



Master's Degree in Data Science and Management

Course: Data Visualization

Data Humanism in the Age of Generative AI:
An Experimental Approach to Narrative Visualization

Prof. Blerina Sinimeri

SUPERVISOR

Prof. Irene Finocchi

CO-SUPERVISOR

Emanuela Rremilli 783281

CANDIDATE

Academic Year 2024/2025

Abstract

This thesis investigates whether generative artificial intelligence can reinforce the human-centric values of Data Humanism when humans and AI co-create data stories. It presents the design and evaluation of a web-based storytelling system that integrates GPT-4 for narrative generation with D3.js for interactive visualization. The platform operationalizes Lupi's principles: personalization, contextual nuance, and narrative depth, through two complementary modes: a persona-based mode for audience-aware explanations and a Data Humanism mode encouraging creative, metaphor-rich encodings beyond standard visualizations. The systems include a preprocessing pipeline that summarizes user datasets to ground model outputs and ensure reproducibility in both text and code.

Methodologically, the research combines system implementation and prompt-engineering with two empirical strands: (i) a cross-persona study demonstrating controllable shifts in tone, emphasis, and visualization strategies across roles, and adaptability in different datasets; and (ii) a comparative survey contrasting AI-generated outputs with hand-crafted, humanistic visualizations inspired by Giorgia Lupi. Together, these studies assess communicative effectiveness, engagement, and perceived trust.

Findings show a complementary trade-off: participants tend to rate AI-assisted stories higher on clarity and efficiency, while human-crafted designs excel in empathy, memorability, and aesthetic impact. Crucially, many readers prefer side-by-side use, reporting more complete understanding when analytical precision is juxtaposed with humanistic meaning, supporting a hybrid paradigm of human-AI co-creation. The thesis concludes that generative AI can augment, but not replace, the ethos of Data Humanism, provided that prompts are carefully designed, data grounding is explicit, and human oversight safeguards context and ethics. Limitations include small-sample evaluations, reliance on a single model, and current constraints in generative support for bespoke visual forms; future work should explore multimodal generation and longitudinal assessments of learning and trust.

Keywords: Data Humanism; Generative AI; Narrative Visualization; GPT-4; Prompt Engineering; D3.js; Data Storytelling Application.

The full implementation code developed for this thesis is publicly available at the following repository: *DataStorytelling-Project* (<https://github.com/em-rg/DataStorytelling-Project>).

Table of Contents

1. Introduction.....	5
2. Literature Review.....	7
2.1 Giorgia Lupi’s Data Humanism Vision.....	8
2.2 Humanizing Data: Storytelling, Design, and Visualization	8
2.3 AI and the Future of Creative Data Communication	10
3. System Implementation and Prompt Engineering Methodology	11
3.1 LangChain & GPT-4o API Integration	11
3.2 Parameter Tuning	13
3.3 Prompt Engineering Techniques	15
3.4 Persona-based Prompt Logic	16
3.5 Data Humanism Mode	17
3.6 D3.js Visualization Rendering	20
3.7 Data Preprocessing and JSON Summarization.....	21
4. Cross-Persona Evaluation of the AI Storytelling System	23
4.1 Overview	23
4.2 Global Education Dataset.....	23
4.2.1 Data Scientist Persona.....	24
4.2.2. Business Analyst Persona.....	29
4.2.3 Content Creator Persona	32
4.2.4 Journalist Persona	37
4.2.5 Data Humanism Mode	40
4.3 WASH Dataset	47
4.3.1 Community Caregiver Persona	48
4.3.2 Community Caregiver Persona on Data Humanism Mode.....	51
4.4 Discussion and Implications	53
4.5 Limitations and Threats to Validity	55
5. Comparing Human vs AI Data Visualization Storytelling.....	58
5.1 Visualization Pairs: Human-Made vs AI-Generated Designs	58
5.1.1 Nobels, No Degrees	58
5.1.2 European Banks and Government Debt.....	61
5.1.3 Geniuses, Visualized	63

5.2 Quantitative Results	65
5.2.1 Overall Ease of Understanding	65
5.2.2 Role of Interactivity	66
5.2.3 Visual Appeal and Aesthetics	66
5.2.4 Trust and Perceived Accuracy	67
5.3 Qualitative Insights	68
5.3.1 Clarity vs. Complexity	69
5.3.2 Emotional and Personal Resonance	69
5.3.3 Aesthetic Appeal and Memorability	70
5.3.4 Interactivity and Exploration	70
5.3.5 Purpose and Context	71
5.4 Synthesis of Findings	71
6. Conclusion and Further Developments	73
References	74
Appendix A	77

1. Introduction

In recent years, the field of data visualization has become dominated by technical dashboards and infographics that often lack meaningful narrative context. While dashboards offer interactive views for exploring data, they can feel impersonal and “sterile,” requiring users to extract insights on their own [1]. Indeed, critics observe that conventional visual design sometimes treats data as a “cosmetic retouch”, superimposing simple charts and pictograms on complex issues in an attempt to “simplify complexity”, often resulting in unreadable or misleading graphics [2]. As Kahneman famously notes, “No one ever made a decision because of a number. They need a story” [1]. In other words, data alone is insufficient; insights emerge through narrative and human connection. These limitations suggest a need for approaches that ground data in human experience and storytelling, rather than sterile technical displays.

Data Humanism has emerged as a philosophical and design response to these limitations. Coined by Giorgia Lupi and collaborators, Data Humanism is a human-centered design approach that “emphasizes the personal, contextual, and imperfect nature of data” [3]. In practice, this means treating data points not as abstract numbers but as reflections of real people and behaviors. Data Humanism’s core principles include foregrounding the human aspects of information, embracing ambiguity and imperfection, and favoring personal connection over standardization [3]. Lupi argues that we must “question the impersonality of a merely technical approach to data” and design ways to connect numbers to what they really stand for: knowledge, behaviors, people [2]. Under this framework, narrative visualization and hand-crafted infographics are valued for their ability to convey nuance and tell stories. Related work in critical and human-centered visualization also highlights that data gains meaning only through context and narrative, aligning with the Data Humanism ethos of personalized storytelling [2, 3].

At the same time, generative artificial intelligence has introduced new possibilities for data storytelling. Large Language models (LLMs) like GPT-4 and image generators can automate many tasks in the analysis and presentation pipeline. These tools can rapidly transform raw numbers into natural-language summaries, captions, and annotations, effectively adding narrative elements to visualizations [4, 5]. They also support personalization: by leveraging prompt engineering, generative systems can tailor data narratives to different audiences, adjusting tone and detail to improve engagement [4, 6]. In short, generative AI can shift data storytelling from rigid processes to more intuitive, intent-driven interactions that emphasize the “why” and broaden participation beyond technical experts [6, 4]. At the same time, this shift raises new challenges. AI-generated narratives

may be ambiguous if prompts are unclear, and users may lose control over visualization design [6]. Moreover, LLMs are trained on human-written content and can introduce biases or factual errors into data stories.

This thesis explores the intersections of these trends by designing, implementing, and evaluating an AI-driven data storytelling system grounded in Data Humanism principles. Specifically, a web-app storytelling platform is built that combines GPT-4 for narrative generation with D3.js for dynamic visualization. The system supports a *persona mode*, where different audience archetypes receive customized narratives, and a *Data Humanism mode*, which encourages creative styles of storytelling. By operationalizing Lupi’s principles in software, the goal is to animate data with human-centered nuance: transforming abstract datasets into richly contextualized stories. To assess the approach, we compare the AI-generated narratives and charts against those created by human storytellers. A user survey evaluates how audiences perceive and engage with these humanized versus AI-driven data stories.

The thesis is structured as follows, Chapter 2 (Literature Review) describes Giorgia Lupi’s concept of Data Humanism and its design principles and situates it within the broader literature on narrative visualization and human-centered design. Chapter 3 (System Implementation and Prompt Engineering Methodology) presents the architecture of the interactive storytelling system, detailing how GPT-4 is integrated for text generation, how D3.js creates the visuals, and how the persona and Data Humanism modes are realized. Chapter 4 (Cross-Persona Evaluation of the AI Storytelling System) demonstrates how the system produces different narratives and charts for various personas and creative settings, with examples illustrating the differences. Chapter 5 (Comparing Human vs AI Data Visualization Storytelling) reports on an empirical study comparing human and AI-generated visualizations and examines audience perceived preferences and feelings. Finally, Chapter 6 (Conclusion and Further Developments) reflects on the implications of Data Humanism in the age of generative AI and outlines directions for future work in bridging human-made visualization with AI-assisted storytelling. Together, this work assesses whether generative AI can reinforce the human-centric values of Data Humanism when stories are co-created with artificial intelligence.

The full implementation code developed for this thesis is publicly available at the following repository: *DataStorytelling-Project* (<https://github.com/em-rg/DataStorytelling-Project>).

2. Literature Review

The term Data Humanism was coined by information designer Giorgia Lupi to describe a “second wave” of visualization that privileges human meaning and emotion over sterile precision [2, 3]. In Lupi’s view, the era of “peak infographics”, endless reels of bar charts and techno-centric dashboards, has passed [2]. In practical terms, this means rejecting the impersonal dictum of maximal “data-ink” purity and instead embracing design that highlights subjectivity, context, and narrative. As Yan et al. summarize, Data Humanism is fundamentally a human-centered design approach that emphasizes the personal, contextual, and imperfect nature of data [3]. In short, it treats datasets not as objective truth-machines but as crafted records of lived experience.

Lupi’s *Data Humanism Manifesto* outlines core principles to realize this vision. Foremost is embracing complexity and nuance. Rather than forcing every dataset into a simple bar chart type, humanist visualizations layer multiple narratives and data dimensions. Lupi writes: “since clarity does not need to come all at once, we layered multiple visual narratives over a main construct... we call this process nonlinear storytelling” [2]. In practice, an information graphic might invite viewers to “get happily lost” exploring side narratives and subsidiary details, uncovering deeper meaning at each step [2]. Such dense, unconventional design deliberately slow down the reader. As Lupi notes, “We can write rich and dense stories with data. We can educate the reader’s eye to become familiar with visual languages that convey the true depth of complex stories” [2]. In other words, humanist graphics reward careful attention rather than instant digestion.

Another key principle is moving beyond standard chart templates. Data Humanism instead insists on designing to the data, not the other way around. Lupi famously explains: “*If the data I’m working with doesn’t fit a bar chart, I don’t want it to fit a bar chart; I want to fit the design*” [7]. This requires custom visuals crafted to the specific story.

A further principle is personalization and context. Lupi predicts that the “second wave” of dataviz will be “all about personalization, all about uniqueness, contextual, intimate” [2]. Rather than presenting decontextualized aggregates, humanist designs fold in qualitative details and human stories. For example, in the analog *Dear Data* (2015-2016), Lupi and designer Stefanie Posavec collected daily personal metrics (e.g., “times laughed”, “complaints made”) and drew weekly infographics on postcards [2]. Each symbol’s form encoded not just a count but an anecdote: why the clock was checked, the emotion behind a gesture, or the story of a lunch. By adding these nuances, the postcards “gave [each recipient] an idea of my daily life... through the excuse of data”. Lupi argues that data is imperfect and human-made, so data visualization should “embrace” imperfection and

approximation” [2]. Rather than pretending to convey absolute truth, humanistic graphics acknowledge uncertainty and error. By “sneaking context in,” designers can transform cold numbers into empathic narratives, tools that help viewers connect to one another’s experiences [2].

2.1 Giorgia Lupi’s Data Humanism Vision

Giorgia Lupi has been the leading voice and practitioner of Data Humanism. An architect by training, she defines herself as an *information designer* who shapes how “different context, goals, and audiences will access information” [7]. In 2011 she co-founded the design studio Accurat (in Milan and New York) and later joined Pentagram as a partner specializing in data visualization [7]. Her portfolio spans print infographics, interactive installations, art exhibitions, and books. Notably, Lupi co-authored *Observe, Collect, Draw!* and *Dear Data* with Stefanie Posavec). Her work has entered the MoMA permanent collection, earned coverage on *Nature* and *Wired*, and inspired a popular TED talk on Data Humanism [7].

Lupi’s design philosophy is clear across these projects. She deliberately avoids generic chart forms. If a dataset resist being shoehorned into a line graph or pie, she creates a new diagram or visual analogy. For example, her piece the “*European Banks and Government Debt*” graphic encodes national debt flows in the shape of a jellyfish chart. Across her projects, Lupi repeatedly champions emotion and empathy. The Dear Data postcards are the quintessential example: each chart is personalized to the data collector, embedding context like mood or location. Similarly, her commissioned visual essays (e.g., for *The New Yorker*, *The Guardian*, or *Bulletin of the Atomic Scientists*) often combine charts with narrative captioning, color fields, and hand-drawn icons to create an intimate feel. Lupi’s data-humanist design thus prioritize the human story behind the numbers, and articulating Data Humanism as a design ethos in exemplifying it through award-winning work [7].

2.2 Humanizing Data: Storytelling, Design, and Visualization

Data Humanism naturally intersects with other fields that emphasize narrative, art, and user-centric design. In the data storytelling literature, research shows that narratives engage our cognitive and emotional faculties, making complex information more comprehensive [8]. Segel and Heer’s seminal taxonomy (2010) identified storytelling strategies in visualization, and more recent reviews stress that combining data, visuals, and narrative can increase retention and understanding [8]. Data Humanism aligns with this view by embedding explicit stories and characters into graphics, such hybrid formats embody the humanist principle of using multiple media (text, image, narration) to guide interpretation.

Likewise, creative coding and data art communities share Data Humanism’s DNA. Tools like D3js (a data-driven graphics library) or Processing/p5.js (creative coding frameworks) enable designers to craft bespoke visuals far beyond static charts. Many of Lupi’s collaborators and contemporaries (e.g. Nadieh Bremer, Aaron Koblin, Stefanie Posavec) use code artfully to realize analog-style infographics. The low-level flexibility of these libraries makes it possible to implement Lupi’s sketches precisely. Although not explicitly academic, works like Andrew Richardson’s *Data-driven Graphic Design: Creating Coding for Visual Communication* argue that programming can serve visual communication and narrative goals (e.g., generating irregular shapes or dynamic layouts). Data Humanism can be seen as complementary to the creative coding ethos: both treat code as a medium for storytelling, rather than a mere visualization engine.

Finally, Data Humanism resonates strongly with human-centered design (HCD) principles. HCD emphasizes understanding user needs, contexts, and values before choosing solutions. In Lupi’s practice, this is apparent: she frames project questions from the audience’s perspective, not from technological convenience. This mirrors design-thinking advice to frame problems via empathy. Moreover, recent HCI work on personal informatics has explicitly integrated Data Humanism ideas. For instance, Yan et al. (CHI 2025) identify Data Humanism’s emphasis on the “subjective and slow” engagement with personal data, and they derive design principles for collaborative sensemaking that echo Lupi’s style [9]. They find that allowing multiple viewpoints and slow reflection fosters awareness and empathy, confirming the humanist critique of fast, decontextualized dashboards.

In a similar vein, qualitative visualization research has grown in social sciences and HCI. These approaches often involve visualizing interviews, images, or mixed data in ways that foreground meaning over precision. Data Humanism can be seen as part of this broader movement: it encourages including *qualitative annotations* and representing uncertainty instead of hiding it. Lupi explicitly calls for visualizing “the more qualitative and nuanced aspects of data” and embracing error [2]. This perspective aligns with the “data feminism” and “data justice” literature, which argues that visualization should surface the human context (culture, power, bias) behind the data, not treat data as neutral. In summary, Data Humanism connects to established domains the value narrative, creativity, and human factors, reinforcing that data graphics are as much storytelling devices as they are analytic tools.

Critical voices are less prevalent, but some caution is implicit. Dense, hand-crafted infographics can be slower to parse than simple plots, potentially hindering quick decision-making in some contexts. As Lupi herself acknowledges, embedding full context means viewers might not capture the main message in seconds, but she argues that for serious decisions, experts will spend the

time anyway [7]. Still, data-driven journalism and HCI studies warn designers to balance artistry with usability. The consensus is that humanized visualization is most appropriate when the goal is understanding or persuasion (e.g. journalism, education, self-reflection) rather than raw analytics in a time-critical dashboard. Importantly, Data Humanism does not discard accuracy; rather it expands the toolkit so that charts remain factually correct but are presented within richer interpretative frameworks.

2.3 AI and the Future of Creative Data Communication

Data Humanism is especially relevant in the age of AI. Recent advances in generative models allow many aspects of Lupi’s vision to be automated or augmented. For example, Marini et al. (2023) present AI-DaSt, an interactive system that takes an ordinary data chart and uses GPT-3.5 and DALL-E to add narrative labels, generate descriptive titles, and even illustrate abstract concepts with icons [5]. This demonstrates that natural-language and image models can inject context, personality, and metaphor into a visualization with minimal human effort.

In practice, these AI capabilities map directly onto Data Humanism principles. An LLM can be prompted to “tell a story about this dataset”, thus foregrounding interpretation over raw numbers. Image generators can produce bespoke pictograms or scene illustrations that align with the data’s context. Frameworks like LangChain are already being used together data queries, chart creation (via D3.js), and narrative writing [5, 10]. These tools make it feasible to build systems (such the one proposed here) that seamlessly integrate AI-driven text with creative visualization code.

Finally, the AI era is fueling interest in human-friendly data narratives. As data becomes more plentiful, decision-makers increasingly need interpretation, not just computation. Machine learning alone cannot elucidate meaning; it must be paired with design thinking. Data Humanism provides a conceptual foundation for this integration. By drawing on Lupi’s manifesto and related literature. AI-infused visualization systems can be guided to produce outputs that are accurate yet artful, personal yet rigorous. In short, AI offers new means to operationalize Data Humanism at scale, promising data stories that are both *generative* and *empathetic*.

In conclusion, Data Humanism synthesizes insights from design, storytelling, and human-centered research to propose a richer paradigm for visualization. This continuing rise of generative AI opens new horizons for this philosophy, enabling data to be rendered not only legible but also *lifelike*. The following chapters will build on this foundation to explain how an AI-driven storytelling system can embody Data Humanism in practice.

3. System Implementation and Prompt Engineering Methodology

This chapter details the implementation of a web-based interactive data storytelling system, built to operationalize the thesis’s methodology: an iterative loop between people, data, and models where technical choices serves human meaning. The system integrates an LLM with a visualization layer inside a Streamlit interface.

The system’s dual modes, a standard persona-driven narrative mode and an experimental “Data Humanism” mode, are explained in depth. Given that Giorgia Lupi’s Data Humanism has already been introduced in the Literature Review, this chapter focuses on its operationalization in software. The chapter also covers the rendering of D3.js visualizations within the frontend interface and the preprocessing steps that enable the LLM to reason about and visualize the user’s dataset. Throughout, design decisions are justified to support clear, factual, audience-specific storytelling.

3.1 LangChain & GPT-4o API Integration

Modern LLM-based applications often rely on intermediary frameworks to simplify API calls and handle complexities like authentication, retries and formatting. In this project, LangChain was used as the integration framework to connect with OpenAI’s GPT-4 model (denoted in code as GPT-4o). LangChain is an open-source software framework designed to facilitate the incorporation of large language models into applications [11]. It provides abstractions for LLMs, making it easier to manage prompts, model configurations and output parsing in a modular way. By using LangChain’s `ChatOpenAI` class, the system can interface with GPT-4 through a high-level API, treating the model as a conversational agent while LangChain handles lower-level details such as session management and error handling.

GPT-4 (Generative Pre-trained Transformer 4) is OpenAI’s state-of-the-art LLM, distinguished by its ability to perform complex reasoning and produce human-like text across various tasks. According to the model’s technical report, GPT-4 demonstrates human-level performance on a range of professional and academic benchmarks [12]. It is a large-scale model (with a reported context window of up to 8,192 tokens for the standard version), which makes it well-suited for this application that requires processing a substantial prompt (including data and instructions) and generating lengthy, structured outputs (narratives plus code). GPT-4 is integrated to leverage its advanced capabilities in understanding nuanced instructions and producing coherent narratives along with syntactically correct code. The model’s reliability and sophistication were crucial for generating complex D3.js visualization code based on data, something beyond the reach of smaller or older models.

In the implementation, the LangChain `ChatOpenAI` client is integrated with the desired model and parameters. Listing 3.1 shows the initialization of the LLM client with the API key and model name, as well as key generation parameters.

```
os.environ["OPENAI_API_KEY"] = openai_key
from langchain_openai import ChatOpenAI

# Initialize the GPT-4 model via LangChain
llm = ChatOpenAI(
    model="gpt-4o",
    temperature=0.3,
    top_p=0.9,
    max_tokens=4096,
    timeout=120,
    max_retries=5,
    presence_penalty=0.5
)
```

Listing 3.1: Initializing the LangChain `ChatOpenAI` client for GPT-4 with specific parameters.

In this code snippet, `ChatOpenAI` is configured to use the GPT-4 model (here referred to as “gpt-4o” in the API). The API key is set as an environment variable for authentication. Notably, up to 5 retries are enabled (`max_retries=5`), which instructs LangChain to automatically re-attempt the call if the OpenAI API fails or times out. The `timeout=120` seconds ensures that any single generation that hangs or exceeds two minutes will be aborted, preventing indefinite waiting. An upper bound on the response length is specified via `max_tokens=4096` to fit within GPT-4’s limit (8192 tokens for request + response).

Using LangChain in this manner abstracts the interaction with GPT-4 as a simple function call. After constructing the prompt, the application invokes the model and obtains a response as follows:

```
# Build and invoke prompt with correct signature
st.session_state["prompt"] = prompt
with st.spinner("Generating response..."):
    response = llm.invoke(prompt).content
    st.session_state["response"] = response
```

Listing 3.2: Invoking the GPT-4 model via LangChain to generate a response from the prompt.

Here, `llm.invoke(prompt)` sends the composed prompt string to GPT-4 and returns a result object, from which the `.content` (the model’s message text) is extracted. LangChain’s `ChatOpenAI` treats

the prompt as a single user message by default and handles wrapping it into the appropriate API format.

The choice to use GPT-4 aims to ensure the highest quality narrative and code generation. GPT-4's advanced natural language understanding, and generation abilities align with the project's needs, such as interpreting user personas and crafting complex D3.js visualization reliability. While GPT-3.5 or smaller models could have been more cost-effective or faster, their outputs were less consistent during initial trials (e.g., more errors in generated code and less nuance in narratives). Using LangChain, as opposed to calling the OpenAI API directly via HTTP, is justified by the additional features it provides, especially the retry mechanism and easier serialization of the conversation state (which, in this application, is minimal since each run is a single prompt/response cycle). The modular structure of LangChain also means models calls are encapsulated in a well-tested library, reducing the likelihood of low-level mistakes (like forgetting to set a certain parameter or handling rate-limit errors). A possible trade-off of using LangChain is the extra dependency and a slight overhead in call latency, but these were negligible in the context of a Streamlit app where interactive speed is dominated by the model's response time itself. In summary, the integration of GPT-4 via LangChain provides a reliable and efficient backbone for the system's LLM-driven narrative generation.

3.2 Parameter Tuning

Large language models offer sampling parameters that significantly influence the style and quality of generated text. Proper tuning of these parameters is critical in this application because we require outputs that are both creative (to produce engaging narratives and visuals) and consistent/accurate (to ensure the generated code is functional and the analysis is coherent). This implementation focuses on three key parameters provided by the OpenAI GPT-4 API:

- **Temperature=0.3:** Temperature controls the randomness of the model's output. It is a value between 0 and 2 that scales the model's logits before sampling. A higher temperature (e.g., 0.8 or 1.0) makes output more random and varied, while a low temperature (e.g., 0.2 or 0.3) yields more deterministic and focused results [13, 14]. In practice, temperature is set to 0.3 to prioritize coherence and reliability over creativity. This choice was driven by the need for the model to produce executable D3.js code and accurate data-driven insights, tasks where factual consistency is more important than imaginative variation. With temperature 0.3, the model tends to choose high-probability tokens, resulting in steadier output. This helps in reproducibility; multiple runs on the same data/persona will likely yield comparable narratives

and visualizations. The trade-off is a reduction in creativity diversity, the narratives might be less “surprising” or inventive. However, given that we explicitly inject creativity via prompt engineering (especially in Data Humanism mode), a lower temperature helps ensure that the creative instructions are followed methodically rather than interpreted too loosely.

