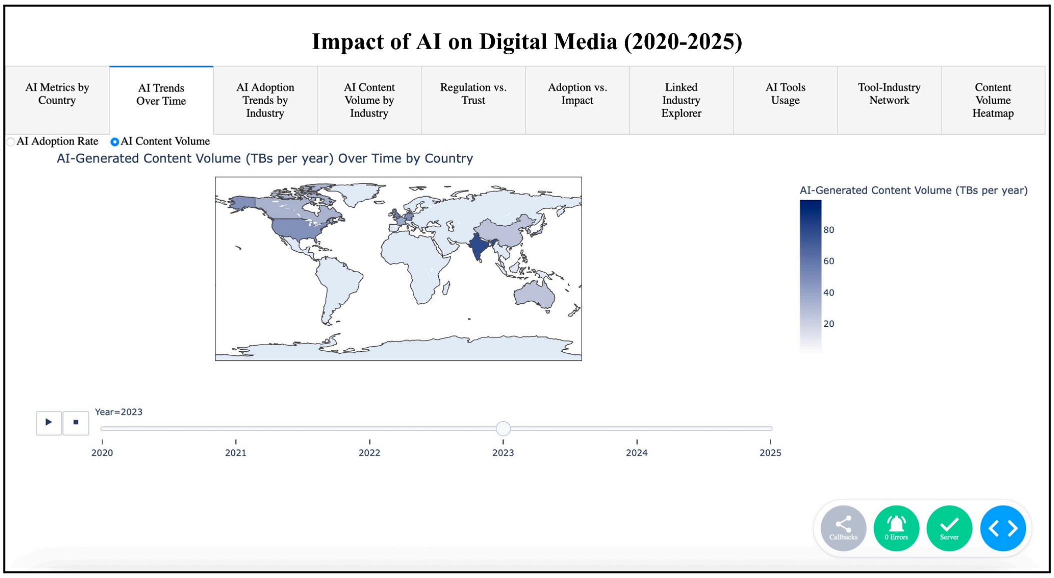
Dashboard Mockup Visualization (Previous)