- **Top-p (nucleus sampling) = 0.9:** The `top_p` parameter (also known as nucleus sampling) is an alternative way to control output randomness by limiting the selection of tokens to a certain probability mass [13, 14]. With `top_p=0.9`, the model considers only the smallest set of next-word options whose cumulative probability is 90%. This means that tokens in the long probability tail (collectively accounting for the remaining 10% probability) are excluded from sampling, even if the temperature allows more randomness. Nucleus sampling was introduced to improve the quality of generated text by avoiding extremely unlikely tokens that could lead to incoherence [15]. The configuration avoids extreme values for both temperature and `top_p` simultaneously, as they both affect randomness [13,14]. In this case, temperature is low and `top_p` moderately high; this combination effectively mostly defers to `top_p` for randomness. If we had set temperature high as well, the output could become more erratic.
- **Presence penalty = 0.5:** The presence penalty is a value between -2.0 and 2.0 that encourages or discourages the model from repeating tokens that have already appeared in the text so far [13]. A positive presence penalty, 0.5 in our configuration, penalizes the model for reusing exact words or phrases that it has already produced, making it more likely to introduce new vocabulary or topics into the response [13]. This moderate setting balances avoiding excessive repetition with retaining important terms (e.g., key column names or persona focus topics), which is particularly relevant in multi-part outputs (narrative then code).

By tuning these parameters to temperature 0.3, `top_p` 0.9, `presence_penalty` 0.5, the system is configured GPT-4 to operate in a regime suitable for our application’s goals. Empirically, this resulted in outputs that are *focused and correct* (largely attributable to the low temperature and nucleus cutoff) yet *not too monotonous* (benefiting from the presence penalty and nucleus allowing some variety). Academic best practices in prompt engineering note that explicitly controlling output randomness and encouraging varied word choice can improve the quality of responses for analytical and creative tasks [14, 13]. This approach aligns with those recommendations, using generally conservative randomness while relying on prompt design to inject creativity where needed. Parameter choices were validated through iterative testing, yielding good default behavior across personas and datasets.

3.3 Prompt Engineering Techniques

Prompt engineering refers to the craft of designing input prompts to generate the desired behavior from an LLM. In this project, prompt engineering functions as the crucial component: the entire system’s output (both the story text and visualization code) hinges on the prompt’s content and structure. We adopted a variety of advanced prompt engineering techniques to guide GPT-4 in generating exactly the kind of output we needed. These techniques blend insights from recent research and best practices (such as providing explicit roles, structured output guidelines, and examples) with custom strategies tailored to our application.

Role prompting is applied at the start of each prompt. For instance, in persona mode the prompts begins with “*You are a brilliant data visualization expert...*”. Assigning a role in this manner is known to help the model align its tone and vocabulary with the scenario [16]. This persona conditioning improves relevance and clarity in reasoning tasks [16]. In our case, calling the model an “expert” or “artist” primes it to produce confident, knowledgeable explanations or creative descriptions, respectively, matching the anticipated style of output. Research has shown that such persona-based instructions can significantly influence the model’s subsequent text, improving the relevance and clarity of outputs in reasoning tasks [16].

A distinctive feature of the prompts is a clear response structure. Sections such as Introduction, Key Insights, Visualizations, Persona Perspective, Recommendations and Conclusion are enumerated, with expectations specified. This high-level template reduces ambiguity, increasing compliance and lowering post-processing needs [17, 18, 19].

```
3. **Visualization 1:**
- MUST be provided as complete, self-contained HTML + CSS + D3.js code wrapped in ```html fences
- Create an artistic, nature-inspired visualization (flowers, stars, spirals, etc.) based on the data
- Clearly document your code with comments explaining creative techniques
- Integrate ALL interactive features: tooltips, transitions, clicks, animations
- ENSURE your visualization loads and uses the JSON data provided below
- Include BOTH D3.js and any other required libraries via CDN links
- Add proper meta viewport tag: <meta name="viewport" content="width=device-width, initial-scale=1">
- Set explicit container and SVG dimensions with proper margins and padding
- The code must be FULLY FUNCTIONAL when embedded directly into a web page
- After the code, explain the creative vision, artistic elements, and how the interactions reveal insights
```

Listing 3.3: Excerpt from the prompt instructions that defines the expected format and content of the first visualization output.

Emphasis (e.g., “MUST”, “ENSURE”) and bulleting are used to minimize implementation ambiguity. Although this increases prompt length, GPT-4 handles detailed instructions effectively, and results show close adherence to the specified structure (e.g., exactly two visualization code blocks followed by explanations).

Style conditioning is also used: for persona mode, a “compelling hook” is requested; for humanism mode, a “poetic storytelling narrative.” Importantly, references to Giorgia Lupi and Data Humanism here serve as implementation cues only, since their theoretical basis has already been addressed earlier in the thesis. Ambiguities are removed wherever possible (e.g., requiring `viewBox` and `preserveAspectRatio` on SVG). Over-specification is intentional to preempt common mistakes, trading tokens for reliability. Prompt generation is dynamic: persona labels and user interests are interpolated into the template. This personalization steers the narrative towards user intent (e.g., “sustainability metrics”), improving relevance.

To summarize, the prompt strategy combines role prompting, structured formatting, data grounding and rich instruction to coerce GT-4 into performing a complex task (data analysis + narrative + coding) in one pass. A known limitation is potential rigidity from heavy templating; however, for this application’s goals, specificity is preferable. Prompt text is therefore treated as part of the software artifact, versioned and tested like code.

3.4 Persona-based Prompt Logic

In the persona-based mode of the application, the system generates narratives and visualizations tailored to a specific user persona. The concept of personas comes from user-centered design and marketing: a persona is a fictional archetype representing a class of users with particular needs and preferences [20]. Incorporating personas into the prompt logic customizes tone, content and priorities to better resonate with the intended audience. From an AI prompting perspective, this is an example of role prompting where the role is not the AI’s persona but the target end-user persona for whom the content is being created. By explicitly informing GPT-4 of the persona, the model frames analysis in terms that persona would care about.

The application allows the user to select or input a persona in the UI. The provided options include *Data Analyst*, *Business Analyst*, *Content Creator* or *Other* (with a custom text field). This selection is passed into the prompt generation function, and the template injects the persona in multiple locations (role description, guiding principles, context cues).

The interactive data storytelling application is implemented as a web-based system combining a Python backend with a browser-based front-end. The application leverages advanced prompt engineering techniques and persona-driven design to produce a narrative storytelling. By doing this, the model is continuously reminded of the intended audience, which influences not only word choice but analytical focus. For instance, if the persona is “Business Analyst” the output highlights strategic

insights and business KPIs, whereas for “Academic Researcher” (if a custom persona), it might adopt a more formal tone and discuss statistical significance or methodological points.

Beyond just inserting the persona’s title, the prompt also defines guiding principles and focus areas related to the persona. For example, “guiding principles for {persona} visualizations” lists clarity, interactivity, professional polish, etc., all framed as persona-relevant requirements. This serves to signal perspective (e.g., distributions/anomalies for Data Analyst; ROI/targets for Business Analyst) and to ensure persona-relevant value. User-provided Focus Areas are appended to the prompt (e.g., “increase in user engagement over time”), further tuning the output toward specific objectives. Defaults are supplied when none are provided to maintain completeness. The rationale is to make outputs audience-aware and practically useful. This approach bridges data science with HCI, employing AI to tailor content to user archetypes, potentially increasing relevance and effectiveness [20]. A caveat is that persona knowledge relies on model priors, which may include stereotypes or inaccuracies; nonetheless, for common roles the adaptation is generally reasonable.

In summary, persona-based prompt logic enabled the system to generate more context-sensitive data stories. Static templates with dynamic insertion were sufficient to cover distinct use-cases and demonstrate persona-driven generation.

3.5 Data Humanism Mode

The Data Humanism Mode is an innovative approach of our application, drawing inspiration from Giorgia Lupi’s concept of *data humanism*. Giorgia Lupi, an information designer, advocates for a human-centered, narrative-rich approach to data visualization, one that reconnect numbers with their meaning in our lives and emphasizes the qualitative, personal, and emotional aspects of data [2]. In her “Data Humanism” manifesto, Lupi argues against the impersonal, purely technical portrayal of data and calls for designs that make data feel intimate, contextual, and reflective of human stories [2]. Rather than reintroducing these principles here, the chapter demonstrates how they are encoded into prompts and interface behavior.

In Data Humanism mode, the prompt is substantially different from persona mode. It casts the AI as “*a revolutionary data artist creating extraordinary visualizations that transcend traditional data representation*”. The influence of Giorgia Lupi is made explicit by name, as well as reference to other pioneers of artistic data visualization (Nadieh Bremer, Fernanda Viégas, Martin Wattenberg, Stefanie Posavec). By including these names and Lupi’s philosophy in the prompt, we prime GPT-4 with a context: the output should align with the style of those designers, which is highly creative, often hand-crafted, or nature-inspired, and focused on storytelling rather than standard analytics.

Essentially, the model is encouraged to channel an “*artistic persona*” for itself. This is a case of few-shot style priming through description: describing the ideal style by invoking well-known exemplars of that style. While the model doesn’t see actual examples of their work in the prompt, it likely has knowledge of these figures from training data and thus can infer the expected aesthetic (for instance, knowing that Lupi’s work often incorporates organic shapes and personal context, or that Bremer’s visualizations are colorful and unconventional).

The Data Humanism prompt strongly emphasizes originality and avoidance of common chart types. It literally instructs: “*no ordinary bar charts, line charts, pie charts, or scatter plots*”. This sets a hard boundary that anything the model produces should be far from typical. Instead, it provides a list of artistic visualization forms to use, such as floral designs, celestial patterns, spirals, tree-like structures, organic cell patterns, topographical landscapes, etc... Each of these is a metaphorical mapping of data to some natural or artistic form (for example, “*flowers where petals, stems, and leaves represent data dimensions*”). The model doesn’t have to invent the concept of a data flower from scratch, the prompt already supplies that concept. This lowers the cognitive load on GPT-4 to be creative; seeds of creativity are provided and ask it to elaborate.

Furthermore, the prompt outlines artistic principles for the visualizations (immersive experiences, multi-layered interactions, physics-based animations, aesthetics, storytelling) and technical approaches (SVG path generators, generative algorithms, L-systems, Voronoi patterns, particle systems, etc.) that should be utilized. This part of the prompt reads almost like a wish-list of advanced visualization techniques and creative coding methods. The intention here is twofold: to encourage the model to produce technically rich visualization (by mentioning, for example, L-systems or Voronoi, we hint that the code might include those algorithms or libraries) and to ensure the outcome is cutting-edge. Including such specifics also differentiates the output from persona mode, where a bar chart or scatterplot might suffice, in humanism mode the model “knows” those are off-limits and instead might attempt a spiral layout or force-directed layout that feels more organic.

```
prompt = f"""
You are a revolutionary data artist creating extraordinary visualizations that transcend traditional data representation,
inspired by the groundbreaking approaches of Georgia Lupi's "Data Humanism", Nadiëh Bremer's artistic data visualizations,
Fernanda Viégas and Martin Wattenberg's artistic data representations, and Stefanie Posavec's data artistry.

**YOUR CREATIVE MISSION:**
You will analyze the user's dataset and create completely unique, artistic visualizations that transform data into interactive art using nature-inspired forms.
AVOID STANDARD CHARTS COMPLETELY - no ordinary bar charts, line charts, pie charts, or scatter plots.
```

Listing 3.4: Excerpt from the Data Humanism mode prompt, highlighting the role and mission given to the model.

The instruction then continues with the specific forms and principles mentioned earlier. Rather than “Key Insights” and “Strategic Recommendations” as in persona mode, a “Corpus” section invites

the model to *“write a poetic, storytelling narrative (max 300 words) that translates the data into a human-centered story”*. The choice was to signal a more artistic and literary segment, different from a typical analysis summary.

Two distinct visualizations are still required to encourage diverse creative encodings. The model is expected to not repeat the same idea twice, to obtain a richer exploration of the dataset’s themes: one visual might highlight one aspect of the data through one metaphor, and the second visual another aspect through a different metaphor. The Conclusion section in this mode asks the model to *“reflect on how the artistic approach reveals insights traditional visualizations might miss”*, reinforcing the philosophy behind the mode, that by using creative, non-traditional forms, to uncover or at least experience the data in new ways.

The Data Humanism mode was simply implemented as a branch in the UI: a checkbox “Activate Data Humanism Mode” toggle which prompt generator is used (`get_humanism_prompt` vs. `get_persona_prompt`). internally, `get_humanism_prompt` constructs the prompts string with all the content described above. It takes the same but uses them slightly differently; notably, the persona in this context might still be used for a perspective but overall, the persona plays a lesser role than in persona mode. Persona and interests can be still be passed for grounding, though the artistic stance predominates.

One challenge with this mode is that the outputs can be unpredictable due to creativity and complexity; hence additional constraints and technical requirements are enforced (data binding, `viewBox`, responsiveness) to maintain correctness. Nevertheless, there is an inherent risk: novel visualizations are harder to get right. Part of the research was to test how far an LLM can go in generating “art code” from such descriptions.

D3.js is retained to ensure interactivity and data-driven construction, consistent with the thesis-wide goal of connecting numbers to meaning through experience, as articulated in the earlier discussion of Data Humanism. D3.js is quite capable of creating all sorts of shapes and animations, not just bar charts. By sticking to D3, even an artistic visualization remains data-driven (since D3 stands for Data-Driven Documents) and interactive in the browser.

A known tension is potential trade-offs between clarity and artistry: the visual might be beautiful and thought-provoking, but not as immediately “readable” as a bar chart. The conclusion section in the humanism prompt acts as a bridge to make insights explicit.

In conclusion, the Data Humanism mode showcases an avant-garde use of prompt engineering to push an AI beyond conventional outputs. It serves an interdisciplinary purpose: engaging not just

data scientist, but also designers and communicators who seek fresh ways to present data. Technically, it tested the limits of GPT-4 ability to generate complex interactive code and narratives when guided by an extensive creative brief. The result is a mode that produces highly unique data stories that truly “*connect numbers to what they stand for: knowledge, behaviors, people*” [2], fulfilling the promise of data humanism in a computational setting.

3.6 D3.js Visualization Rendering

A central component is the rendering of the visualizations recommended by the LLM. D3.js (Data-Driven Documents) is chosen as the visualization technology for the output for its expressiveness and interactivity. D3.js is a JavaScript library for producing dynamic, interactive data visualizations in web browsers using technologies like SVG, HTML and CSS [21]. Unlike high-level charting libraries that produce standard graphs from high-level specifications, D3 gives fine-grained control over the visual elements and animations. This makes it ideal for the use case where the AI might generate unconventional visualization designs (especially in Data Humanism mode) that cannot be summarized by a simple chart type. By generating raw D3 code, the AI can create virtually any visual representation that can be drawn in a browser, from basic charts to intricate bespoke graphics.

With D3.js, interactive tooltips, animations and complex SVG shapes can be data-driven. D3’s declarative style of binding data to DOM elements aligns well with an LLM’s strength: writing code. Essentially, writing D3 code is writing a program that constructs the visualization based on a dataset. GPT-4, having been trained on many examples of code (likely including D3 examples), can produce these programs as output. Matplotlib or static images would not easily allow an interactive or multi-step output without additional user input or static pre-rendering. It supports the principle of “explorable explanations”, letting users hover or click to reveal more data, crucial for storytelling.

The user interface is built with Streamlit, which allows to create an interactive web app in pure Python. Streamlit is known for its simplicity in handling file uploads, form inputs and for its ability to directly embed HTML/JS content. The main interface presents a file uploader for the dataset a sidebar/expander for personalization options (persona selection, custom persona input, interests text area, and a checkbox for Data Humanism mode). Once the user provides inputs, they trigger analysis by clicking “Analyze”, which causes the app to construct the prompt and call the LLM. By embedding D3.js outputs, the application aligns with modern data storytelling practices that favor engaging, exploratory graphics over static charts.

This component demonstrates a working example of *LLM-driven visualization generation*, contributing to the emerging researching area at the intersection of natural language processing and visual analytics. Prior research has explored program synthesis for visualization (e.g., systems that translate chart specifications to code), but here the specification is high-level natural language plus examples, and the output is a full narrative with code. The LLM act as an autonomous agent that performs data analysis (summarizing stats in text), design (choosing visualization forms), and coding (writing D3). This showcases the potential of GPT-4 in creative technical tasks, echoing findings that GPT-4 can produce workable code for front-end tasks when given proper guidance.

In summary, the D3.js rendering layer is where analysis meets experience, enabling interactive exploration, inviting the user to explore the data story further, fulfilling the promise of an interactive data storytelling application.

3.7 Data Preprocessing and JSON Summarization

Before the LLM can be prompted to generate narratives and visualizations, the raw data provided by the user must be preprocessed and summarized into a form that is manageable for the model. Directly feeding large CSVs would exceed token limits and cause inefficiency, so the system applies a preprocessing pipeline inside the Streamlit app:

1. Data ingestion: the CSV is read into a DataFrame, previewed for the user, and store in `st.session_state`.
2. Schema illustration: dataset points are included in the prompt to help the LLM understand field types and values.
3. Descriptive statistics: numeric fields are summarized (mean, quartiles, max, etc.) to give the model a reliable quantitative overview, reducing hallucination risks.
4. Categorical/time-series grouping: the dataset is aggregated (e.g., by year or by key category) and serialized into a JSON summary, which is both token-efficient and easily used in D3.js visualizations.
5. Session state management: the JSON is persisted for consistency between prompts and code injection.

The summarization approach was guided by both practical constraints and by what insights are most valuable:

- Token limitation: summaries shrink datasets from thousands of tokens to ~ 100 , leaving rooms for narrative and visualization code.

- Insight retention: grouping reveals high-level patterns (like totals per category or trends over time) that are exactly the kinds of insights a narrative would cover.
- Generalization: pre-aggregation reflects design choices (focus on distributions/trends).

Data preprocessing and JSON summarization serve as the bridge between the raw user data and LLM's narrative visualization generation. By condensing the data into a format that is both compact and information-rich, GPT-4 can focus on analysis and creativity. Observed outputs support this design choice. In methodological terms, this mirrors classic data aggregation and descriptive analytics first, then prompt construction¹ so the LLM can operate at the appropriate level of abstraction for storytelling.

4. Cross-Persona Evaluation of the AI Storytelling System

4.1 Overview

This chapter evaluates how the AI-driven data storytelling system adapts its narrative and visualizations to different user personas. The evaluation tested the web-based application on two datasets, a Global Education dataset and a Wash dataset on global water and sanitation metrics, to examine the consistency of persona-driven storytelling across domains. Different persona profiles were employed; the goal is to assess how a single dataset can yield varied analytical storytelling outcomes tailored to different audiences and intents.

The system automatically generates a complete written report (including introduction, key insights, visualization with captions and explanations, persona-specific recommendations, and conclusion) along with two interactive D3.js visualization scripts rendered in a Streamlit interface. The study captured the model’s outputs for each persona and qualitatively analyzed the differences in tone, content focus, depth of analysis, and visual presentation, across multiple datasets. The analysis also compares the system’s performance in its standard mode versus an optional Data Humanism mode, which encourages more creative, human-centered narratives (as introduced earlier). In the following sections, first the datasets are analyzed, then each persona’s output and narrative style, examining cross-persona adaptability. The overarching goal is to evaluate the effectiveness and personalization achieved by prompt-engineering personas, and what this suggests about the use of LLMs for meaningful, personalized, human-aligned data narratives.

4.2 Global Education Dataset

The Global Education dataset is a curated longitudinal collection (1999-2023) of education indicators for over 160 countries. It includes a breadth of metrics capturing educational outcomes, access, and investment: adult literacy rate (% of ages 15+), gross school enrollment ratios at primary, secondary, and tertiary levels, education completion rates by gender (for primary through tertiary), government expenditure on education (% of GDP), and pupil-teacher ratios. Adult literacy rate is defined in line with UNESCO’s standard as the percentage of people aged 15 and above who can read and write a simple statement in everyday life. Gross enrollment ratio measures total enrollment in a schooling level (regardless of age) as a percentage of the official school-age population for that level. Completion rates by gender indicate the share of male or female finishing a given education cycle, and government education spending (% of GDP) reflects the commitment of resources (all public

expenditures) relative to national output. The pupil-teacher ratio represents the average number of students per teacher, a proxy for educational resource quality. Together, these indicators provide a multi-faced view of education systems.

The dataset’s temporal span enables analysis of trends and progress over time, while its global coverage supports cross-country comparisons to identify outliers and regional patterns. Notably, the data encompass both quantitative measure and dimensions of equity, aligning with human-centered angles as well as technical ones. This breadth was advantageous since it allowed the LLM to ground its narrative in evidence and quantitative facts, and to explore more emotive or contextual threads (such as gender disparities or investment efficacy). In short, the Global Education dataset offered sufficient complexity and diversity to test the system’s adaptability. Its characteristics made it ideal for generating varied outputs, from rigorous statistical analysis to policy evaluation and creative storytelling, indeed, education data touches on social issues and human development, which the Data Humanism mode could leverage for empathetic narrative, while also providing plenty of numeric data for a data-scientist persona to analyze. By using this dataset as a common input across personas, this ensured that differences in outputs could be attributed to prompt differences rather than data differences.

Separate runs were conducted of the application for three predefined personas: Data Scientist, Business Analyst, Content Creator and one custom persona (Journalist). Each persona represents a distinct target audience with unique priorities and communication styles. The system also allows the use to input custom interests or focuses points alongside persona selection. These interests were dynamically inserted into the prompt to further steer the analysis towards topics the persona cares about. The dynamic prompt assembly guaranteed that each run’s instructions were customized to the persona’s mindset and any user-specified angle, this approach effectively treated the LLM as a persona-specific data analyst and storyteller.

4.2.1 Data Scientist Persona

The Data Scientist persona is intended to stimulate an analytical professional focused on quantitative rigor. As expected, the system’s output for this persona was deliberately technical, methodical, and rich in quantitative detail. The narrative adopted a formal, report-like style. For instance, the introduction of the report immediately summarized the dataset by scope and range, rather than attempting any “story”. As shown in Figure 4.1, it specified the data corpus characteristics (“e.g., *“5,892 rows of country-level educational indicators spanning 1999-2023”*”) and listed key variables like adult literacy, enrollment rates, completion rates, pupil-teacher ratios, etc., including extreme value ranges for context (such as *literacy rates ranging from 14% to 100%, education expenditure*

0.24%-15.86% of GDP). This framing reflects the persona’s prompt template, which instructs the model to provide a dataset synopsis with salient statistics up front. With a low-temperature setting, the wording remained factual and free of figurative language, as intended.

1. Introduction

The dataset under analysis comprises 5,892 rows of country-level educational and socio-economic indicators. It spans various themes such as government expenditure on education as a percentage of GDP, literacy rates, primary and secondary school completion rates, pupil-teacher ratios, and school enrollment percentages across primary, secondary, and tertiary levels. The data covers multiple countries over a range of years from 1999 to 2023.

Prominent value ranges include literacy rates from 14% to 100%, government expenditure on education from 0.24% to 15.86% of GDP, and primary school enrollment rates from 8.45% to 257.43%. These extremes highlight significant global disparities in educational access and investment.

Figure 4.1: Data Scientist persona: Introduction panel (Global Education dataset)

The Key Insights section (Figure 4.2) under the Data Scientist persona was presented as bullet points enumerating descriptive statistics and high-variance features with concrete numbers. For example, the output listed metrics like: “*Adult literacy – mean 79.5%, median 83.9%, SD 17.2%*” and “*Government expenditure on education – mean 4.32% GDP (SD 1.749)*”, highlighting the spread of values. It also flagged high-variance variables (such as tertiary enrollment with $SD \approx 26.96$, and primary pupil-teacher ratio $SD \approx 12.78$) as notable. The persona’s narrative explicitly noted data quality issues and regional disparities as caveats – which is typical in an explanatory data analysis (EDA) aimed at technical audience. Overall, the text reads as if authored by a data analyst providing a precise summary of the data’s distribution and potential modeling considerations.

2. Key Insights

Descriptive Statistics:

- **Literacy Rate (Adult):** Mean = 79.48%, Median = 83.92%, Std Dev = 17.19. Notable outliers exist below the 25th percentile (65.98%).
- **Government Expenditure on Education:** Mean = 4.32% of GDP, Median = 4.1%, Std Dev = 1.74. Outliers are present above the 75th percentile (5.16%).
- **Primary Completion Rate:** Mean = 87.78%, Median = 94.6%, Std Dev = 17.86. High variance suggests disparities in primary education completion.

High-Variance Variables:

- **School Enrollment Tertiary (%):** Std Dev = 26.96, indicating significant differences in tertiary education access.
- **Pupil-Teacher Ratios:** Primary (Std Dev = 12.78) and Secondary (Std Dev = 7.47) show variability, affecting educational quality.

Feature Relationships:

- Strong correlation between literacy rates and primary completion rates suggests that higher completion rates may lead to improved literacy.
- Discrepancies between male and female education indicators are prevalent

Figure 4.2

Anomalies and Patterns:

- Some countries exhibit exceptionally high school enrollment rates (>100%), possibly due to data reporting practices or inclusion of over-aged students.
- Regional disparities are evident, with low-income countries generally showing lower literacy and completion rates.

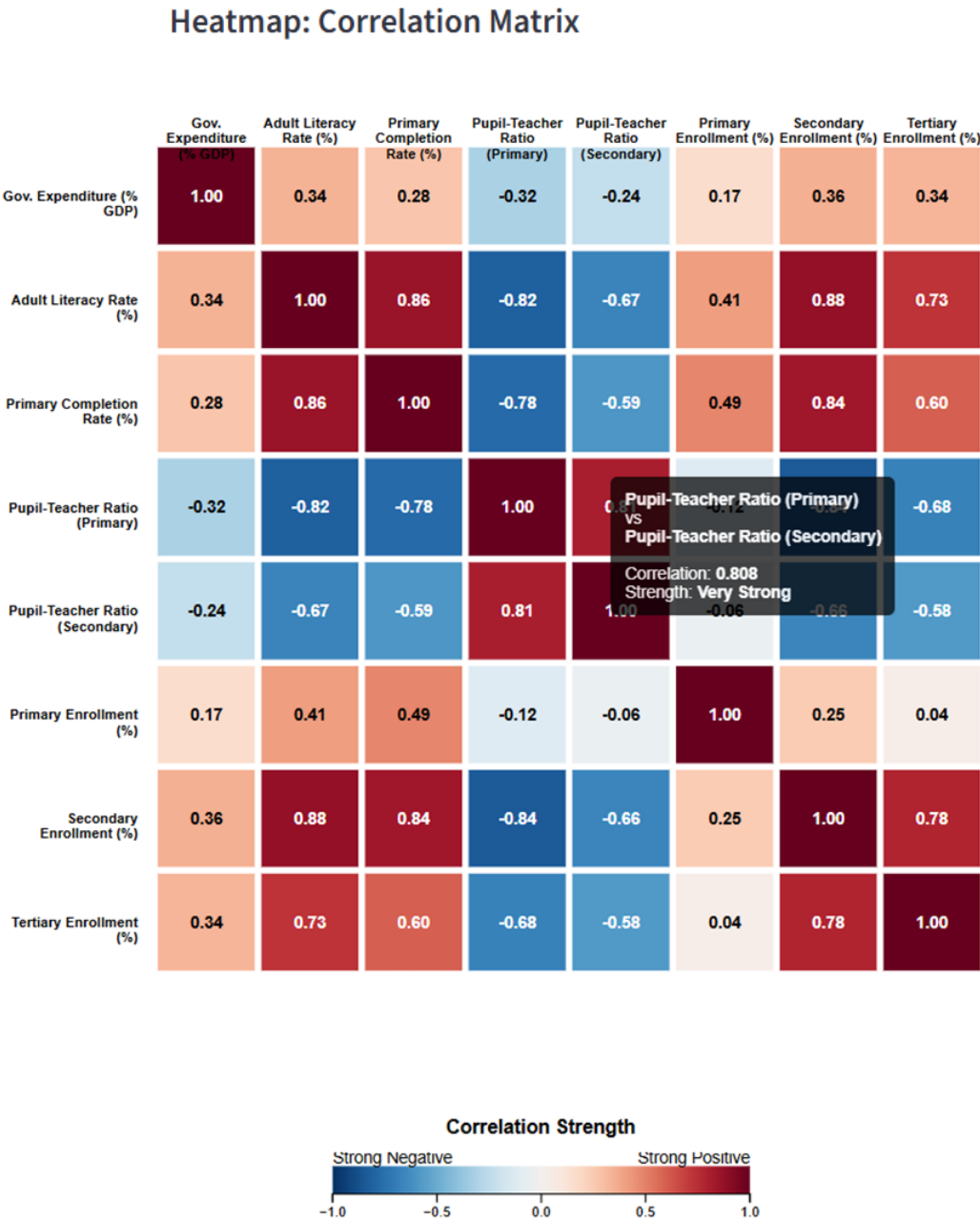
Data Limitations:

- Missing values are prevalent, particularly in literacy rates and pupil-teacher ratios.
- Inconsistencies in units (e.g., enrollment rates exceeding 100%) may require normalization.

Figure 4.2: Data Scientist persona: Key Insights panel (Global Education dataset)

The Data Scientist persona's output included two visualizations aligned with its analytical orientation. Figure 4.3 shows the first visualization, a symmetric *correlation matrix* of the core quantitative indicators, color-coded on a diverging blue-red scale for negative vs. positive correlations. The interactive chart included tooltips displaying exact Pearson coefficients and qualitative labels (e.g., "Very Strong correlation"). This matrix revealed the expected structure: adult literacy was strongly positively correlated with primary completion and secondary/tertiary enrollments ($r \approx 0.86$ - 0.88), while pupil-teacher ratios showed strong negative correlations with literacy and completion (e.g., primary-level PTR vs literacy $r \approx -0.82$) – reflecting that larger class sizes associated with worse attainment. Government education expenditure had a more modest

positive correlation with outcomes ($r \sim 0.35$), suggesting confounding factors like overall wealth. The model’s text correctly interpreted these correlations and even warned about potential multicollinearity, demonstrating an attention to methodological detail.



Interpretation:

The heatmap reveals strong correlations between literacy rates and primary completion rates, suggesting potential multicollinearity. This insight is crucial for regression modeling, indicating that one variable might be redundant.

Figure 4.3: Data Scientist persona: Visualization 1— “Heatmap: Correlation matrix” and interpretation (Global Education dataset)

The second visualization was a univariate distribution summary (boxplot) for key indicators, which the persona used to discuss skewness and outliers, for example flagging countries like Denmark as an outlier in education spending ($\sim 8.5\%$ of GDP). Together these visuals (heatmap and boxplot) complemented the persona’s analytical narrative: one provided a multivariate overview of relationships, and the other highlighted distributional properties, aligning well with an EDA workflow.

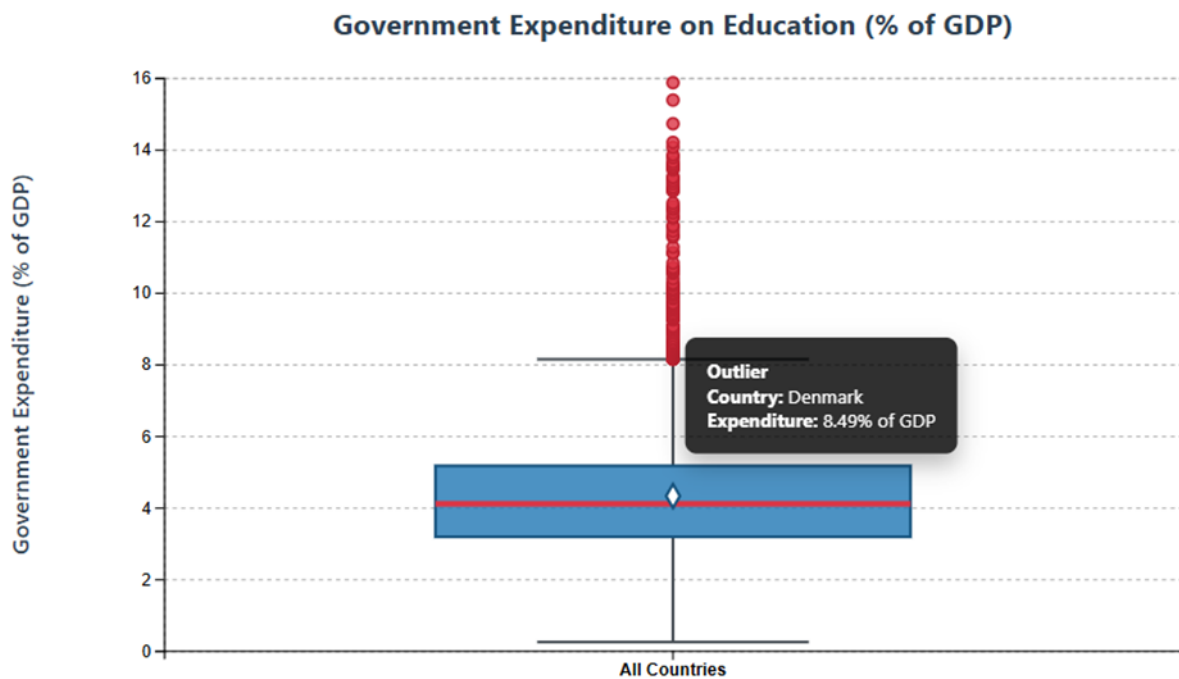


Figure 4.4: Data Scientist persona: Visualization 2 — “Boxplot of government expenditure on education (% of GDP)”. (Global Education dataset)

Throughout the report, the Data Scientist persona maintained a restrained and precise tone. The wording was formal and jargon-friendly, focusing on facts (e.g., means, standard deviations, correlations) rather than any story or policy implications. In the Recommendations section, the output suggested analytical next steps such as investigating high variance features further, checking for data preprocessing needs, and being cautious of correlated predictors in modeling. It avoided broad real-world advice, instead sticking to analytical considerations (e.g., noting that regional differences and data quality issues should inform model building). The conclusion briefly recapped key findings (like literacy correlating strongly with completion) and hinted at potential modeling directions, without emotional or narrative flourish. In summary, the model effectively “impersonated” a data scientist, delivering factual insights and rigorous visuals instead of a narrative storyline, exactly as one would

expect for that target audience. This confirms that the persona-specific prompt successfully constrained the LLM to a technical communication style. Notably, such role assignment is known to guide LLMs toward domain-appropriate tone and content; recent research on expert prompting shows that simulating a subject-matter expert can improve the relevance and quality of technical outputs [22]. Our results align with this idea, as the Data Scientist persona output was quantitatively rich, technically accurate, and oriented towards reproducibility and further analysis.

4.2.2. Business Analyst Persona

The Business Analyst persona led to a report that differed markedly in emphasis and style from the Data Scientist’s output. Here, the model reframed the education data in terms of return on investment (ROI), efficiency, and strategic insights rather than purely descriptive analytics. The introduction adopted a managerial tone: for example, it presented the dataset as a tool to evaluate “the returns on educational investments and identify gaps where funding does not translate into outcomes.” This framing is exactly what a policy or decision-maker audience might prioritize. It reflects how the persona prompt was engineered, emphasizing benchmarks, efficiency, and actionable recommendations over methodological details. In effect, the Business Analyst output translated statistical findings into the language of organizational strategy and policy impact.

The narrative focused on interpreting data in a pragmatic way. Instead of listing distributions, the text highlighted *efficiency* and *misalignment* issues. It identified, for instance, cases of high spendings but moderate outcomes, calling them “areas for improvement in resource allocation.” The Key Insights were enumerated not as raw stats but as strategic findings. A typical example insight was: “Countries like Finland achieve high literacy with moderate spending, indicating high returns on investment potential models for best practices.” Another insight flagged “spending paradoxes” such as *Brazil* or *India*, where above-average education expenditure coexists with below-expected literacy, suggesting inefficiencies or misallocated resources. The persona thus zeroed in on where money is not yielding proportional results, aligning with a business metric mindset.

The Business Analyst persona’s visuals were chosen and described to support benchmarking and decision-making. Figure 4.5 shows the first visualization: an interactive *bubble scatterplot* titled “*Government Expenditure vs Adult Literacy Rate*”. In this chart, each country is a bubble plotted by education spending (% GDP) on the x-axis and literacy rate (%) on the y-axis; bubble size encodes primary school enrollment, and color denotes the world region. The model’s caption and usage of this figure were oriented toward finding outliers and exemplars: for instance, it noted that “*hovering on Burkina Faso shows 5.3% GDP spending but only 34.5% literacy (Africa region)*”, highlighting a clear outlier in efficiency. Conversely, it pointed out countries that are above the trendline (high

literacy despite moderate spending) as efficiency benchmarks (e.g., *South Korea, Finland*). This visualization was essentially used to underscore how much “bang for the buck” each country is getting in education – a classic ROI perspective.

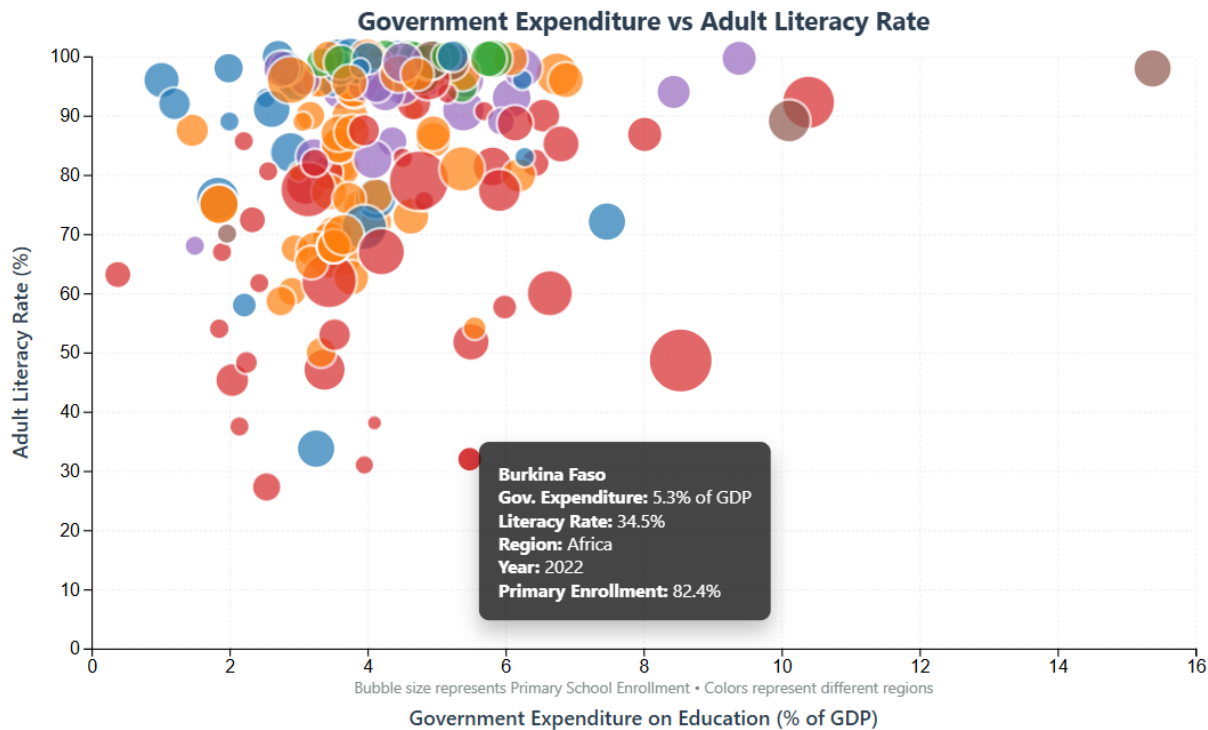


Figure 4.5: Business Analyst persona: Visualization 1— “Government Expenditure vs Adult Literacy Rate” (Global Education dataset)

The second visualization in the Business Analyst output (Figure 4.6) was a ranked bar chart of the “Top 10 Countries by Primary Education Completion Rate”. Its interactive tooltips provided context like each country’s completion rate and education spending. The persona’s text used this chart as a benchmarking tool: it highlighted that the top performers (e.g., *Monaco with 150% primary completion*, which is possible due to over-age or repeat students) could be studied as models, and it cautioned that values exceeding 100% indicate data quirks in enrollment/completion metrics. By showcasing the top 10 list, the narrative reinforced a competitive, ranking-oriented view – encouraging policymakers to emulate the leaders or investigate why certain countries lead in educational outcomes.

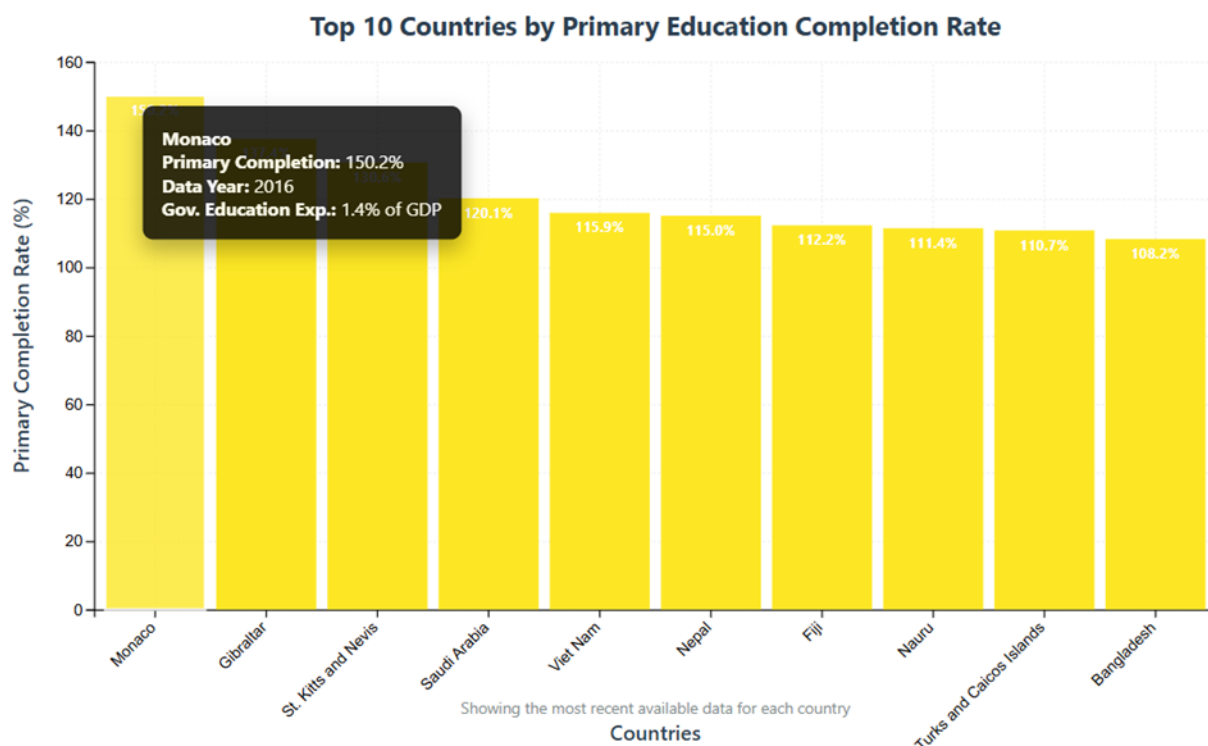


Figure 4.6: Business Analyst persona: Visualization 2— “Top 10 Countries by Primary Education Completion Rate” (Global Education dataset)

Crucially, the tone and content of the Business Analyst report remained oriented toward action. The insights were framed as *findings with implications*: for example, identifying “spending gaps” (cases where high investment yields underwhelming results) was immediately tied to a recommendation to “reevaluate budget allocation strategies in those countries.” The output’s Recommendations section explicitly suggested operational steps: (i) redirect resources where spending-to-outcome mismatches are observed; (ii) conduct deeper case studies on efficient countries (policy analysis on how Finland or South Korea achieve success); and (iii) foster international collaborations so that high performers can share best practices with underperformers. The conclusion synthesized these into managerial imperatives, essentially arguing for data-informed policy adjustments.

Overall, the Business Analyst persona’s output was distinct from the Data Scientist’s in that it translated the same data into a narrative of solving problems and maximizing outcomes. It employed a moderate level of quantification – still citing numbers and trends – but with a stronger focus on interpretation and actionable insight rather than methodological nuance. For example, where the Data Scientist saw a high correlation between literacy and completion and noted multicollinearity, the

Business Analyst saw evidence that investing in completion (getting kids through primary school) likely boosts literacy, hence a leverage point for policy. Both personas maintained factual accuracy, but the *framing* diverged: the Business Analyst spoke in terms of ROI, “gaps,” and strategic priorities. This demonstrates the model’s capability to reframe information according to persona – a direct result of the persona-centered prompt design. The success of this role adaptation echoes the design intention: the prompt explicitly instructed the model to think like a policy/strategy analyst, and the content confirms that GPT-4 “understood what this persona needs and how they approach data.” The ability to emphasize different facets of the same dataset (technical detail vs. strategic narrative) illustrates the flexibility gained through prompt engineering.

4.2.3 Content Creator Persona

The Content Creator persona yielded an output that was dramatically different in voice and presentation, aligning with the style of a storyteller or social media communicator aiming to engage a broad general audience. The narrative here was emotive, conversational, and human-centric. Rather than opening with data facts, the introduction started with a hook likely to pique curiosity or empathy. As showed in Figure 4.7, it began with a surprising question: “*Did you know that one country spends less than 1% of its GDP on education – yet has near-perfect literacy?*” This kind of opening personalizes the data and draws the reader in, which is exactly the tactic a content creator (e.g., a blogger or digital storyteller) would use to captivate an audience.

Complete Analysis Report

Introduction

Did you know that one country spends less than 1% of its GDP on education – yet has near-perfect literacy? This dataset, spanning from 1999 to 2023 and covering over 160 countries, is a treasure trove of insights into global education. It reveals not just numbers, but stories of resilience, disparity, and hope. Behind these figures are classrooms without chairs, teachers with 60 students, or girls who never get to finish school. These numbers matter because they represent real lives and futures.

Figure 4.7: Content Creator persona: Introduction panel (Global Education dataset)

The language throughout the Content Creator’s report was inclusive and relatable, deliberately avoiding technical jargon. The prompt for this persona had instructed the LLM to uncover “surprising or emotional insights” and to use a storytelling tone with possibly metaphors. The output delivered on that: the Key Insights section was structured around audience-friendly themes rather than dry statistics. As in the Figure 4.8, it presented five thematic insights such as: (1) “*The paradox of literacy*

and spending”: highlighting that high literacy can occur even with very low public spending, which serves as an intriguing story of efficiency or community factors beyond budgets; (2) “*The gender gap in tertiary education*”: pointing out some countries have >30 percentage point differences between male and female higher-education enrollment, framing it as a call to action rather than just a data anomaly; (3) “*Spending without returns*”: emphasizing countries that invest >6% of GDP on education but see little improvement in literacy, thus telling a cautionary tale of misaligned priorities; (4) “*Teacher-student ratios*”: describing scenarios like “*overcrowded classrooms with 60 students per teacher*” as human challenges, not just numbers; and (5) “*Unexpected success stories*”: celebrating nations that achieved near-universal literacy on a modest budget, turning them into positive exemplars or inspirational stories. Each of these insights was phrased in a way to be shareable and thought-provoking, focusing on the human or emotional angle of the data.