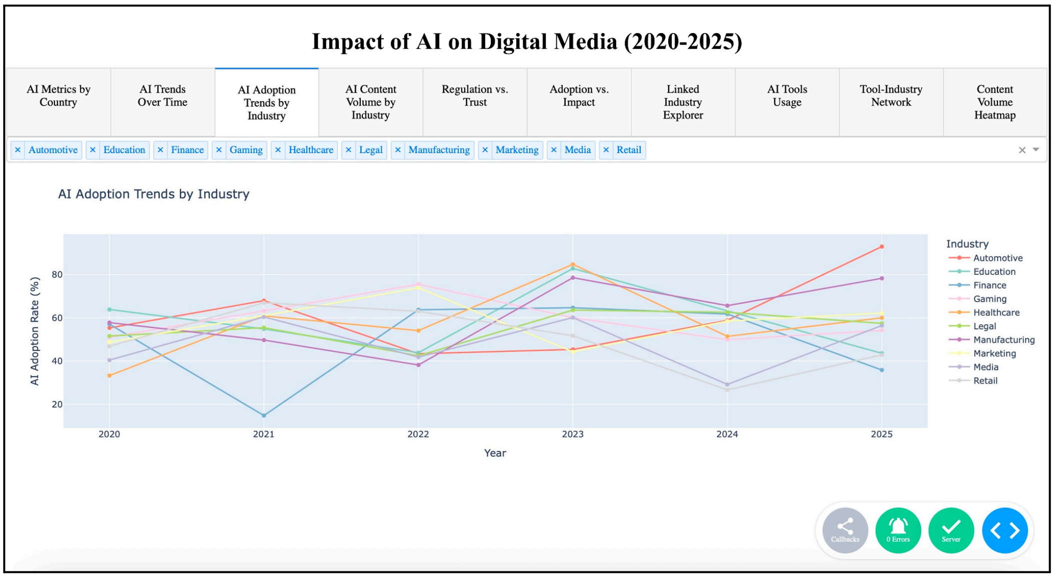


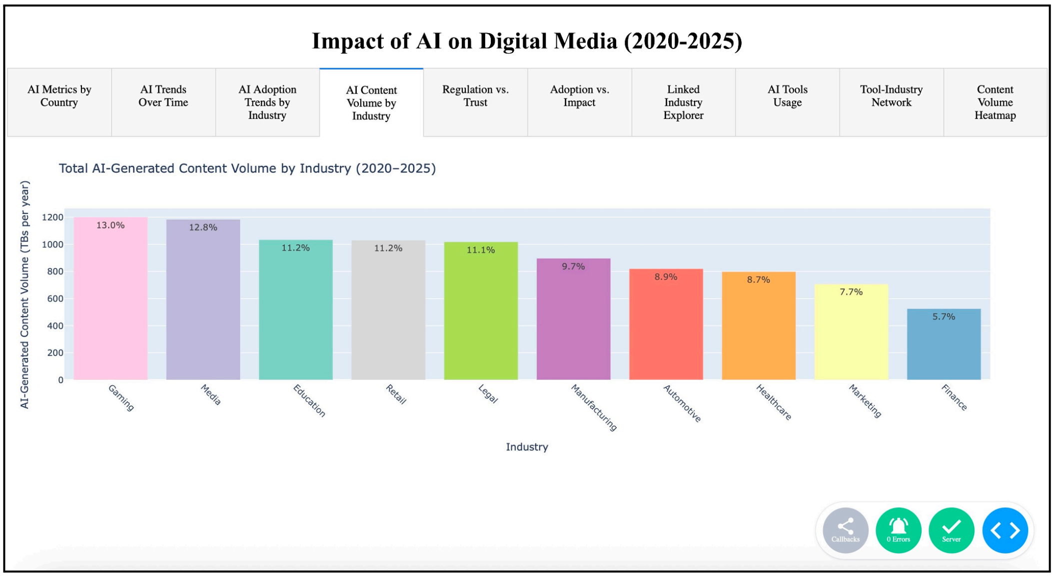
Since the previous dashboard layout displayed all visualizations at once, I decided to switch to a tab-style layout to reduce visual clutter and improve user navigation. This change allows users to focus on one analysis at a time, enhances readability, and makes the dashboard more intuitive and responsive.

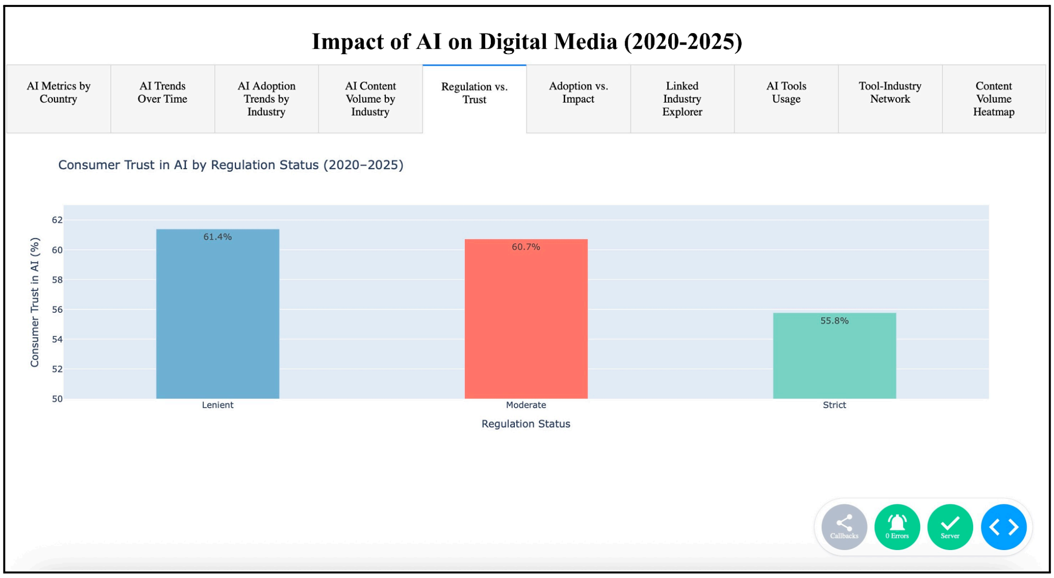
Dashboard Visualization (Final)

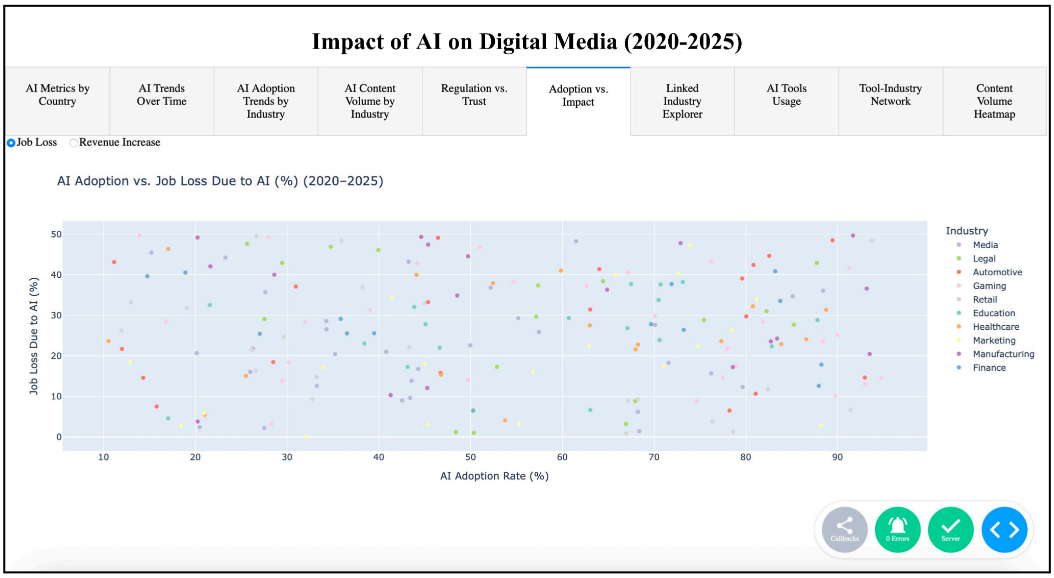


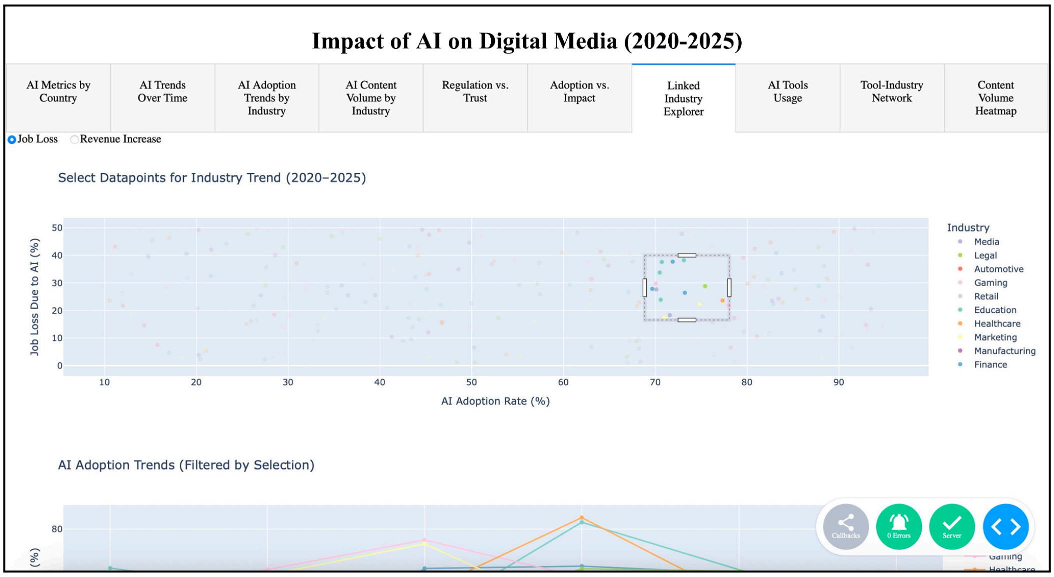


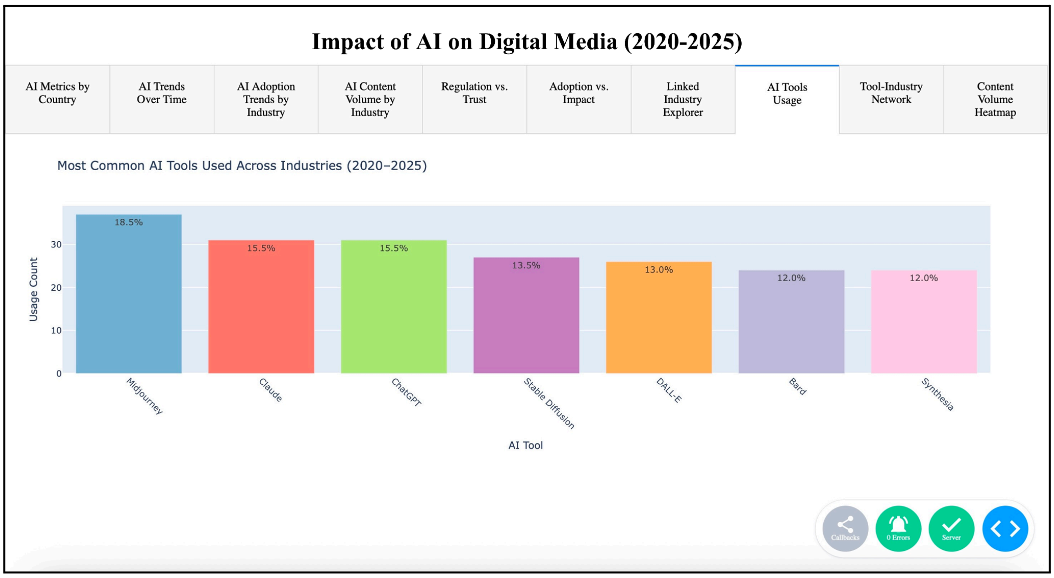


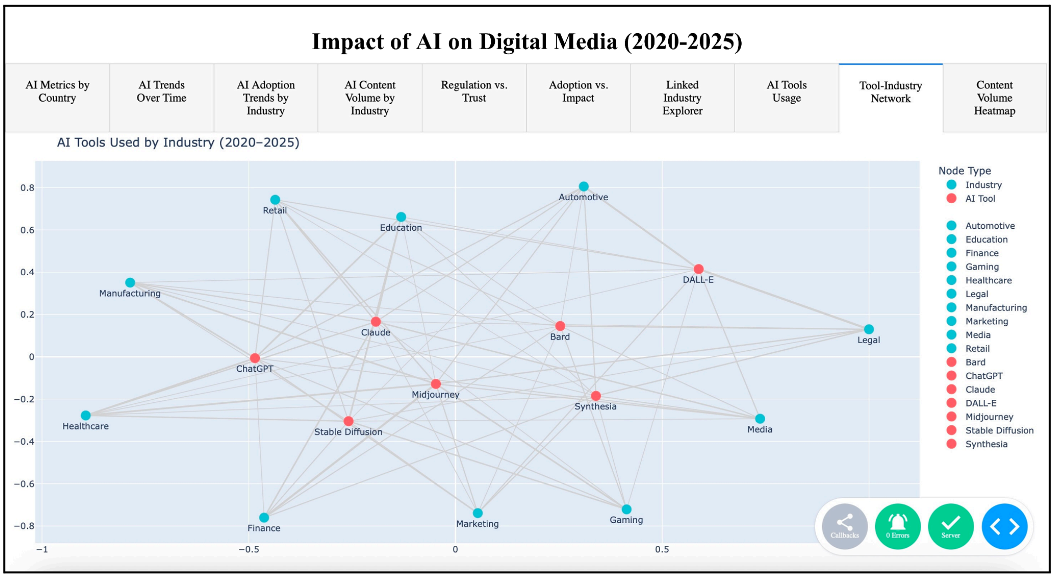

Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
Total AI-Generated Content Volume by Industry (2020–2025)
AI-Generated Content Volume (TBs per year)
13.0%
12.8%
11.2%
11.2%
11.1%
9.7%
8.9%
8.7%
7.7%
5.7%
Gaming
Media
Education
Retail
Legal
Manufacturing
Automotive
Healthcare
Marketing
Finance
Industry