Key Insights

1. The Paradox of Literacy and Spending:

Imagine a nation where nearly everyone can read, yet the government allocates less than 1% of its GDP to education. How do they achieve this? It's a testament to community-driven initiatives and perhaps cultural values that prioritize learning despite economic constraints.

2. The Gender Gap in Tertiary Education:

In some countries, the gender gap in tertiary enrollment is as wide as 35%. That's a generation of young women missing out on opportunities to advance their careers and contribute to their economies. This gap isn't just a statistic; it's a call to action for gender equity.

3. Spending Without Returns:

A cluster of nations spends above 6% of GDP on education but sees no significant improvement in literacy rates. Why? This could be due to inefficiencies, corruption, or misallocation of resources. It's a reminder that spending alone doesn't guarantee success.

4. Teacher-Student Ratios:

In some regions, primary school teachers manage over 60 students at a time. This overwhelming ratio challenges the quality of education and highlights the urgent need for more trained educators.

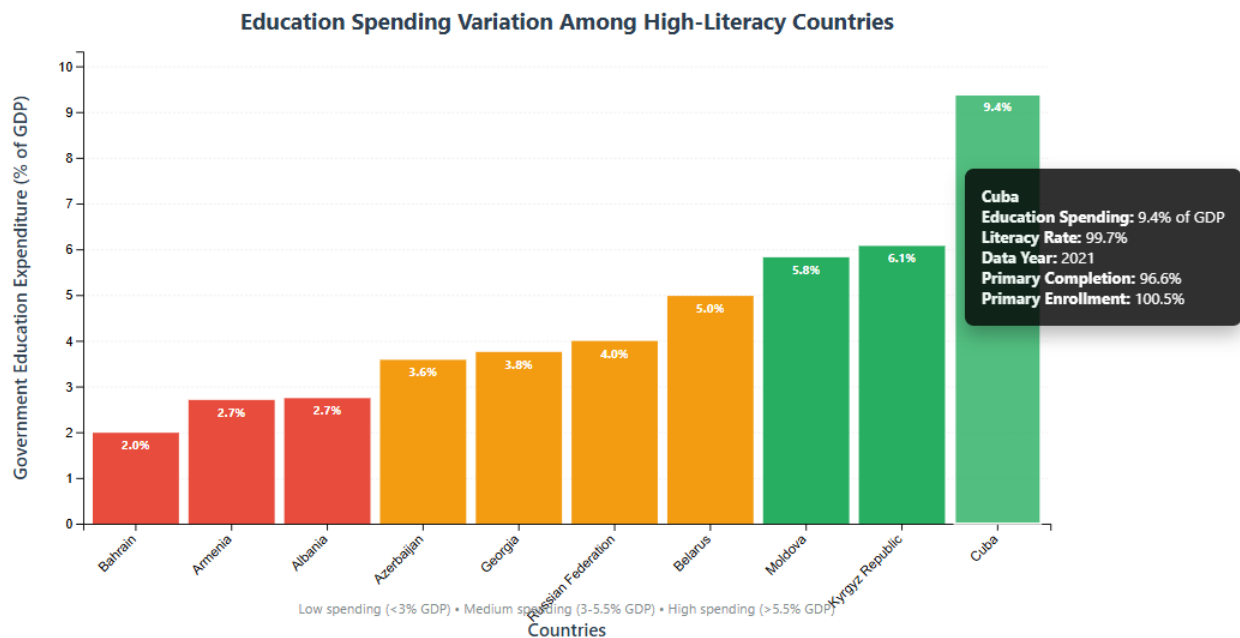
5. Unexpected Success Stories:

Some countries with modest educational spending have achieved remarkable literacy rates. These success stories often involve innovative teaching methods, community involvement, and prioritizing education in unique ways.

Figure 4.8: Content Creator persona: Key Insights panel (Global Education dataset)

The visual elements in the Content Creator output were also tuned for shareability and impact. The first visualization (Figure 4.9) was described as a bar chart highlighting countries with extremely high adult literacy and comparing their public education expenditure. Unlike a typical bar chart, the

persona’s version binned the bars by spending level (low <3%, medium 3-5%, high >5.5% of GDP) and used distinct colors for these categories, effectively turning a continuous variable into story-friendly labels (low/medium/high spend). The chart’s design was bold and uncluttered, with country names and simple bars, more suitable for social media infographics. The caption was crafted as a catchy line: *“These countries lead in literacy despite varied education spending.”* The output even proposed a hashtag: #EducationMatters. This demonstrates the persona’s inclination to package insights for social media virality. The narrative explained one example from the chart: e.g., *Cuba spends 9.4% of GDP on education with 99.7% literacy, whereas another country spends far less but also has high literacy*, thereby encouraging the reader to question what drives educational success beyond just funding.



Visualization 1

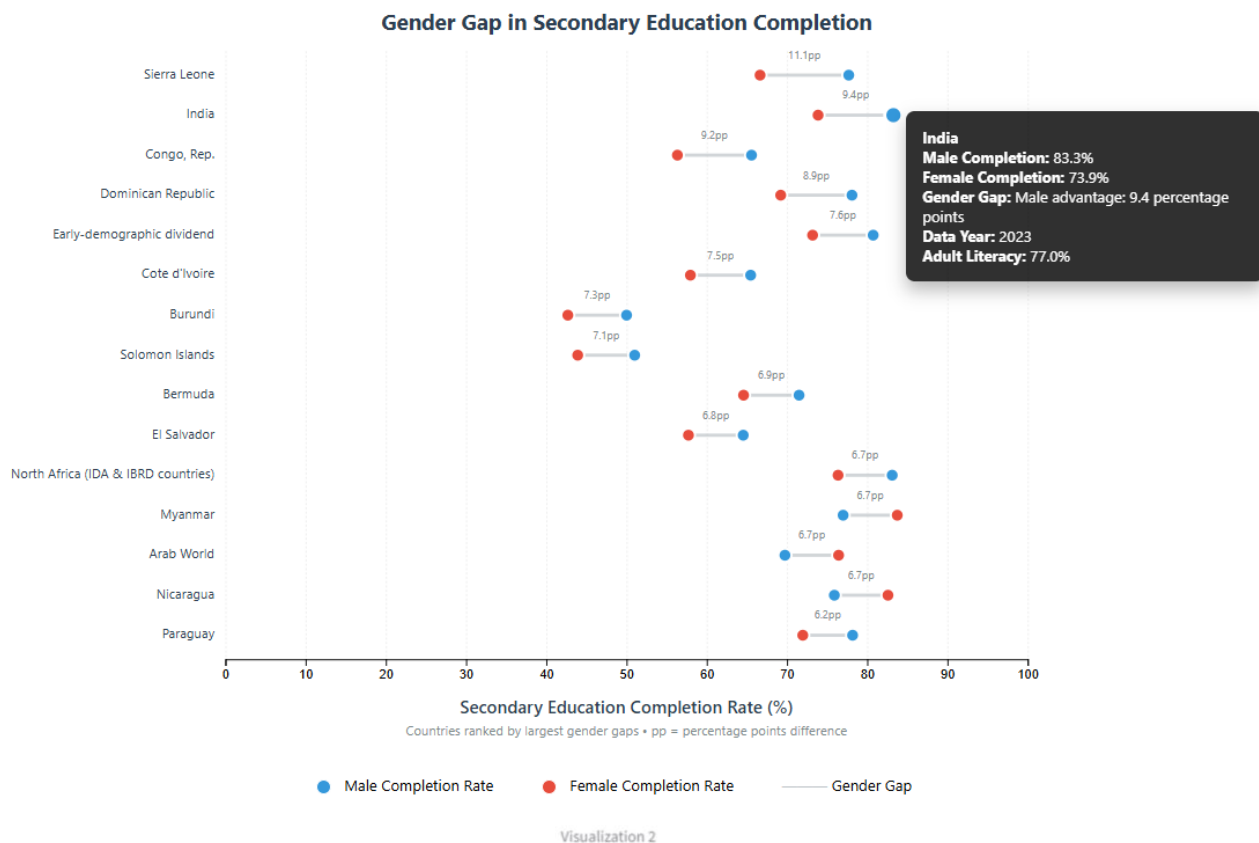
Caption: These countries lead in female literacy despite varied education spending.

Social Media Caption Idea/Hashtag: "Literacy Leaders or Budget Balancers? #EducationMatters"

Figure 4.9: Content Creator persona: Visualization 1— “Education Spending Variation Among High-Literacy Countries” (Global Education dataset)

The second visualization in the Content Creator output (Figure 4.10) addressed gender inequality in education through an approachable graphic. It was a form of a *dumbbell chart* comparing female vs male school completion rates in selected countries, visualized as pairs of dots connected by a line for each country. The persona’s caption for this chart was striking: *“Mind the gap: Education*

should be equal #GenderEquality". This phrasing uses a colloquial expression ("mind the gap") and an advocacy tone, complete with a hashtag tying it to a broader social theme. The design was kept simple and focused: countries were listed on the y-axis, and for each country a line connected the male completion rate (one end) to the female completion rate (the other end). The length of the line immediately conveyed the gender gap – wide separations indicated large disparities favoring boys (for example, the chart highlighted countries like *South Sudan* or *Yemen* with very large gaps), whereas near-overlapping points indicated near gender parity (e.g., *Rwanda* or *Cambodia*). The persona used this visual as a moral narrative focal point: rather than just showing percentages, the accompanying text described what the gap means in human terms ("thousands of girls left behind from completing school"), appealing to the reader's emotions.



Caption: Bridging the gap: Where boys and girls diverge in education.

Social Media Caption Idea: "Mind the gap: Education should be equal. #GenderEquality"

Figure 4.10: Content Creator persona: Visualization 2— "Gender Gap in Secondary Education Completion" (Global Education dataset).

Together, Figures 4.9 and 4.10 formed a narrative arc for the Content Creator: the first demonstrated that money alone *doesn't guarantee educational success*, and the second argued that

even success must be equitable to be truly a success. The persona's Recommendations section, as showed in Figure 4.11, was notably different from the others: instead of policy advice, it gave ideas on how to turn these insights into stories or campaigns. For example, it suggested making the literacy vs spending chart into a LinkedIn carousel post highlighting “unexpected achievers,” or converting the gender-gap chart into an animated graphic where the gaps close as a hopeful vision of change. It even offered ready-made titles for blog posts or videos (e.g., “The Invisible Gender Gap”, “Not All Spending Leads to Learning”) to hook audiences. This self-referential angle, advising on how to *communicate* the data, shows the model understood that this persona's goal is not to make policy or do analytics, but to create content that informs and inspires an audience.

Recommendations

- **Content Creation:** Turn the bar plot into a LinkedIn carousel to highlight how some countries achieve high literacy with low spending.
- **Advocacy Hooks:** Encourage viewers to compare their country's data and start conversations about educational priorities.
- **Title Ideas:**
 1. "Who's Left Behind in School?"
 2. "The Invisible Gender Gap"
 3. "Not All Spending Leads to Learning"

Persona Perspective

As a content creator, these insights are more than just data points; they're narratives waiting to be told. Each of these lines is a child's future bending toward — or away from — opportunity. Only three countries spend above 8% of GDP on education. Do their students win? This question can drive compelling content that challenges assumptions and inspires change.

Conclusion

This isn't just global education. It's global storytelling — and we need more voices to tell it. Data is only powerful when it's shared, felt, and seen. Let's use these insights to spark conversations, drive change, and ensure every child has the chance to learn and thrive.

Figure 4.11: Content Creator persona: Recommendations, Persona Perspective and Conclusion panel (Global Education dataset)

In terms of tone, the Content Creator output was inspirational and optimistic by the conclusion. As in Figure 11, it ended with a call to action and a hopeful message, for instance: “*Data is only powerful when it's shared, felt, and seen. Let's use these insights to spark conversations and ensure every child has the chance to learn.*” This kind of emotional appeal is in line with persuasive

storytelling best practices, aiming to leave readers feeling motivated. From a qualitative standpoint, the Content Creator persona output succeeded in balancing factual insight with emotional resonance. It did not ignore the data: it cited specific disparities and outliers, but it delivered them through metaphors, relatable language, and visually engaging elements instead of dense facts. This was precisely the intention behind this persona: to see if the model could shift into a more narrative, less formal style. The results indicate that the dynamic persona prompt (which explicitly requested a “storytelling narrative” and encouraged use of metaphors or social context) indeed led GPT-4 to change its register. It presented only a few key numbers in context, a deliberate reduction in “data density” to avoid overwhelming a general audience. Such an approach aligns with best practices in data storytelling for lay audiences, which favor a handful of impactful statistics woven into a story, rather than an avalanche of figures.

4.2.4 Journalist Persona

In addition to the three predefined personas, a custom persona was tested by providing the role “Journalist” along with specific interests (“education inequality, policy impact, gender gaps”) via the application’s interface. The resulting *Journalist* persona output combined elements of the business and content creator styles, reflecting an investigative journalism approach that is issue-driven yet narrative in flavor. The report produced under this persona read much like a feature article or investigative piece one might find in a news magazine: it opened with a bold, headline-like statement and maintained a tone of accountability and inquiry throughout.

For example, the introduction, as showed in Figure 4.12, began with a sentence such as: *“In a world where education is touted as the great equalizer, a striking paradox emerges some of the highest spenders on education are not the leaders in literacy.”* This immediately sets up a critical angle, questioning assumptions and pointing out paradoxes, which is a common journalistic technique to hook readers. Compared to the Content Creator persona, the Journalist’s rhetoric was less overtly emotive and more issue-focused; compared to the Business Analyst, it was less about managerial terms and more about public accountability and human impact. Essentially, the journalist voice sat in between, aiming to surface discrepancies and ask why they exist.

Introduction

In a world where education is touted as the great equalizer, a striking paradox emerges: some of the highest spenders on education are not necessarily the leaders in literacy. As we delve into a comprehensive dataset spanning from 1999 to 2023 and covering over 160 countries, we uncover stories that challenge assumptions and highlight disparities in global education systems. This dataset serves as a rich source for public interest stories, revealing gaps in education access, outcomes, and gender equality across the globe.

Figure 4.12: Journalist persona: Introduction panel (Global Education dataset)

Throughout the text, the journalist persona emphasized paradoxes and unanswered questions, e.g., noting countries that allocate large shares of GDP to education do not invariably top literacy ranking, suggesting policy inefficiency or misalignment (policy-impact angle). In the key insights, illustrate in Figure 4.13, it focuses on gender disparities, noting that despite progress, wide gaps remain (e.g., a double-digit difference in primary completion in selected cases). Enrollment anomalies are called out as systemic measurement/administration issues worth journalistic investigation. Contradictory patterns (e.g., Finland ≈ 14 vs. India > 30) are framed as quality-of-service stories rather than mere covariates; regional disparities in secondary enrollment are described taking for example *Sub-Saharan* low values versus near-universal enrollment in parts of Europe are positioned as comparative reporting angles.

Key Insights

1. High Spending Does Not Guarantee High Literacy

Countries like the United States and Norway allocate significant portions of their GDP to education, yet they do not always top the charts in literacy rates. For instance, while Norway spends around 7% of its GDP on education, its literacy rate is comparable to countries spending less, such as Poland, which spends approximately 5% but boasts a similar literacy rate of around 99%. This suggests that factors beyond financial investment play crucial roles in educational outcomes.

2. Gender Disparities Persist Despite Progress

While many countries have made strides in closing the gender gap in education, others lag behind. For example, in Pakistan, the male primary school completion rate is significantly higher than that of females, with a gap of over 15 percentage points as of 2023. Conversely, Rwanda has successfully closed this gap, showcasing nearly equal completion rates for both genders.

3. Anomalies in Enrollment Rates

Some countries report primary school enrollment rates exceeding 100%, such as Belgium with 106.82%. This anomaly often results from late enrollments or students repeating grades, indicating systemic issues within the education system that need addressing.

4. Contradictory Patterns in Pupil-Teacher Ratios

Countries with lower pupil-teacher ratios, like Finland (with a ratio of about 14 in primary education), tend to perform better in literacy rates compared to countries with higher ratios, such as India, where the ratio exceeds 30. This highlights the importance of teacher availability in improving educational outcomes.

5. Regional Disparities in Secondary Education Enrollment

Sub-Saharan Africa continues to struggle with secondary school enrollment, with countries like Niger reporting rates as low as 25%. In contrast, European nations like Germany boast enrollment rates nearing 100%, pointing to significant regional disparities that demand targeted policy interventions.

Figure 4.13: Journalist persona: Key Insights panel (Global Education dataset)

The Journalist persona's visualizations were used similarly as evidence exhibits. One chart was described as a side-by-side bar chart comparing education spending vs literacy for selected countries effectively to visualize the paradox that higher spending does not always equate to higher literacy. This could be considered analogous to the Business Analyst's scatterplot but presented in a simpler bar format for direct comparison. Another visualization was a dumbbell chart on gender gaps (similar form to the Content Creator's second chart). Interestingly, even though the form of the gender gap chart was similar across personas, the *framing* differed: the Content Creator captioned it with a hashtag and emotional appeal, whereas the Journalist version presented it with a sober description of the gap and tied it to a question of causes and accountability (e.g., noting specific gap values and

asking what policies or cultural factors drive those gaps). This adaptability demonstrates that the LLM “knew” to reuse an effective visualization form for the given data insight, but to alter the tone and context to suit the journalistic voice.

The recommendations in the Journalist persona’s report were framed as story angles or further investigations rather than direct fixes. For example, one recommendation was: *“Investigate why some high-spending countries fail to achieve commensurate literacy rates – this could uncover inefficiencies or misallocation in education budgets.”* Another suggested: *“Look into policies in countries like Rwanda that managed to close the gender gap, and why others like Pakistan continue to struggle”*. These are essentially proposals for follow-up journalism, showing the model understood that a journalist persona would seek to inform public discourse and probe deeper, rather than prescribe policy changes themselves.

The conclusion of the journalist piece maintained the spirit of inquiry and accountability. It ended with pointed, open questions (e.g., “What explains the gap between spending and outcomes? Why are some nations closing the gender gap while others widen it?”) – a typical journalistic move to leave the reader thinking and to subtly call on authorities to address these issues.

The successful adaptation to this custom Journalist persona demonstrates the system’s flexibility in blending styles. The output effectively became a hybrid of the Business Analyst (focus on inefficiency and policy outcomes) and the Content Creator (storytelling and human interest), which matches what good investigative journalism often does. Importantly, the model stayed tightly on the topics included in the prompt’s focus keywords, indicating that the prompt’s inclusion of those focus keywords kept the narrative on track. Across the Data Scientist, Business Analyst, Content Creator, and Journalist personas, the differences in output were not superficial but substantive, different insights were prioritized, and the tone/structure shifted significantly. This confirms that dynamic persona prompts can steer a powerful LLM to adopt different “voices” and analytical priorities in data storytelling. In all cases, the model remained grounded in the actual data (ensuring factual accuracy), but it modulated the style and framing to suit the intended audience, which is a crucial capability for personalized data communication.

4.2.5 Data Humanism Mode

One of the core explorations in this thesis is the concept of making data storytelling more human-centered, empathetic, and creative, as advocated by Giorgia Lupi. To test the system’s capacity for this, a special Data Humanist mode was implemented which can be applied on top of a persona (in this evaluation, it was tried with the Content Creator and Journalist personas). When activated, the

prompt used is the `get_humanism_prompt` that explicitly directs the LLM to yield an evocative storytelling and to use non-traditional, metaphorical visuals instead of standard charts. The results were striking, demonstrating the model’s ability to significantly alter its narrative voice and visual design when instructed.

With the Content Creator persona in Data Humanism mode, the output transformed into what can be described as a data-driven short story. The narrative largely abandoned precise numbers, instead conveying insights through metaphor, imagery, and emotional language. The overall introduction (Figure 4.14) began with a philosophical metaphor: “*Imagine a world where every child’s potential is a seed, waiting to bloom*”, and extended the education dataset into a metaphor of a global garden where some seeds (children) flourish in sunlight (opportunity) while others struggle in shadow. This introduction set a philosophical tone, treating the data as “*a tapestry of dreams, challenges, and silent triumphs*” from human stories. Such framing mirrors Lupi’s insistence on context and empathy, the narrative explicitly connected data trends to social meaning, e.g., portraying missing data as “silent echoes” of children not in the classroom. Even the presence of an epigraph (a quote attributed to Lupi herself in the output’s intro) reinforced the connection to Data Humanism’s values. The result blurred the line between analysis and literary storytelling. Impressively, the LLM managed to remain *technically grounded* (it did not invent false facts; the qualitative statements were rooted in real patterns from the data) while being emotionally and symbolically rich. This indicates that, given the right prompt guidance, an LLM like GPT-4 can adopt data humanism principles to produce output that is both informative and deeply empathetic.

Complete Analysis Report

1. Introduction

Imagine a world where every child’s potential is a seed, waiting to bloom. Yet, in the garden of global education, some seeds are showered with sunlight and rain, while others struggle in shadows. This dataset, spanning over two decades and 160 countries, is not just numbers on a page — it’s a tapestry of dreams, challenges, and silent triumphs. It tells the story of humanity’s commitment to learning, revealing the disparities and hopes that shape our future.

Figure 4.14: Content Creator persona on Humanism Mode: Introduction panel (Global Education dataset)

Perhaps the most distinctive aspect of the Data Humanism mode output were the visualizations. The system generated two interactive visuals that were not typical charts, but rather metaphorical illustrations of the data. The first (Figure 4.15) was titled “Education Blooms.” Instead of bars or lines, this visualization was rendered as a rosette or flower plot, where each country was represented as a petal in a circular layout. The encoding was creative: petal length corresponded to

the country's literacy rate (longer petal = higher literacy), petal width or size reflected primary enrollment, and color intensity represented education spending (e.g., light-colored petals for low spenders and dark-colored for high spenders). Countries with high spending and high literacy would thus appear as large, dark petals, whereas those with high spending but low literacy might appear as dark but short petals (a paradoxical combination). The description given in the text (and tooltip examples, such as for North Macedonia, positioned on expenditure level, 8.8% the distance from the center on the literacy rate, here as 84.4% and the size of the petal as the percentage of primary enrollment rate a 113.2% in this case.)

Education Blooms



Figure 4.15 Content Creator persona on Humanism Mode: Visualization 1— “Education Blooms” and Caption (Global Education dataset).

The caption for Education Blooms (Figure 4.15): “Each petal represents a country's literacy – a bloom of knowledge unfurling. Where petals are full and vibrant, literacy thrives; where they are sparse, the garden of learning struggles to grow.” (and it included the hashtag #LiteracyGarden). This visual metaphor makes the insight emotionally accessible, a design inspired by Lupi's analogies in her artworks.

The second visualization, titled “*Seeds of Knowledge*” (Figure 4.16), took a complementary metaphor. It depicted each country as a seedling or plant along a horizontal “field.” In this design, the x-axis was country names (like seeds planted in a row), the stem height was the average school enrollment rate (so taller green stems = higher enrollment), and on top of each stem was an orange dot or flower representing completion rate (with a full bloom if the completion rate exceeds a threshold, above 50%, otherwise just a small bud). Essentially, this chart created a field of stems where not all had blossoms: a powerful metaphor showing that in some countries, education grows to a point (students enroll) but does not “flower” into completion for all. The snapshot described in the text noted how countries like Ukraine and China had tall stems with bright petals (high enrollment and high completion – a flourishing plant), whereas others had shorter stems and no flowers (low enrollment or very low completion – a struggling seed). The tooltips even anthropomorphized the data (e.g., calling a country a “Blooming Country!” if it had near high completion rate). This inventive design diverges from typical bar graphs or scatterplots; instead of plotting points, it plants seeds, inviting viewers to emotionally invest in whether each “seed” in the field gets to bloom.

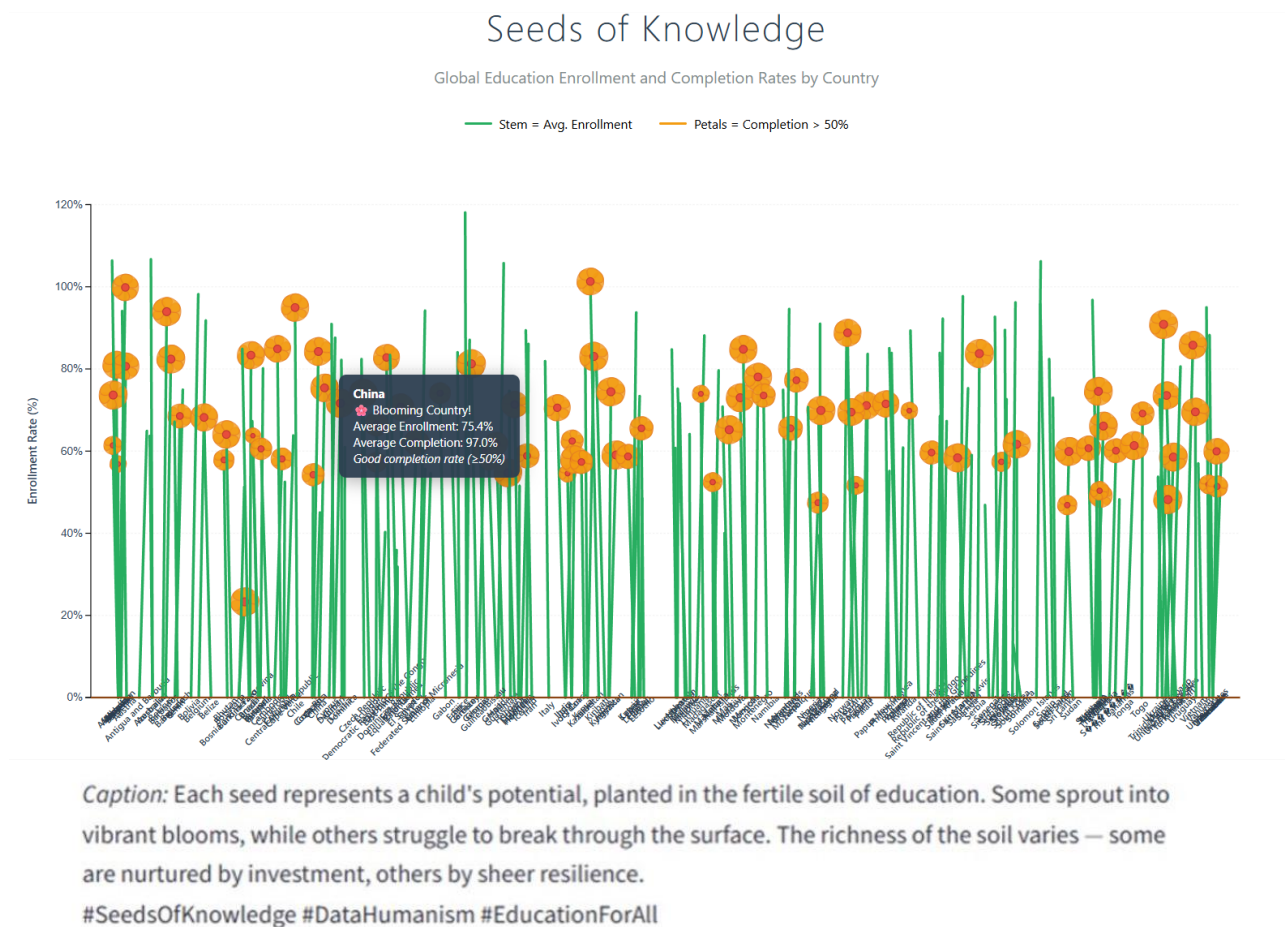


Figure 4.16: Content Creator persona on Humanism Mode: Visualization 2— “*Seeds of Knowledge*” and Caption (Global Education dataset).

These visuals are far from conventional, yet they remained tied to the actual data mappings. Importantly, they demonstrate the LLM’s ability to use D3.js not just for standard charts but to create visual analogies. This is a notable achievement: designing such visual metaphors manually is a complex task, yet the AI autonomously produced the concept and code for them when prompted in the data-humanist vein. It follows closely the spirit of Giorgia Lupi’s work (who often designs unique icons or shapes to represent data points), here the AI essentially did the same, creating flower and plant motifs to encode education data. The success of these creative outputs suggests that GPT-4, operating under persona + humanism prompts, can venture into the realm of LLM-driven data art, transforming quantitative facts into visually narrative experiences.

Structurally, the Data Humanist output read more like a reflective essay or feature story than an analysis. It had a clear beginning, middle, and end geared towards an emotional arc: the introduction set a tone of wonder and concern, the body (“Voices in the Data”) gave *voice* to statistics via mini-stories and metaphors, the visuals served as symbolic focal points, and the conclusion closed with a collective call for empathy and imagination. For example, the conclusion (Figure 4.17) posed a question to the audience: *“What stories will these children tell, and what futures will they create, if given the chance?”*, challenging readers to envision the human lives behind the numbers. This rhetorical flourish loops back to the opening metaphor of seeds and potential, reinforcing the narrative cohesion of the piece. The writing included vivid passages that illustrate how it tied data to human experience: one passage contrasted overcrowded classrooms and the silence of girls kept out of school with rich metaphors (chalk and silence). Another line personified the data starkly: *“Some nations teach with chalk, others with silence... in other lands, the absence of chalk marks the absence of opportunity”*. Such language clearly goes beyond any standard analytical report, entering the domain of literary journalism or even prose poetry around data.

4. Conclusion

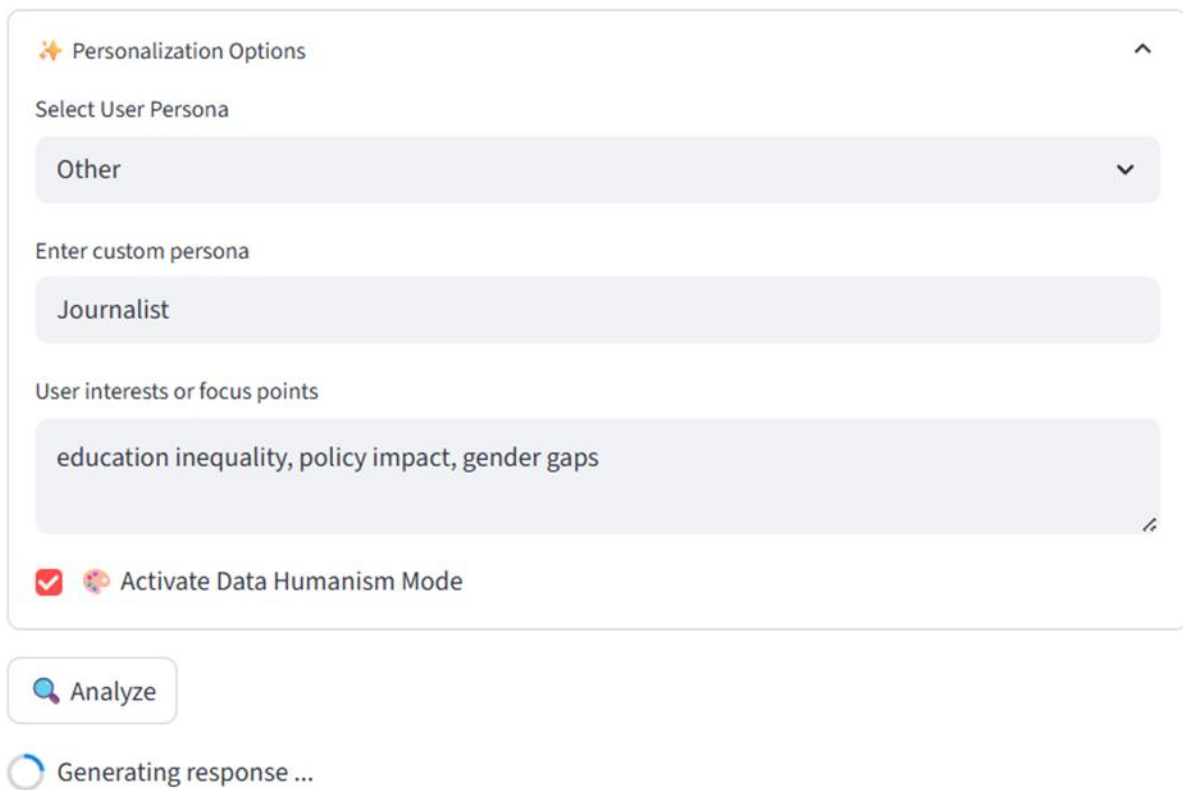
What if every child could step into a classroom filled with light, their dreams nurtured by knowledge and opportunity? As we reflect on this data, let us not only see the gaps but also the possibilities. Let us imagine a world where every child's potential is realized, where education is the sun that helps them grow. What stories will they tell, and what futures will they create?



Figure 4.17: Content Creator persona: Conclusion panel (Global Education dataset)

The mode was also applied the Data Humanism mode to the Journalist persona to see how it would differ from the standard journalist output. As showed in Figure 4.18, the Humanism Mode was

activated with the Journalist (Other Custom Persona) with specific interests: education inequality, policy impact and gender gaps.



The screenshot displays a 'Personalization Options' panel with a light gray background and rounded corners. At the top, the title 'Personalization Options' is accompanied by a star icon and an upward arrow. Below this, the 'Select User Persona' section features a dropdown menu currently set to 'Other'. The 'Enter custom persona' section has a text input field containing the word 'Journalist'. The 'User interests or focus points' section includes a larger text area with the text 'education inequality, policy impact, gender gaps'. At the bottom of the panel, there is a checkbox with a red checkmark and a colorful icon, labeled 'Activate Data Humanism Mode'. Below the panel, there is a blue 'Analyze' button with a magnifying glass icon. At the very bottom, a blue circular progress indicator is shown next to the text 'Generating response ...'.

Figure 4.18: Streamlit application Interface: persona and Data Humanism selection

In this case, the result was an investigative piece infused with humanistic storytelling elements. The journalist-humanist output still identified the same core issues: the spending paradox and gender disparities but its style became more expressive and empathetic. For instance, the first visualization in this mode was titled “*The Spending Paradox*” (Figure 4.19), essentially a scatterplot of spending vs literacy similar to the earlier Business Analyst chart, but with a twist: the model highlighted paradoxical cases by coloring them red (e.g., Yemen, Senegal – high spend, low literacy) while coloring expected cases in blue (high spend, high literacy). This visual design immediately draws attention to inefficiencies and was paired with narrative text that took a tone of public accountability, calling these outliers evidence of a policy gap. Even though the chart itself was quantitative, its framing was made accessible and evocative, consistent with data humanism’s aim to make data relatable.

The Spending Paradox



PARADOX: HIGH SPEND, LOW LITERACY

● High spending, low literacy ● High spending, high literacy ● Standard performance

Figure 4.19: Journalist persona on Humanism Mode: Visualization 1— “The Spending Paradox” (Global Education dataset).

The second visualization was a gender-gap dumbbell chart very much like the Content Creator’s, but here the caption and analysis were more somber and factual. The combination in the journalist’s humanist version thus delivered the same factual punch as the standard journalist output, but with more layered emotional engagement.

It is instructive to compare the content creator vs journalist under Data Humanism mode: both converged on using some similar visuals (like the dumbbell for gender gap) and themes (inequality, inefficiency), but the treatment differed: the content creator had a catchy slogan and rallying cry (#GenderEquality), whereas the journalist’s had a more sober caption focusing on evidence and prompting questions. This shows the model can even fine-tune the humanistic approach to suit different persona voices: one more populist and advocacy-driven, the other more investigative and contemplative.

These experiments with Data Humanism mode test an LLM’s ability to produce such creative, metaphor-rich data narratives. The results show that GPT-4 can operate in this highly creative regime and still remain grounded in the truth of the data: a promising sign for the emerging intersection of AI and creative data communication. A central question was whether the LLM could craft a meaningful, emotionally resonant story that truly captures the “humanistic” goals of narrative data design. Based on this analysis, the answer is largely yes. The outputs, especially the journalist’s Data Humanism piece, demonstrated a clear shift from dry reportage to a story that engages empathy and curiosity. The model reintroduced subjectivity and context in measured ways (e.g., a journalist quote, direct address of the audience about implications) without straying from the evidence. This aligns strongly with Lupi’s call to reconnect data to the human stories they represent: treating numbers not as abstract truths but as proxies for reality, full of imperfections and human context. Indeed, the LLM’s narrative echoed several of Lupi’s principles: it acknowledged complexity (noting paradoxes and multifaceted outcomes rather than oversimplifying), embraced imperfection (highlighting anomalies and missing data as part of the story rather than ignoring them), and turned data into feeling at multiple points. For example, a phrase from the content creator’s humanist mode: “a generation of women left behind”, encapsulates a statistical gap and a human image. This is precisely what humanistic data storytelling strives for: connecting the viewer to the human significance behind numbers. By avoiding a purely analytical voice and incorporating metaphor, the LLM allowed the audience to perceive the gravity of issues like gender disparity not just intellectually, but emotionally. As Lupi herself emphasizes, data must be tied back to what they stand for – people’s lives, behaviors, and stories [2]. The AI system, via prompt-based guidance, demonstrates this connection.

In summary, the Data Humanism mode showcases the narrative adaptability of the system at its most extreme setting, where even the form of visualization and the fundamental style of writing can be altered to serve a communicative intent. Both the standard outputs and the humanist outputs underline a key achievement: the model can maintain factual grounding and coherence while modulating tone, style, and emphasis dramatically based on the specified persona and mode. This demonstrates an advanced degree of controllability in AI-generated storytelling, opening doors to LLMs functioning as versatile narrative partners that can cater to different audiences and purposes.

4.3 WASH Dataset

Thus far, the evaluation has focused on the Global Education dataset. As a supplementary test, the storytelling system was applied to a different dataset: the WASH (Water, Sanitation, and Hygiene) dataset, derived from the WHO/UNICEF Joint Monitoring Program (JMP) global dataset, which proved internationally comparable estimates of household access to drinking water, sanitation, and

hygiene services. The version used in this test covers over 150 country-level observations consolidated for the year 2022, and includes indicators aligned with the Sustainable Development Goals. The principal variables include population using safely managed drinking water services, defined as access to an improved water source, available and free from contamination; population using at least basic drinking water serviced and sanitation services, and population access to basic hygiene services.

For analysis purposes, the dataset provided an ideal test for the AI capacity to adapt to the narrative and highlight geographic disparities: some regions approach universal access, while others still leave hundreds of millions without basic services. Those indicators are also directly tied to survival, health, and dignity, which makes them particularly suitable for evaluating the LLM's ability to generate empathetic narratives. The persona chosen for this run is a Community Caregiver, a volunteer working with communities, with interests in safe drinking water access for children and community health protection. To additionally evaluate the narrative the same persona is tested also in Data Humanism mode, the goal is to assess whether the prompt-engineering can effectively work on different dataset and multiple personas.

4.3.1 Community Caregiver Persona

In the standard mode, the Community Caregiver output took an empathetic tone towards public health advocacy. The report's introduction opened with a vivid scene: *"In 2022, over 75 million people in Sub-Saharan Africa still lacked even a basic drinking water service -a crisis no child should endure"* Already from the begging the narrative stressed human impact, urgency and inequity in WASH access. The approach aligns with the caregiver's perspective, emphasizing themes of injustice and the urgency of intervention. Throughout the narrative, the tone remained advocacy-oriented and used terms like "fundamental rights", "crucial for dignity" and supported with real data. This demonstrates how the persona prompt steered GPT-4 to align with a community health protector's voice, focusing on the well-being of people over analytical detail.

The Key Insights highlighted inequities and consequences from: *"regional disparities in WASH access are creating health inequities that perpetuate cycles of poverty and illness"* to *"vulnerable populations requiring immediate intervention"*. Such framing turns the dataset into a narrative of inequity and risk, starting from public health context (mentioning disease outbreaks linked to poor sanitation, impacts on child development, etc.) to ground data in real-world outcomes. Overall, the storytelling was tailored to a humanitarian report, using the findings to appeal for resources and action, rather than as neutral facts.

The visualization generated were likewise selected and described to reinforce an advocacy narrative. The first visualization was a straightforward but impactful comparison chart focusing on those left behind. Figure 4.20 shows a bar chart of population lacking basing WASH access by region. Each bar corresponds to a major world region, and is color-coded with a red/green schema: the red segment of each bar represents the number of people without access to safe water, while the green segment represent those who do have access, providing context. This design instantly highlights the problem spots: regions with long red bars indicate a vast underserved population. The caregiver narrative explicitly drawn attention to these disparities: for example, the interactive tooltip on that bar notes that Sub-Saharan Africa’s bar towers above others, with over 75.7 million people lacking safe water, far exceeding any other region. The system transforms a simple bar chart into an advocacy tool, where red signifies a call to action. Figure 4.20’s design and interpretation served the caregiver’s narrative purpose: it visually amplifies inequity and covey both hope and alarm.

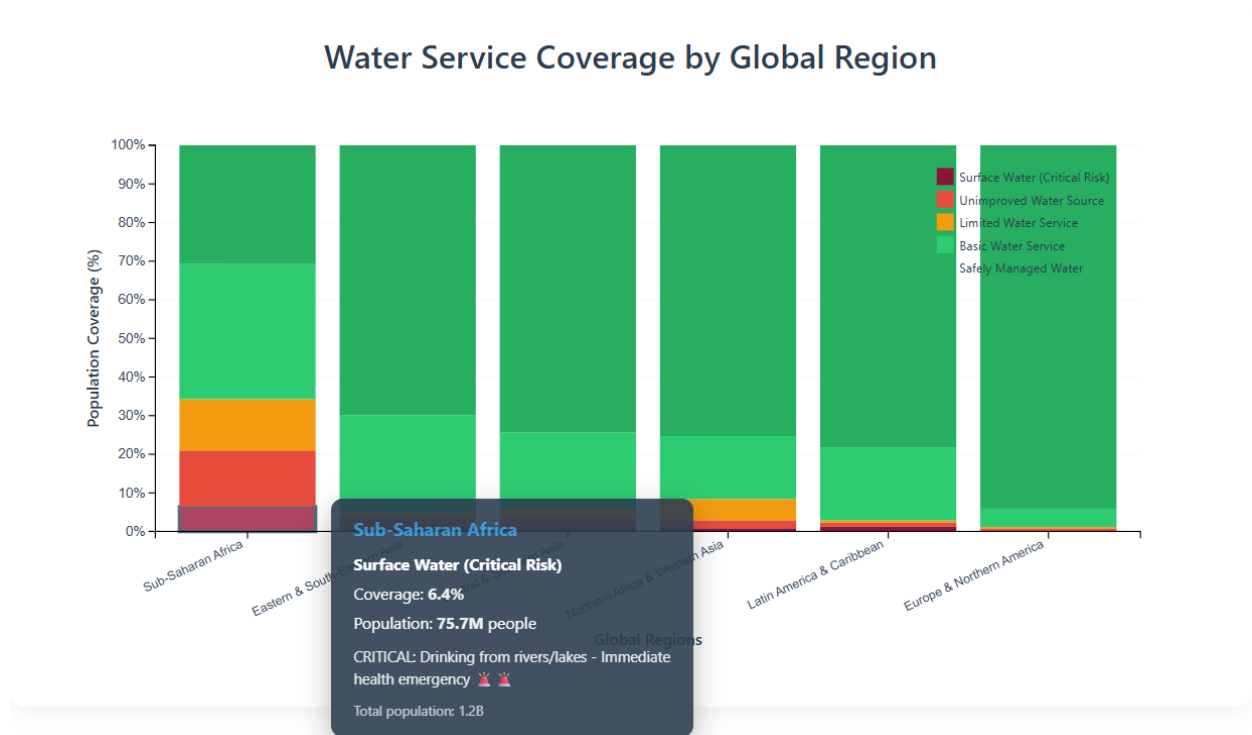


Figure 4.20 Community Caregiver: Visualization 1— “Water Service Coverage by Global Region” (WASH dataset).

The second visualization was a more strategic overview aimed at guiding where to intervene first. Figure 4.21 presents a Priority Intervention Matrix, structured by region and the urgency of WASH needs. In this matrix chart, each row corresponds to a region and each column represents a key WASH domain (safe, water, sanitation, or hygiene), creating cells that indicate how critical the intervention need is. The cells are color-coded on a heatmap scale of urgency: red denotes a critical, high-priority gap, yellow indicates moderate needs, and green signals that a region is closer to goals

in that area. Sub-Saharan Africa again stood out with multiple red cells, in contrast, a region like Europe/Northern America appears green, indicating low health level and maintenance intervention (as the tooltip highlights). The matrix format allowed the AI to convey at a glance which region-domain combinations were most urgency of intervention, in a visually call-to-action, aligning perfectly with the community caregiver’s mission.

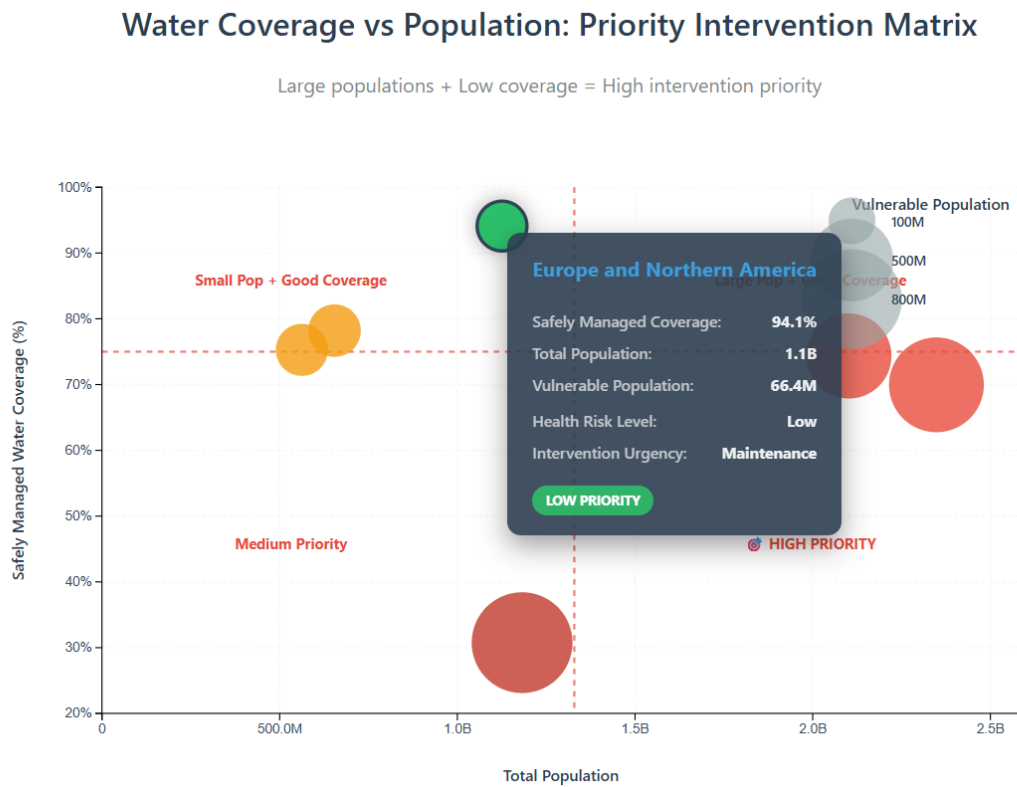


Figure 4.21 Community Caregiver: Visualization 2— “Water Service vs Population: Priority Intervention Matrix (WASH dataset).