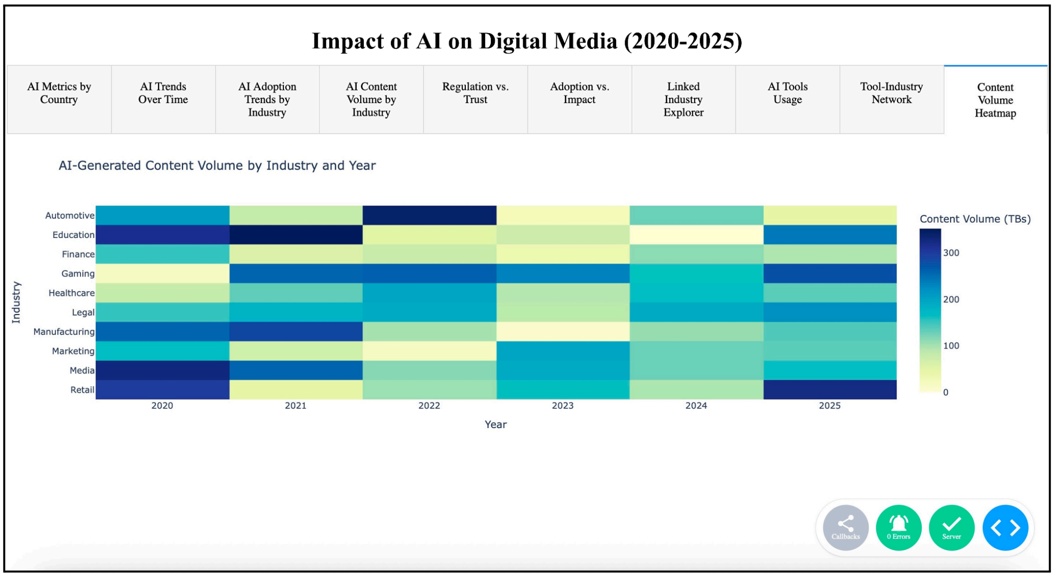

Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
Consumer Trust in AI by Regulation Status (2020–2025)
Consumer Trust in AI (%)
61.4%
60.7%
55.8%
Lenient
Moderate
Strict
Regulation Status


Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
Job Loss
Revenue Increase
AI Adoption vs. Job Loss Due to AI (%) (2020–2025)
Job Loss Due to AI (%)
AI Adoption Rate (%)
Industry
Media
Legal
Automotive
Gaming
Retail
Education
Healthcare
Marketing
Manufacturing
Finance


Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
Job Loss
Revenue Increase
Select Datapoints for Industry Trend (2020–2025)
Job Loss Due to AI (%)
AI Adoption Rate (%)
Industry
Media
Legal
Automotive
Gaming
Retail
Education
Healthcare
Marketing
Manufacturing
Finance
AI Adoption Trends (Filtered by Selection)


Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
Most Common AI Tools Used Across Industries (2020–2025)
Usage Count
18.5%
15.5%
15.5%
13.5%
13.0%
12.0%
12.0%
Midjourney
Claude
ChatGPT
Stable Diffusion
DALL-E
Bard
Synthesia
AI Tool


Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
AI Tools Used by Industry (2020–2025)
Node Type
Industry
AI Tool
Automotive
Education
Finance
Gaming
Healthcare
Legal
Manufacturing
Marketing
Media
Retail
Bard
ChatGPT
Claude
DALL-E
Midjourney
Stable Diffusion
Synthesia


Impact of AI on Digital Media (2020-2025)
AI Metrics by Country
AI Trends Over Time
AI Adoption Trends by Industry
AI Content Volume by Industry
Regulation vs. Trust
Adoption vs. Impact
Linked Industry Explorer
AI Tools Usage
Tool-Industry Network
Content Volume Heatmap
AI-Generated Content Volume by Industry and Year
Automotive
Education
Finance
Gaming
Healthcare
Legal
Manufacturing
Marketing
Media
Retail
Industry
2020
2021
2022
2023
2024
2025
Year
Content Volume (TBs)
300
200
100
0