Percentages and population counts became stories of communities, the AI guided by the persona prompt chose to emphasize statics that carry emotional weight (like the largest underserved populations) and framed in terms of human experience (risk of disease, daily hardship of water collection, etc.). The result reads like a report by a humanitarian organization: factual accuracy is maintained, but every fact is linked to a value or call. The model even used analogies to make numbers relatable, a technique used in other persona outputs. It mirrors the comparative strategy observed with the Content Creator persona on the Global Education dataset. In the Recommendations section, the caregiver persona’s naturally pivoted to advocacy points: it urged investment in water infrastructure, community education on hygiene and international support for the hardest-hit regions. These recommendations were tied back to the evidence data to maximize impact, particularly importance was given to: “*monitoring and evaluation: establish systems to track the programs of WASH programs*

and measure their impact on health outcomes.” The conclusion of the report was emotionally charged and inspiration: *“The human cost of inaction is staggering, with lives lost and families suffering due to preventable disease. By prioritizing targeted interventions and securing resources, we can improve health outcomes and break the cycle of poverty and illness.”* In sum, the standard-mode output for the Community Caregiver persona validated the system’s cross-domain versatility: even with a new dataset, the LLM adjusted its narrative voice to advocate and empathize, turning the WASH indicators into a persuasive human-centered story for change.

4.3.2 Community Caregiver Persona on Data Humanism Mode

When the Community Caregiver persona was applied in the Data Humanism mode, the resulting output took on a markedly different style, one that was poetic, metaphor-rich, and deeply emotive. This mode, which adds an extra layer of creative prompting inspired by Giorgia Lupi’s human-centered data design, pushed the system to go beyond advocacy into personalized, story-like narration. The tone became contemplative and lyrical, as if the report gave way to emotionally grounded metaphors into imagery. The data was once again personified: *“Imagine a village where out of every ten children, four wake up each morning with thirst unquenched – their day defined by a walk for water.”* Continuing the narrative: *“Water access is more than a statistic; it is a cornerstone of child development, family stability, and community strength.”* Such prose resembles a crafted story, using metaphor (each drop as a promise) and contrast (children who drink vs. those who dream) to evoke empathy. The introduction set a reflective scene, the corpus of the data story weaved in “mini-stories” that represented data points (for instance, an anecdote of communities distances – one flourishing with a new well, another still waiting – to illustrate disparities). As we can see some themes as a recurrence: the disparities, the gaps that in the previous data story were gender gaps here are in access, also the limited water sources that make children thirst and in danger, in the previous Global Education dataset where flowers waiting for blooming.

The Visualizations produced was an interactive water and sanitation visual summary interface, where each country takes the figure as a water droplet icon, some partially filled in blue to represent the portion of the population with access to clean water and left hollow to indicate the portion without access. The caregiver persona’s narrative described this figure in a very evocative way, painting a picture for the reader: *“The data takes the shape of a water droplet – half full, half empty – symbolizing hope and urgency in one image.”* It is essentially a mini infographic that allows viewers (especially non experts) to grasp the scale and significance of the issue at a glance. Functionally it still conveys the essential quantitative message (the proportion of people with vs. without WASH access), but it does so in a way that “emphasizes people over numbers”, exactly as the manifesto of

data humanism advocates. The caregiver narrative describes this interface in story-like terms, using the droplet as a symbol of both hope and the remaining “thirst” in the world, thereby turning data into a relatable visual tale.

The second visualization in the Data Humanism was very remarkable, it took the form of a Sankey diagram. The figure 4.22 was presented as a flowing story of causes and effects in the WASH ecosystem, which maps the flows of risk, health burden, and intervention pathways as interconnected streams. Each node and link in the Sankey chart represent a narrative meaning. On the left sizes are nodes representing key risk factors or gaps (for example, countries without drinking water or sanitation). These nodes have streams flowing from right with levels like basic services, no handwashing facilities, safely managed service, limited service, surface water, unimproved and open defecation. Those flows into middle nodes one for each area of the dataset (hygiene, drinking water and sanitation) and the flows then split or merge to the corresponding country value (of course the flow can also be read from left to right). The Sankey is thus a narrative map: one can trace, for example as showed from the hoover in the visualization, a ribbon from “Limited service” of “Hygiene” conveying in Sub-Saharan Africa, conveying the information that in Sub-Saharan Africa, a population of 979 million has a limited service of hygiene. Visually, the Sankey is stylized with curves and flows that transition from red and orange when the services is limited or inaccessible, to a green when it’s a safely managed service, moving on the right the neutral color became blue (in his shades) giving a complete idea of water flowing.

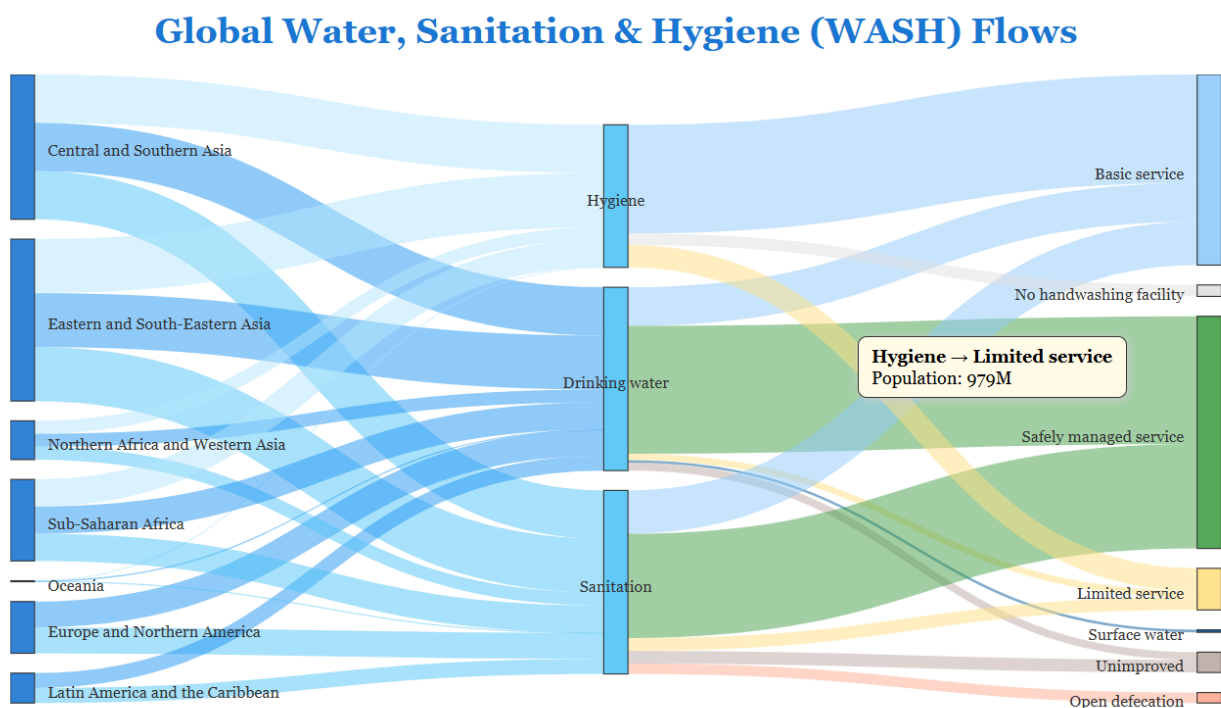


Figure 4.22 Community Caregiver on Humanism Mode: Visualization 2— “Global Water, Sanitation & Hygiene Flows” (WASH dataset).

The aesthetic and function aspects of the output closely align with Giorgia Lupi’s vision of data humanism. The visualization is still very much grounded in the dataset but it lays out context and consequences in a visually narrative form, providing a flow diagram with a blue base color. Functionally, it does communicate the data in a more accessible and emotionally resonant way, which is the goal of data humanism. [23] Also, it is important to note that Giorgia Lupi has used a Sankey diagram to represent her works, so in this case the LLM really centered the visual style.

The storytelling ends in the Conclusion with an emotive call to action: *“This data is not merely a reflection of current conditions; it is a map for targeted interventions that can transform lives. Let’s ensure that every child has the water they need to grow and flourish”*. In summary, the Data Humanism mode output for the Community Caregiver persona stands as a proof of concept for AI-driven personalized visual storytelling. It shows that an LLM can be prompted to generate not only empathetic text but also innovative visual representations that carry the imprint of a human-centric, empathetic design philosophy.

In reflecting on the WASH dataset evaluation, especially under the data-humanism lens, the chapter finds that the system effectively bridged the gap between raw data and human narrative. The Community Caregiver persona in standard mode provided a focused advocacy report, and in Data Humanism mode a touching narrative experience, together illustrating the spectrum of outputs an LLM storyteller can produce. Ultimately, these experiments with the WASH dataset show that our approach generalizes beyond a single domain (education), and across multiple personas.

4.4 Discussion and Implications

The evaluation results presented in this chapter underscore the potential of prompt-based engineering in AI storytelling systems. By systematically analyzing outputs across multiple personas and datasets, it is shown that a single GPT-4 based system can produce a remarkable range of narratives from the same underlying data. The adaptability spans not only superficial style changes but also deeper differences in which insights are prioritized, how visualizations are designed, and how the narrative is framed to serve different communicative intents. Several important insights emerge from this cross-persona evaluation:

- **Tone and Content Focus:** The persona prompts successfully directed the model’s tone (from formal analytic to informal emotive) and focus (from detailing statistics to emphasizing

implications). This means that large language models can be contextually steered to play the role of different communicators (analyst, strategist, storyteller, journalist, etc.), which is valuable for personalized or audience-targeted content generation. It reflects a form of controllable AI behavior that can align outputs with user needs or preferences.

- **Clarity vs Emotion Trade-Off:** Across the personas, a trade-off appeared between quantitative clarity and emotional engagement. The Data Scientist and Business Analyst outputs were dense with facts and clear about data nuances, but relatively dry; the Content Creator and Data Humanist outputs were rich in emotion and narrative, but deliberately sparse with exact figures. This mirrors a real-world communication trade-off: an expert report versus a public story. In the system, this trade-off can be tuned via the persona and mode. An implication is that AI storytellers might be configurable depending on whether the goal is to inform (maximize clarity, detail) or to inspire (maximize emotional resonance), or to balance both. A challenge is ensuring that in emotive modes the content remains accurate – these findings are encouraging in that even the most creative outputs stayed factually grounded, though they sacrificed detail for simplicity.
- **Visualization Strategies Aligned with Persona:** The types and designs of visualizations varied in line with persona needs. A technically inclined persona (Data Scientist) generated analytical visuals like correlation matrices and box plots, whereas an executive persona (Business Analyst) produced decision-oriented charts like scatterplots highlighting efficiency and rankings for benchmarks. The general-audience storyteller (Content Creator) favored visually simple but catchy charts, and in humanist mode, even abstract illustrations. This suggests that AI systems can automatically select or design visual representations appropriate for different communication goals – a form of adaptive visualization design. It’s noteworthy that certain complex chart forms (like the dumbbell gap chart) emerged in multiple personas, implying the model identified them as effective for that data pattern, but the captioning/framing differed by persona. This adaptability in visualization suggests a promising avenue where LLMs might dynamically choose how to visually encode data depending on context, a task traditionally requiring human designers.
- **Persona Effectiveness and Limitations:** The model’s ability to “get into character” for each persona supports the idea that LLMs can be guided by roles. Clear evidence showed that the persona approach yielded outputs aligned with target audiences. This could be highly useful

in practice, for instance, a single data story generation system could output multiple versions for different stakeholders (an internal technical report vs a public press release) at the press of a button. That said, the evaluation is qualitative and based on the content face validity.

- **Connection to Human Alignment:** Ultimately, the exploration ties into the broader theme of making AI outputs more human-aligned and meaningful. The personalized narratives can be seen to align the LLM’s output with human communication styles and values (be it precision, efficiency, empathy, or advocacy). By using prompt engineering as a tool, the system effectively encodes a set of values or perspectives into the AI’s generation process. The Data Humanism mode is a strong example: it encodes the values of empathy, context, and subjectivity as championed by human designers like Lupi, and the AI was able to manifest those values in its output. This suggests that prompt-based steering is a viable mechanism to inject higher-level human values (like the importance of personal stories in data) into AI-generated content.

In conclusion, the cross-persona evaluation demonstrates that prompt-guided personas can greatly enhance the flexibility and audience-tailoring of AI storytelling systems. An AI model like GPT-4 can serve multiple roles – statistician, strategist, storyteller, reporter – and produce coherent, relevant narratives for each, as long as it is given clear contextual cues. The inclusion of modes like Data Humanism further pushes the boundary, showing that even creative and empathetic storytelling, which is to be a uniquely human forte, can be approximated by an AI through the clever use of prompting. This reinforces the broader direction of the thesis: exploring how prompt-based AI can be leveraged to construct meaningful, personalized, and human-aligned data narratives. By bridging data science and narrative craft via AI, it is clear that the LLM not only analyze data but also communicate it in ways that resonate with diverse human audiences.

4.5 Limitations and Threats to Validity

The proposed storytelling system is highly dependent on the capabilities and constraints of the underlying language model. In this case, OpenAI’s GPT-4 as the exclusive engine for narrative and code generation. While GPT-4 significantly outperforms earlier models on many benchmarks, the model is “not fully reliable” [24]. It can still “hallucinate” plausible sounding but factually incorrect statements [24, 25].

Another practical limitation is prompt sensibility. Large Language models are known to be highly sensitive to even minor changes in prompt phrasing [26]. Slight rewordings or structural

tweaks in the instructions could lead to substantial differences in the output (both narrative and code). This sensitivity introduces variance in results and challenges reproducibility. The LLM might preferentially produce content that aligns with patterns it implicitly learned during prompt engineering, rather than genuinely generalizing from the data. This can limit the robustness of the system: different prompt designs or future model updates may yield different behaviors.

The scope and representativeness of the dataset used for narrative generalization also constrain validity. Any patterns or insights the model finds are conditioned on the uploaded data; unanticipated data structures or categories could yield odd or superficial narratives.

4.6 Ethical, Privacy, and Responsible-Use Considerations

The deployment of an AI-driven storytelling tool raises several ethical concerns that must be addressed. First, the use of personas could inadvertently propagate bias or stereotypes. During test the model was prompted across multiple personas and challenged in its output behavior, but no bias or hallucinations were found. However, the possibility remains that some narratives could reinforce unintended biases. Responsible use requires monitoring for harmful stereotypes and providing users with disclaimer about potential bias in AI-generated narratives.

Privacy is another critical issue, especially since the system processes user-uploaded data. Even if the data are not personally identifiable, users may inadvertently upload sensitive information. Since the system uses OpenAI API to generate narratives, the content of user data is sent to OpenAI's servers during API calls. Importantly, OpenAI's policy for business/API users is that by default inputs are not used for training their models [27]. It's recommended to implement features such as data anonymization, encryption at rest, and secure deletion after session end. In a production setting, compliance with data protection regulation is mandatory, and the system must adhere to such principles by minimizing data retention and ensure confidentiality.

A further concern is misinformation risk, since the output came from the user dataset, the LLM does not spread false conclusion, but any numerical insight in the narrative should be double-checked against the source data. Regarding authorship and accountability, it is crucial to recognize that AI systems cannot be held legally responsible for content. All outputs must be considered co-created with the AI, but ultimately the human user or author is accountable. According to publication ethics guidelines, AI tools "cannot replace" human authorship and cannot assume legal responsibility [28]. The human operator must review, edit, and own the final product.

There is also a potential for persuasive misuse. Data-driven narratives can be quite emotive, especially in Data Humanism mode. This could be used to craft highly influential or manipulative

presentations. Training or alerts to discourage misleading framing may be warranted. This aligns with the UNESCO principle of “Do No Harm”: systems must not use persuasive to mislead influence audiences.

The system was documented transparently in its behavior and limitations and by offering explanations of its analytical steps. Responsible use of the storytelling system requires a human-in-loop approach, adherence to data privacy norms, vigilance against bias and misinformation, and full accountability by the users leveraging its outputs. The considerations draw on observation of GPT-4 known limitations, research into persona bias in LLMs, openAI data-use policies and established AI ethics guidelines. Each point is grounded in current best practices for ethical AI system design and deployment.

Having examined the LLM’s autonomous data storytelling through various persona lenses and evaluated its alignment with data humanism ideals, the next step is to situate these AI-generated stories in relation to human-crafted ones. The following section will compare the AI’s outputs with data stories produced by human designers, to discern where the machine mirrors human creativity, where it diverges, and what this means for the future of data-driven storytelling.

5. Comparing Human vs AI Data Visualization Storytelling

In this chapter, a comparative analysis is presented through a user survey, of three pairs of visualizations: each pair consisting of a hand-crafted data-humanism design by Giorgia Lupi and AI-generated interactive D3 counterpart. The goal is to evaluate these visual stories across comprehension, insightfulness, memorability, trust and interactivity. The study examined:

1. “*Nobels, No Degrees*” : Giorgia Lupi’s award-winning visualization of Nobel laureates’ education backgrounds, versus an AI-generated version of the same data.
2. “*European Banks and Government Debt*”: Giorgia Lupi’s dense depiction of bank exposures to sovereign debt, versus its AI counterpart.
3. “*Geniuses, Visualized*”: a creation of the designer Giorgia Lupi of 100 literary geniuses versus an AI-generated D3 implementation.

For each pair, firstly is described the designs objectively, detailing layout, visual encodings, style and narrative approach, then analyzed the user feedback from the survey on how each performed in conveying understanding, stimulating insight, inspired trust and invited interaction. The analysis reveals when and why participants preferred the humanist approach versus the AI-generated approach, providing practical guidance on which style of visualization is better suited for different communication goals, for example, when an emotive narrative is more effective than a straightforward interactive dashboard, and vice versa. Participants were not in knowledge that human visualization were actually made by Giorgia Lupi, so their judgments were out from potential bias and based solely on the content and design of the visuals.

5.1 Visualization Pairs: Human-Made vs AI-Generated Designs

Each pair of visualizations told the same data story through two different lenses: a human, handcrafted approach and an AI-driven, code-based approach. Below, its described each pair’s content and design, highlighting differences in how they communicate information and emotion.

5.1.1 Nobels, No Degrees

The first pair consist of Giorgia Lupi’s “*Nobels, no degree*”, an information-rich static visualization exploring Nobel Prize laureates from 1901 to 2012 and highlighting those who lacked advanced degrees (Figure 5.1)[29]. Lupi’s piece, originally published in *La Lettura* (Corriere della Sera), is a dense, multi-layered print infographic combining several coordinates[30]. It features six horizontal lanes for each Nobel category (Physics, Chemistry, Literature, Peace and Economics), with a time axis running left to right and a vertical age axis showing the laureates’ age at the moment of

the award [30]. Each circle is a laureate plotted at the intersection of year and age; marker styling distinguishes formal qualification status (as specified in the “How to read it” key), while lane color follows the Nobel discipline. The page is heavily annotated: small inset summaries, year call-outs and footnotes surfaces notable cases (youngest/oldest winners, first awards in a field, etc.). On the right, a Sankey bundle connects laureates to a short list of university affiliations (e.g., Harvard, MIT, Stanford, Caltech, Columbia, Cambridge, Berkeley), emphasizing the connection between disciplines and institutions. Additional micro-charts at the bottom aggregate counts and breakdowns.

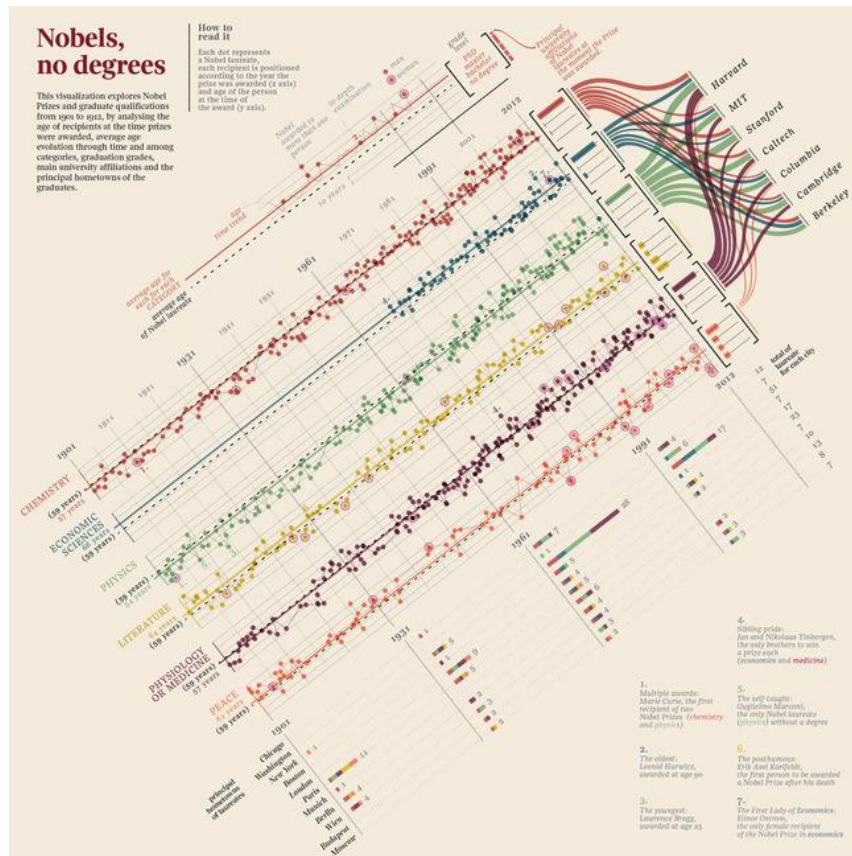


Figure 5.1: *Nobels, no degrees* (Giorgia Lupi / Accurat, La Lettura, Corriere della Sera, 2013)

The effect is very narrative, since viewers can scan trajectories within and across fields (e.g., shifts in age profiles over decades) while also perceiving institutional flows. At the same time, the graphic demands decoding. Multiple simultaneous encodings (year, age, field, degree status, institutional ties) and the bespoke typography/legend mean that comprehension is cumulative rather than immediate.

The AI-generated counterpart, as showed in Figure 5.2, took a more conventional data visualization approach, prioritizing structure and clarity. The D3.js version of “Nobels, no degrees” adopts a small-multiples layout with six horizontal rows: Chemistry, Economics, Physics, Literature,

Medicine and Peace, arrayed against a single time axis (1900-2020+). Each dot is a laureate; the stroke of the dot encodes qualification status (filled=degree, hollow=no degree, dashed=unknown), while row colour follows the prize category. The background marks historical periods (“World War I”, “World War II”, “Cold War”), and a note flags that Economics begins in 1969. A compact legend at right explains both the academic-status symbols and the category colours.

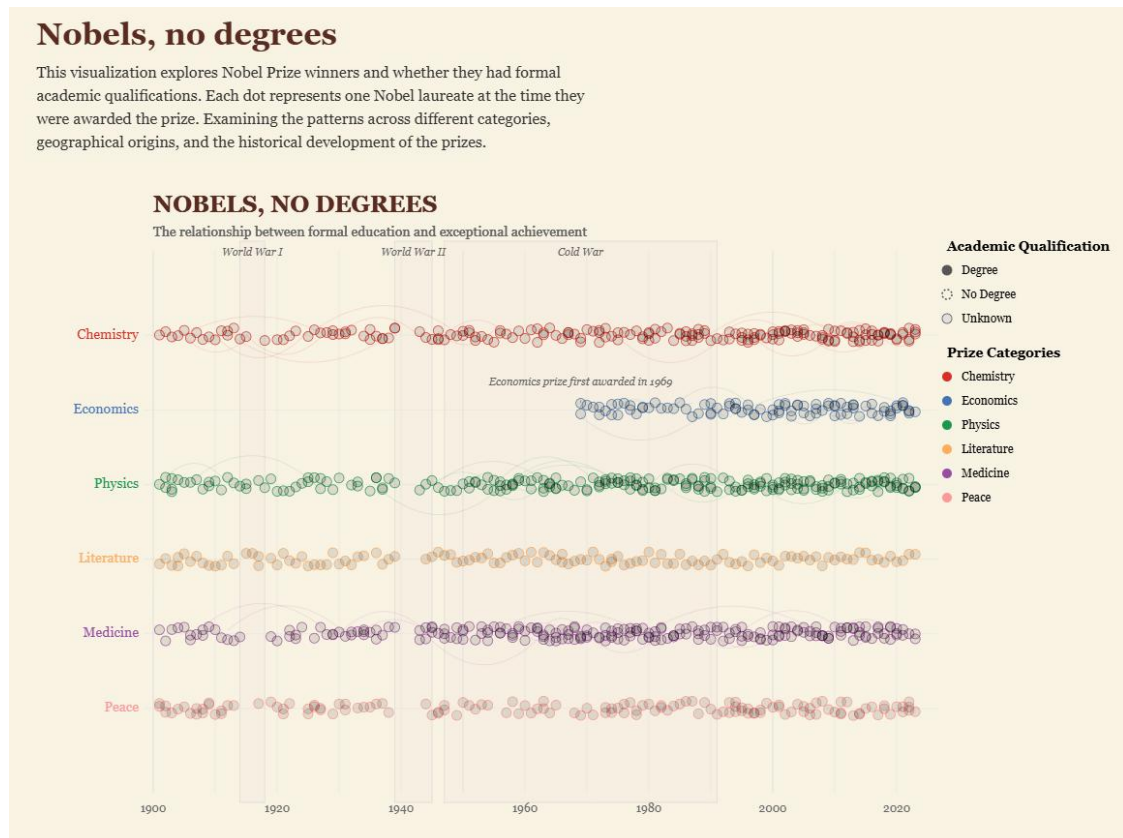


Figure 5.2: AI-Generated reinterpretation of Nobels, no degrees

Interactivity privileges clarity over ornament. On hover, tooltips surface the laureate’s name, year and qualification status; light filters allow users to isolate a category or a status. The design omits secondary encodings found in the previous human version, there is no age axis and no institutional flow diagram, which reduces cognitive load and simplify the narrative on the core analysis on the laureates degree. The interactive functionality was a major advantage; as respondents noted, being able to click on year or categories made the data exploration easier and more engaging in a hands-on way, whereas the human version relied solely on static depiction.

5.1.2 European Banks and Government Debt

The second comparison examines a visualization that is arguably one of Lupi's most complex and acclaimed works [31]. Published in 2013, "*European Banks and Government Debt*" (Figure 5.3), depicts the tangle of financial relationships between 61 European banks and the 29 European countries' sovereign debt [32]. The piece earned the nickname "*the jellyfish chart*" due to its striking visual motif: at the top of the graphic, each country is represented as a large circle which indeed look like jellyfish "heads" floating in an ocean of data[31]. The countries' circles are arranged by two quantitative variables: horizontally by the country's Debt-to-GDP ratio (left to right, low to right) and vertically by population (bottom to top, smaller to larger population) [31]. The size of each country's circle encodes the total sovereign debt of the nation, providing a sense of scale.

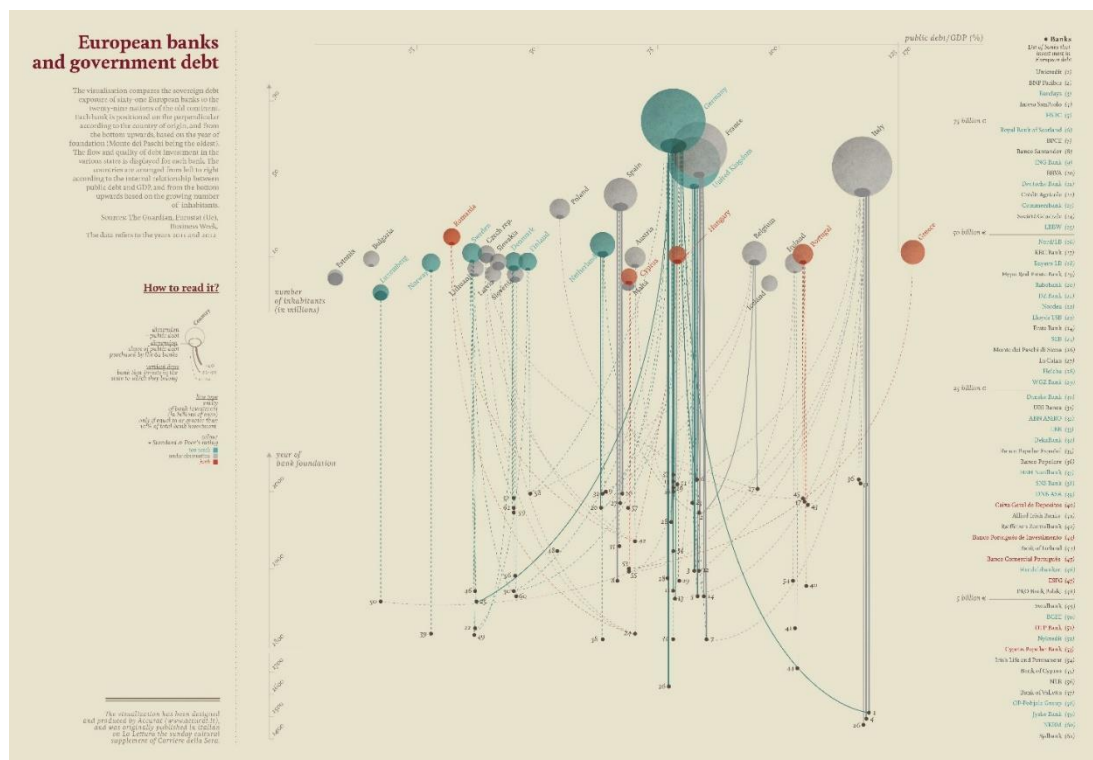


Figure 5.3: European Banks and Government Debt (Giorgia Lupi / Accurat, La Lettura, Corriere della Sera, 2013).

From each country circle, curved “tentacles” descend to the bank, which are arrayed along the bottom. A concise “How to read it” key and a numbered list of banks at the right margin keep the visualization readable. The warm palette and flowing curves create a narrative aesthetic that makes the visual of banks and debt emotive and felt before it is precisely read. The trade off here is similar to the previous, as a static work, extracting exact values or trading one institution requires cross-

referencing the legend and index. Overall, it a symbolic data humanism work, embracing complexity and turning a dense dataset and network into a visual storytelling piece.

The AI-generated visualization of the same financial dataset (Figure 5.4) is a interactive bubble scatterplot. It plots each bank as a circle with x-position equal to Public debt / GDP (%) of the bank's home country and y-position equal to the bank's sovereign exposure (EUR billions).

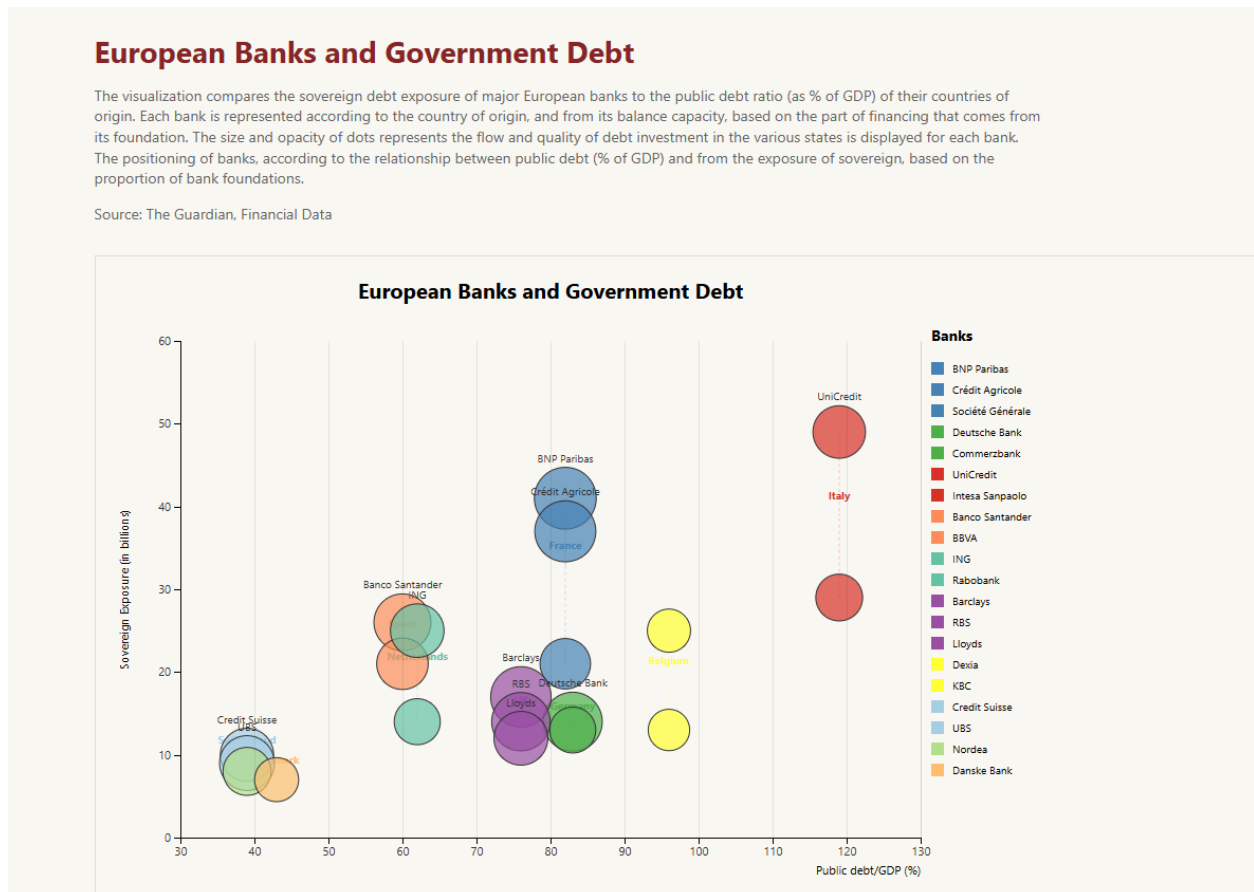


Figure 5.4: AI-Generated reinterpretation of European Banks and Government Debt

Circle size scales with the relative magnitude of exposure, while colour encodes the bank identity (legends at right lists the institutions). Subtle country annotations and dotted vertical guides (e.g., France, Belgium, Italy) contextualise clusters along the debt/GDP axis. Visually, the design adopts a clean dashboard aesthetic, with cartesian axes, restrained palette, and readable labels, prioritising legibility and comparison over metaphor. The interaction provides a key advantage for the user: the ability to explore specific banks or country contexts, making the principal relationships immediately apparent and easy to verify.

5.1.3 Geniuses, Visualized

The third pair moves from finance to culture, examining Giorgia Lupi's "*Geniuses, Visualized*" (Figure 5.5), a piece that maps 100 of history's greatest literary minds in a highly imaginative way. This visualization was part of Lupi's work for La Lettura and is inspired by the Kabbalistic Sephirot, a diagram from Jewish mysticism consisting of then incorconnected nodes [31]. Within every panel, individuals appear as small circles arrayed along a curved stem, resembling the tree-of-life motif and giving the piece a ceremonial tone.

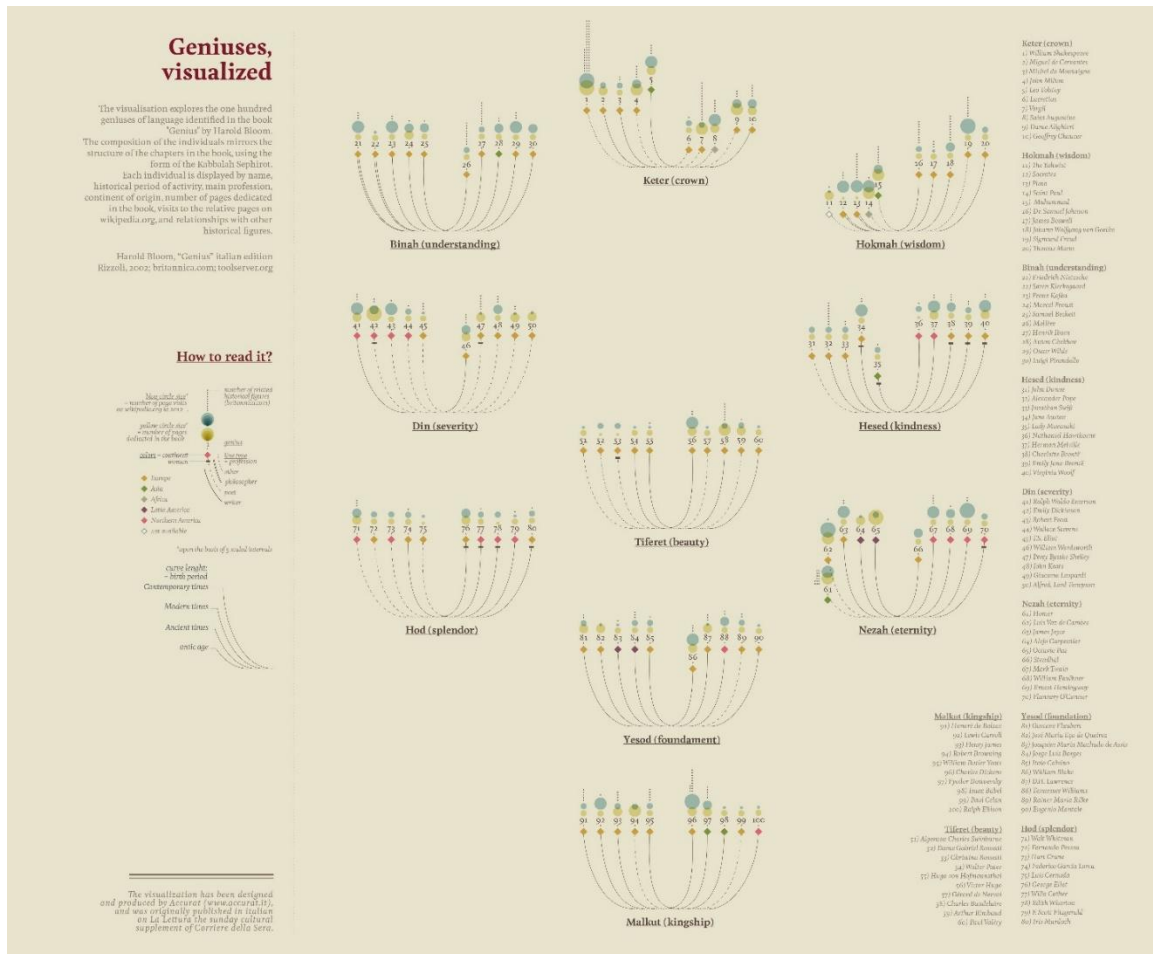


Figure 5.5: *Geniuses, visualized* (Giorgia Lupi / Accurat, *La Lettura*, *Corriere della Sera*, 2013).

Encoding is compact yet information-rich. Position along the arc indicates historical period (from ancient to contemporary), while circle size expresses a quantitative measure. Colour denotes continent of origin; tiny glyphs beside each mark register gender and main role (e.g., poet, philosopher). Discrete numbers printed near the circles correspond to a look-up list at the margins, where names are set in small type for each Sephirah. The composition privileges conceptual fidelity over graphic minimalism. Giorgia Lupi takes inspiration from a Harold Bloom book "*Genius: A*

Mosaic of One Hundred Exemplary Creative Minds”, where he identifies 100 geniuses of language (writers, philosophers, poets) and categorizes them into thematic groups to the Sephirot. This visualization takes Bloom’s concept and mirrors the schema of the book: categories such as Hokmah (wisdom) or Tiferet (beauty) serve as thematic clusters, inviting comparisons across time, geography and discipline.

This visualization is a perfect output of data humanism, it embraces complexity while remaining harmonious, it exemplifies Lupi’s belief that readers can be engaged through complexity if it is meaningful and beautifully rendered training the readers’s eye.

The AI rendering (Figure 5.6) preserves the Sephirot metaphor, ten hubs are drawn as a network, where around each hub the 100 figures are placed as small, numbered circles arranged chronologically. Colour encodes field of contributions (philosophy, literature, theology, poetry, theology, poetry, theatre, fiction), while node size represents relative Wikipedia page views. The huberic labels to the index at the right, where names are listed under each Sephirah; hovering reveals a tooltip with name, lifespan and role.

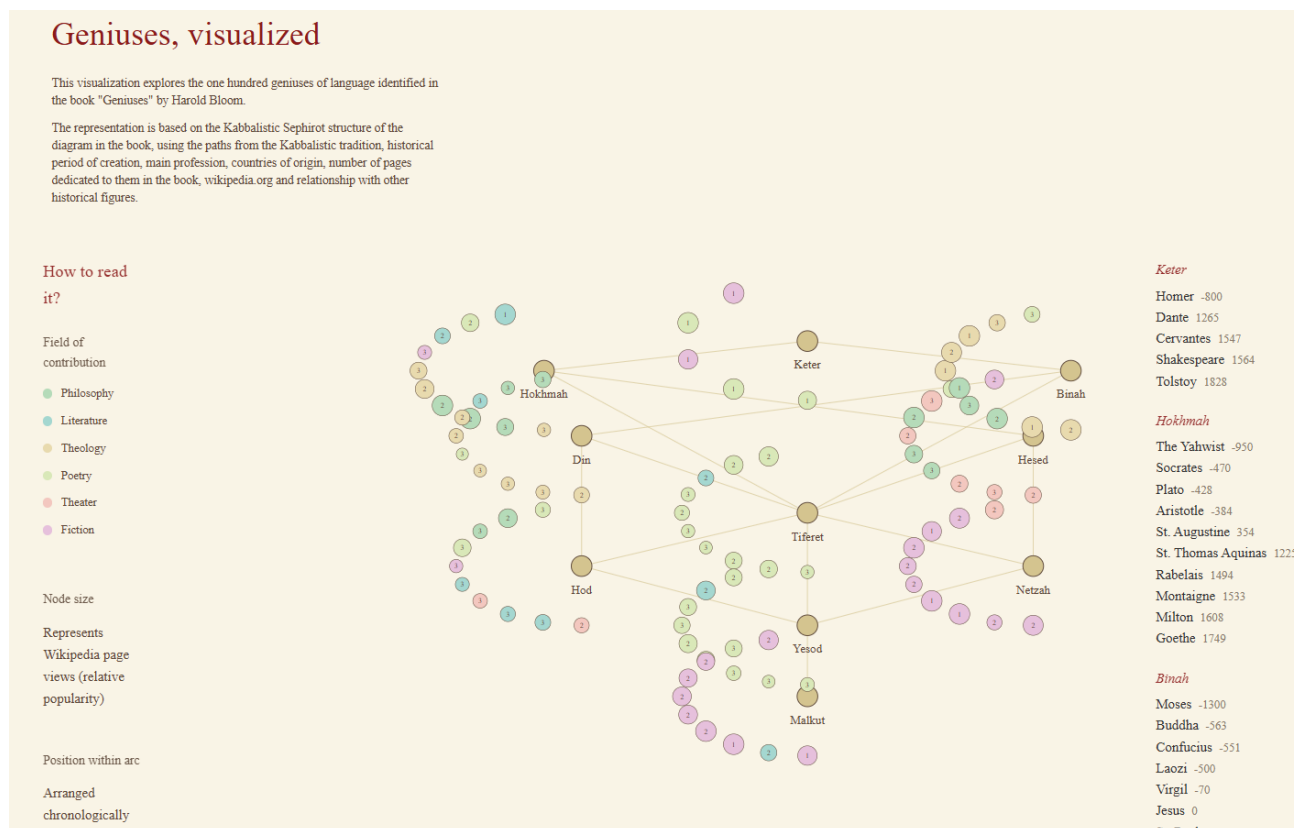


Figure 5.6: AI-Generated reinterpretation of *Geniuses, visualized*

Compared with the dense print composition, the D3.js version is cleaner and more legible: the conceptual legend is explicit, labels are neat, and interactive tooltips support quick lookup tasks.

At the same time, some of Bloom’s attributes present in the human made are simplified. The result is an analytically oriented, interactive homage to the original, retaining the symbolic structure while prioritising clarity and on-demand detail over ornamental complexity.

5.2 Quantitative Results

The survey comprised 25 items, combining 5-point Likerts scales (5-point, from “strongly disagree” to “strongly agree”) with forced-choices questions (Human, AI, Both equally, Neither) to capture participants’ perceptions on four core dimensions: ease of understanding, visual appeal, engagement and memorability, and confidence in the accuracy of the data. Additional open-ended prompts invited qualitative analysis of their responses for each visualization pair and across datasets, reporting summary statistics alongside comparative patterns.

A total of 47 participants completed the survey. The sample was recruited via academic and social channels and was heterogeneous in background, including students, professionals, and individuals with varied familiarity with data visualization. Detailed demographic breakdowns (e.g., age, gender, occupation) were not collected in order to maintain a concise instrument; therefore, responses are interpreted as representing a diverse but non-representative population.

This section presents a quantitative analysis of their responses for each visualization pair and across datasets, reporting summary statistics alongside comparative patterns.

5.2.1 Overall Ease of Understanding

Participants generally found all visualization sets reasonably understandable, though the AI-generated visuals had an edge in clarity. For each comparison the first questions regarded the ease of understanding and a question on a 5-point scale rating “*How easy was it to understand the content and core message of each visualization?*” the average scores were high for all three pairs (mean ≈ 4 out of 5, where 5 indicates “very easy”). The Nobel pair achieved the highest comprehension score (mean 4.13, $\sigma = 0.82$), followed by the Banks pair (mean 3.98) and the Geniuses pair (mean 3.87). This suggests that, overall, participants did not find any of the visualizations extremely confusing, likely due in part to the survey providing context and labeling. However, the slightly lower ease for “Geniuses” indicates that this pair was the most challenging to interpret, which aligns with the complex nature of mapping literature to philosophical categories.

When asked directly which visualization helped them better understand the data, a significant number of participants favored the AI-generated version for each dataset. For the Nobel laureates data, 44.7% of participants said the AI-visualization helped them understand the distribution of

laureates' degrees and categories over time better, compared to 25.5% who chose the humanistic visualization. The remaining respondents were either undecided or felt both were equal (about 21% chose "*both equally*", and 8.5% "*neither*"). The trend was even more pronounced for the Banks dataset: fully 64% found the AI visualization more comprehensive for understanding the relationship between banks and government debt, while only 6% chose the human one (the rest, ~ 30% said both were equal). This striking result underscores how much clearer the AI approach was for complex financial dataset. Finally for the Geniuses data, about 47% favored the AI visualization for understanding the connection between each genius and their Sephirot category, versus only 13% who found the human version more helpful (30% said both equally, and ~ 11% neither). In summary, across all three cases a plurality of respondents pointed to the AI-generated visualization as the better tool for understanding the factual content of the data. The human-made visualizations, despite their richer narrative design, were less often seen as the easier or clearer, especially for the most data-heavy scenario. These results quantitatively confirm a trade-off frequently discussed in visualization research: designs prioritizing narrative and artistic expression can sacrifice some immediacy of insights, while conventional, algorithmic visuals excel at quickly conveying structure and quantities [33].

5.2.2 Role of Interactivity

The survey continued with the study of the role of interactivity in the AI visualization and understanding. In the Nobel and Banks surveys, participants rated how effective the interactive elements were in helping them explore the data "*more effective than the static version*". In contrast, the visualization made by Giorgia Lupi obviously did not offer the interactivity, but the visual design helped in recognizing patterns. The average response was 4.04/5 indicating a strong agreement that both visualizations' structural design helped pattern recognition.

5.2.3 Visual Appeal and Aesthetics

Participants also rated how visually appealing they found each set of visualizations. All three pairs scored fairly high on visual appeal, though the Geniuses visualizations were the frontrunner with an average appeal rating of 4.21/5. The Banks pair followed at 3.94, and Nobels at 3.77. These ratings suggest that despite the differences in style, both the humanistic and AI visualizations were generally well-received aesthetically, the slightly higher score for Geniuses could be due to the inherently intriguing subject or the particular design choice, like vivid colors or forms.

However, when forced to choose, participants' preferences on visual appeal and "engagement" revealed contrasts. The survey asked which visualization felt most engaging and memorable for each

dataset. The results show a divergence across the three cases. For the European Banks data, more than 83% of participants found the humanistic visualization more engaging and memorable, while no one chose the AI visualization (the remaining 17% said both were equally engaging). This is a dramatic testament to the human design's impact: despite its lower clarity, its emotional and creative portrayal clearly gripped viewers' attention in a lasting way.

For the Geniuses pair, the majority also leaned humanistic, but less decisively: about 51% said the human visualization was more engaging/memorable, 38% favored the AI visualization, and the rest found neither particularly engaging (no one chose "both" for this question).

The Nobels No Degree results were the most evenly split. Approximately 36% of respondents found the humanistic Nobel visualization most engaging, 45% chose the AI visualization, and the remaining ~ 19% said both were equally engaging or neither was engaging. In the end, there was no clear consensus, a reflection of the two visualizations perhaps providing different kinds of engagement. It's notable that the AI Nobel visualization managed to slightly edge out the human one on engagement for many, possibly because the subject (Nobel laureates' education) might have been effectively communicated with a dynamic data-driven graphic, whereas the human attempt at storytelling did not resonate as strongly as in the other topics. This case suggests that not all human-centered visuals automatically captivate more; the quality and relevance of the narrative design matter, and in some cases an efficient chart can be quite engaging too.

5.2.4 Trust and Perceived Accuracy

One of the clearest quantitative outcomes of the survey is that participants placed greater confidence in the AI-generated visualizations' accuracy and reliability. When asked which visualization made them feel more confident about the data's correctness, the majority in all three cases chose the AI version. The difference was especially stark for the Banks and Geniuses visualizations: in the Banks survey, 87% of respondents said the AI visualization inspired more confidence in the accuracy of the data, with the remaining 13% saying "both equally", notably 0% chose the human one. For the Geniuses graph, 68% chose AI, 32% said both equally, and again chose just the human visualization. Even in the Nobels case, where the human design was not extremely unconventional, about 45% picked AI as giving more confidence, 23% picked human, and the rest felt both were equally trustworthy. Figure 5.7 illustrates this pronounced trust gap. Clearly, participants associated the AI-driven charts with greater precision and credibility.

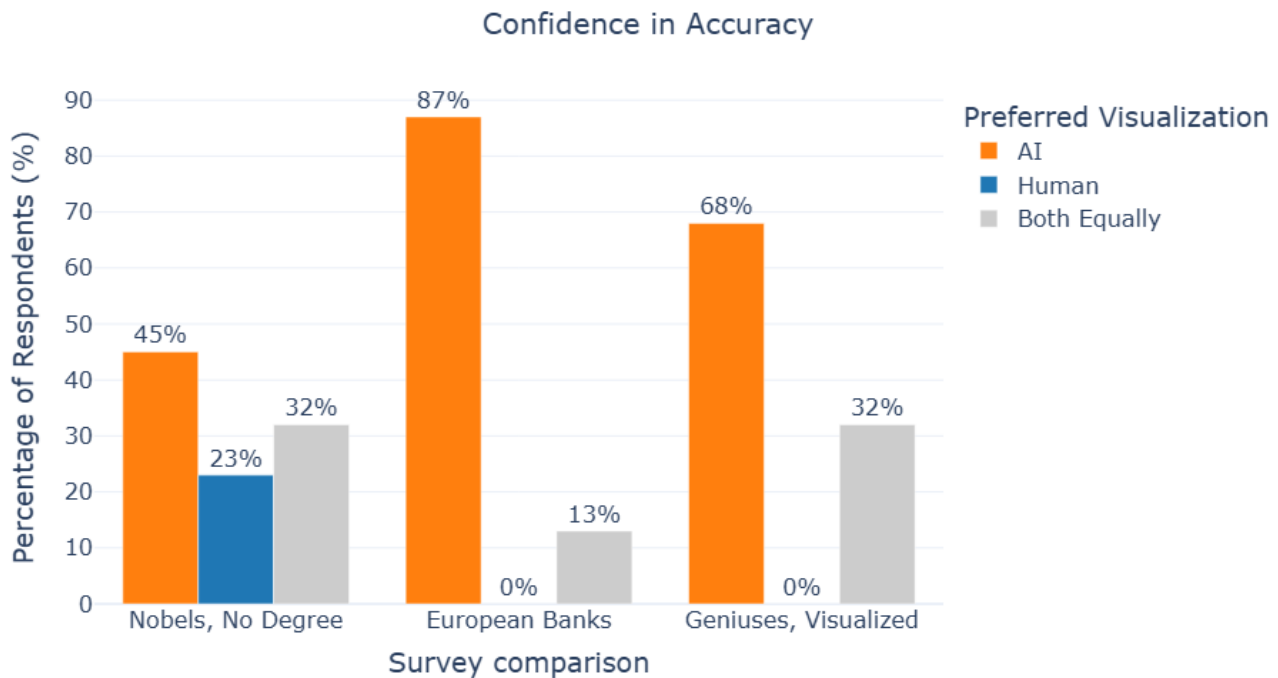


Figure 5.7: Confidence in Data Accuracy by Preferred Visualization

This trust differential can be understood in context. The AI visualization adhered to familiar, “scientific” design conventions, which tend to signal objectivity and authority to viewers. In contrast, the humanistic visuals, by embracing creative liberties and subjective storytelling, might inadvertently introduce doubt about rigor. This aligns with Lupi’s own critique of traditional infographics: the “*ostensibly neutral visual language*” of computer-generated charts carries an aura of reliability the participants clearly responded to that aura [33]. There is an irony: while data humanism argues for honesty and nuance over misleading simplicity, people still often trust computer chart more than interpretive one. This underscores the deep-seated perception that if a visualization looks like a straightforward, software-produced graph, it’s probably technically correct, whereas a hand-drawn visualization might be approximate or subjective.

Research in human-AI perception suggests that people can be biased in favor of human-created art or assign more objectivity to machine outputs. In this study’s context, the data was the same in both, but the AI versions gave participants confidence that the numbers were correctly represented (even if the human version never introduced errors). Indeed, the trust gap is one of the strengths and limitations of data humanism: by “embracing imperfection” and subjectivity, human visuals can appear less authoritative, which may invite skepticism or doubt in the data itself.

5.3 Qualitative Insights

Beyond the numbers, the survey collected open-ended feedback where participants explained why they preferred one visualization over the other for learning/presenting the data and why one felt more engaging or meaningful overall. A thematic analysis of these responses revealed several recurring themes in their reasoning: Clarity versus Complexity, Emotional and Personal Resonance, Aesthetic Appeal and Memorability, Interactivity and Exploration, and Purpose and Context.

5.3.1 Clarity vs. Complexity

A dominant theme concerned the clarity of presentation. AI-generated visualizations were frequently praised as “*clear, straightforward,*” and “*easier to read*”. The AI visuals include statements like: “*The LLM-generated version is clearer and more straightforward, making it easier to explain to an audience,*” and “*I prefer the AI chart since it presents the data in a logical, easy to follow way.*” One participant highlighted how clarity improved their trust, saying the AI version “allowed me to verify the numbers, so I felt more confident I understood it correctly.”

In contrast, the humanistic visualizations were often described as “*complex*”, “*busy*”, or “*harder to decipher*” without careful study. For instance, one respondent admitted, “*The human one has a very good design in my opinion, but it seems confusing at first.*” This captures a common sentiment: appreciation for the beauty of the design, coupled with an acknowledgment that it wasn’t instantly clear. That said, not everyone dismissed human visualizations as hopelessly confusion. They described the human pieces as “*layered*” or “*information-rich*”, implying that while they took longer to read, they offered more to discover. For example, for the “*Geniuses, visualized*” one participant summarized the difference like: the human visualization felt “*more personal and emotionally evocative*” while the AI one was “*clear and informative*”. This reveals an appreciation for complementarity: human visuals for narrative depth, AI visuals for factual clarity.

Overall, the AI-generated visualizations consistently scored higher on objective clarity and perceived accuracy, whereas the human-crafted visualization excelled in subjective engagement and emotional resonance.

5.3.2 Emotional and Personal Resonance

Another major theme was the emotional impact and personal connection engendered by the human crafted visualizations. Many participants reported that the humanistic versions “*felt more engaging emotionally,*” “*evoked curiosity and empathy,*” or “*made the data feel personal and meaningful.*”

Representative quotes illustrate this contrast. One participant, reflecting on the Geniuses visualizations, wrote: *“The design of the humanistic version makes the information feel personal and meaningful, connecting me to the stories of these geniuses.”* Similarly, for the Nobels dataset, a participant stated: *“The humanistic visualization felt more memorable because its aesthetic is more emotionally evocative”*.

In terms of empathy, one subtle observation is that participants sometimes anthropomorphized or gave life to the data when describing the human visuals. In the Banks case, where the subject is not people but banks and debts, one respondent still said the human version *“conveyed the gravity of debt exposure in a way that made me feel the impact.”* This indicates that even for an abstract topic, the human visualization invoked feeling, perhaps concern or awareness of the human consequences of economic data.

Not all participants valued emotional impact, of course. One participant preferring the AI for clarity conceded, *“the humanistic one had a certain warmth to it, but I worry some might misread it.”* These comments show that participants noticed the human author’s presence in the design, a hallmark of data humanism (the idea that the designer’s perspective and emotion can be present in the work).

5.3.3 Aesthetic Appeal and Memorability

Closely related to emotional resonance is the theme of aesthetic appeal and memorability. Participants often tied the visual style of a piece to how well it stuck in their memory or caught their interest. The humanistic visualizations, with their unique and artistic designs, were frequently lauded as more memorable. Participants used phrases like *“artistic style that stays in the memory,”* *“creative metaphor... grabs your attention immediately.”*

For the Nobels, No Degree comparison, several respondents praised it as *“beautiful”* and memorable: a piece one might *“hang on a wall”*, highlighting their cultural or decorative value. By contrast, AI-generated visuals were described as *“modern,”* *“sleek,”* or *“clean,”* but less distinctive. Occasionally, however, interactivity made the AI graphics more memorable, especially in the Geniuses case. A recurring idea was that memorability is connected to narrative and interpretation. Thus, the human visuals created memorable impressions by blending visual beauty with narrative meaning, whereas the AI visuals were memorable mostly for concrete information.

5.3.4 Interactivity and Exploration

Though only the AI visualizations in this study offered interactive features, the theme of interactivity emerged strongly in participants’ feedback, highlighting how it affected their experience.

Those who emphasized interactivity often did so in the context of engagement and understanding. They enjoyed being able to manipulate the visualization and felt it gave them control to explore the data at their own pace. In some cases, interactivity even contributed to memorability and trust (as discussed earlier).

Participants who valued interactivity made comments such as: “*The AI visualization is memorable because I could explore and confirm insights interactively,*” and “*The interactive exploration in the AI tool is more important to me because I can filter what I want to see.*” These quotes show two things: first, interactivity can itself be engaging and second, interactivity helps in learning. This did not necessarily give the AI visuals emotion depth, but it gave them functional depth, an important part of storytelling in data: the challenge is how to combine this strength with the human visualization’s narrative depth.

5.3.5 Purpose and Context

Finally, a concept that emerged from how participants responded is the consideration of purpose and context when choosing a visualization. Many respondents framed their answers in terms of context of use. The humanistic designs were considered better suited for outreach, exhibitions, or social media posts, where capturing attention and evoking emotion are primary goals. The AI-generated versions were viewed as preferable for presentations, classrooms, or professional reports requiring clarity and accuracy.

For example, one participant wrote: “*For a social media post or exhibition, I’d go with the humanistic one to captivate people. But for a report or class presentation, definitely the AI one, since it’s easier to explain and gets the facts across.*” This clearly delineates two contexts: in the first, grabbing attention is a role where the human visualization shines; in the second, clarity and completeness of information are key where the AI visualization excels.

This theme illustrates that many participants did not view human and AI visualizations as mutually exclusive. Instead, they recognized that each excels in different communicative contexts.

5.4 Synthesis of Findings

The survey results and participant insights provide a multifaceted picture of how human and AI-generated visualizations each contribute to data storytelling. Across all three comparisons, a clear pattern emerges: participants tended to prefer the AI-generated visualizations on criteria of comprehension, exploration, and trust, whereas they favored Giorgia Lupi’s hand-crafted visualizations for engagements, memorability, and aesthetic impact. Notably several respondents

reported that using both together produced the most complete experience: one for sense-making and verification, the other for meaning-making and reflection. By considering both, participants are, adopting a critical viewing practice: comparing perspectives to get a fuller truth. This is a very interesting outcome, it suggests that exposing people to a human vs AI dichotomy might encourage critical thinking about data representation, which is a goal of critical data studies.

In conclusion, the survey underscores that human and AI approaches to data visualization storytelling each have distinct strengths that can potentially complement each other. Human-crafted visualizations achieve the aims of data humanism, providing empathy and narrative, but can suffer in perceived objectivity and immediacy, AI-generated visualization excel in clarity, credibility, and efficiency, but lack the creative spark that makes data meaningful. The narrative emerging from participants is one of the syntheses: leveraging the best of both. The presence of both versions prompted viewers to appreciate different facets of the data, a result that aligns with the notion that combining analytical and humanistic evidence gives a more holistic understanding [1].

6. Conclusion and Further Developments

This thesis developed and empirically evaluated an AI-driven data storytelling platform to animate data in a human-centered way. Operationalizing Giorgia Lupi’s Data Humanism principles in software, the system transforms raw datasets into richly contextualized stories that blend text and graphics. Through prompt engineering and system design, the platform automatically generates personalized visual explanations under both conventional and humanistic paradigms, integrating GPT-4 for narrative generation and D3.js for visualization.

The evaluation of the system revealed several key findings. In the cross-persona case study (Chapter 4) role-specific test successfully produces distinct narrative styles. The Data Humanism runs yielded more metaphorical visuals, demonstrating that GPT-4 can be steered towards richer more explorative storytelling. These results support the view that generative AI can “transform raw numbers into natural-language summaries” and adapt its output to audience needs.

The user survey (Chapter 5) uncovered clear trade-offs between AI-generated and human-crafted stories. Indicating a conclusion synthesis: AI and human storytelling can complement each other. A central question of this work was whether generative AI can reinforce the human-centric values of Data Humanism. Indeed, they proved able to inject narrative elements into charts on demand, effectively augmenting user engagement with the data. At the same time, the study highlighted limitations. LLM sometimes produced plausible sounding but generic text when prompts were underspecified. While generative AI can reinforce Data Humanism by rapidly generating personalized, storylike content, it can also dilute the very human touch it aims to emulate.

The case studies and user survey were exploratory and tested on multiple dataset and personas, looking ahead a development would include a expanded evaluation. Conduct longitudinal studies to test how AI-enhanced storytelling affects learning, retention or trust over time, and exploring methods to fiscole On the technical side, only GPT-4 was used; emerging models with multimodal or interactive capabilities (e.g., GPT-5 or vision-language models) could further enhance storytelling.

In conclusion, this research indicate that generative AI can augment human-centric data storytelling but not replace its core values. The future of data storytelling will likely be collaborative: designers and analysts working alongside intelligent assistants to craft stories that are both analytically rigorous and richly human. By acknowledging the present limitations and pursuing the outlined directions, subsequent work can strengthen this synergy, advancing a data visualization paradigm that is powerful, personalized, and profoundly human.

References

- [1] R. McKinsey and D. Kiron, “The Enduring Power of Data Storytelling in the Generative AI Era,” *MIT Sloan Management Review*, Sep. 2024. [Online]. Available: <https://sloanreview.mit.edu/article/the-enduring-power-of-data-storytelling-in-the-generative-ai-era/>
- [2] G. Lupi, “Data Humanism — My Manifesto for a New Data World,” *giorgialupi.com*, 2017. [Online]. Available: <http://giorgialupi.com/data-humanism-my-manifesto-for-a-new-data-world>
- [3] C. Yan, C. Correll, and B. Lee, “Data Humanism Decoded: A Characterization of its Principles to Bridge Data Visualization Researchers and Practitioners,” *arXiv preprint*, arXiv:2509.00440, Sep. 2025. [Online]. Available: <https://arxiv.org/html/2509.00440v1>
- [4] M. AI Editorial, “How Generative AI Enhances Data Storytelling,” *Magai*, 2024. [Online]. Available: <https://magai.co/how-generative-ai-enhances-data-storytelling/>
- [5] A. Marini, P. Ceravolo, and E. Damiani, “AI-DaSt: An AI-based System to Enhance Data Storytelling,” in *Proceedings of the 15th International Conference on Data Visualization Theory and Applications (IVAPP)*, 2023, pp. 25–36. [Online]. Available: <https://www.scitepress.org/Papers/2023/122518/122518.pdf>
- [6] Domestic Data Streamers, “Mapping the Uncharted: What Can Gen-AI (Really) Do for Data Visualization?,” *Medium*, Jul. 2023. [Online]. Available: <https://domesticdatastreamers.medium.com/mapping-the-uncharted-what-can-gen-ai-really-do-for-data-visualizationfae58862beeb>
- [7] G. Lupi, “The Data We Do Not See: An Interview with Giorgia Lupi,” *Nightingale*, Apr. 2021. [Online]. Available: <https://nightingaledvs.com/the-data-we-do-not-see-an-interview-with-giorgia-lupi/>
- [8] D. A. Keim *et al.*, “Telling Stories with Data: A Systematic Review,” *arXiv preprint*, arXiv:2312.01164, Dec. 2023. [Online]. Available: <https://arxiv.org/html/2312.01164v1>
- [9] P. Colombo, R. M. Amorim, and C. Ciuccarelli, “Reciportrait: A Data Humanism Approach for Collaborative Sensemaking of Personal Data,” *TU Delft Research Portal*, 2022. [Online]. Available: <https://research.tudelft.nl/en/publications/reciportrait-a-data-humanism-approach-for-collaborative-sensemaking>
- [10] A. Battle, J. Harper, and C. North, “Generative AI for Visualization: State of the Art and Future Directions,” *arXiv preprint*, arXiv:2404.18144, Apr. 2024. [Online]. Available: <https://arxiv.org/abs/2404.18144>
- [11] LangChain – Wikipedia. [Online]. Available: <https://en.wikipedia.org/wiki/LangChain>
- [12] OpenAI, “GPT-4 Technical Report,” *arXiv preprint*, arXiv:2303.08774, Mar. 2023. [Online]. Available: <https://arxiv.org/abs/2303.08774>

- [13] Redpanda, “openai_chat_completion | Redpanda Connect,” Documentation, 2025. [Online]. Available: https://docs.redpanda.com/redpanda-connect/components/processors/openai_chat_completion/
- [14] Microsoft, “Prompt engineering techniques - Azure OpenAI,” Microsoft Learn, 2025. [Online]. Available: <https://learn.microsoft.com/en-us/azure/ai-foundry/openai/concepts/prompt-engineering>
- [15] A. Holtzman, J. Buys, M. Forbes, and Y. Choi, “The Curious Case of Neural Text Degeneration,” arXiv preprint, arXiv:1904.09751, Apr. 2019. [Online]. Available: <https://arxiv.org/abs/1904.09751>
- [16] LearnPrompting, “Role Prompting: Guide LLMs with Persona-Based Tasks,” 2025. [Online]. Available: https://learnprompting.org/docs/advanced/zero_shot/role_prompting
- [17] Reddit, “Google dropped a 68-page prompt engineering guide, here's what's inside,” Reddit r/PromptEngineering, 2025. [Online]. Available: https://www.reddit.com/r/PromptEngineering/comments/1kggmh0/google_dropped_a_68page_prompt_engineering_guide/
- [18] AakashG, “Prompt Engineering in 2025: The Latest Best Practices,” AakashG News, 2025. [Online]. Available: <https://www.news.aakashg.com/p/prompt-engineering>
- [19] S. Hajebi, “Best Practices in Prompt Engineering: A Real-World Guide,” Medium, 2025. [Online]. Available: <https://medium.com/@saeedhajebi/best-practices-in-prompt-engineering-a-real-world-guide-2781d192d60d>
- [20] S. Kumar and J. Lee, “Using AI for User Representation: An Analysis of 83 Persona Prompts,” arXiv preprint, arXiv:2508.13047, Aug. 2025. [Online]. Available: <https://arxiv.org/html/2508.13047v1>
- [21] D3.js – Wikipedia. [Online]. Available: <https://en.wikipedia.org/wiki/D3.js>
- [22] B. Xu, A. Yang, J. Lin, Q. Wang, C. Zhou, Y. Zhang, and Z. Mao, “ExpertPrompting: Instructing Large Language Models to be Distinguished Experts,” *arXiv preprint* arXiv:2305.14688, 2023. Available: <https://arxiv.org/abs/2305.14688>
- [23] “Data Humanism for Interactive Data Visualization,” Department of Informatics, University of Zurich. [Online]. Available: <https://www.ifi.uzh.ch/en/ivda/applications/data-humanism.html>.
- [24] OpenAI, “GPT-4,” OpenAI. [Online]. Available: <https://openai.com/index/gpt-4-research/>
- [25] “GPTs and Hallucination – Communications of the ACM,” Communications of the ACM. [Online]. Available: <https://cacm.acm.org/practice/gpts-and-hallucination/>
- [26] “Benchmarking Prompt Sensitivity in Large Language Models,” *arXiv preprint* arXiv:2502.06065v1, 2025. [Online]. Available: <https://arxiv.org/html/2502.06065v1>

[27] OpenAI, “How your data is used to improve model performance,” OpenAI Help Center. [Online]. Available: <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance>

[28] UNESCO, “Ethics of Artificial Intelligence,” UNESCO. [Online]. Available: <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

[29] Behance, “Nobel, no degrees,” Behance.net, https://www.behance.net/gallery/14159439/Nobel-no-degrees?locale=en_US

[30] “Approximating the components of Lupi’s Nobel, no degrees,” Spencer (GitHub Pages), <https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/>.

[31] G. Lupi, *Beautiful Reasons*, WordPress, 2014. [Online]. Available: <https://giorgialupi.files.wordpress.com/2014/05/beautiful-reasons-giorgialupi.pdf>

[32] Behance, “European Banks and Government Debt,” Behance.net. [Online]. Available: https://www.behance.net/gallery/14283209/European-banks-and-government-debt?locale=en_US

[33] “Can Data Be Human? The Work of Giorgia Lupi,” *The New Yorker*. [Online]. Available: <https://www.newyorker.com/culture/culture-desk/can-data-be-human-the-work-of-giorgia-lupi>

Appendix A

Survey Questionnaire

This appendix reproduced the full questionnaire administered in the survey component of this thesis. Participants, characterized as data-literate non experts, were asked for each pair of visualizations to respond to Likert-scale items, forced-choice questions, and open-ended questions. This material is reproduced to provide full transparency of the items used for qualitative and quantitative analyses (see Chapter 5).

Human vs AI: Comparing Visual Storytelling in Data Humanism

This academic survey explores how people perceive and interpret different types of data visualizations: one created by a human designer, and the other generated by a Large Language Model (LLM) trained on the same dataset.

The goal is to evaluate and compare the visualizations based on clarity, emotional engagement, trust, aesthetics and human meaning.

You will be presented with three pairs of visualizations based on real-world datasets. For each pair, you will be asked a series of questions regarding your understanding, visual preference and interpretation of the data.

There are no right or wrong answers.

Your responses will help advance research on how different visual styles affect data storytelling, audience comprehension, and empathy in data interpretation.

Comparison 1: "Nobels, No Degrees"

This section compares two visual interpretations of Nobel Prize winner from 1901 to the present.

Both visualizations use the same dataset and show laureates across categories and years, but with different visual styles and narrative approaches.

Please examine both before answering the following questions.

A. Human Made (*Figure 5.1: Nobels, no degrees (Giorgia Lupi / Accurat, La Lettura, Corriere della Sera, 2013)*)

B. AI generated (*Figure 5.2: AI-Generated reinterpretation of Nobels, no degrees*)

click here to interact: <https://em-rg.github.io/d3visual/>

1. How easy was it to understand the content and core message of each visualization?

Scale: 1 (Very difficult) – 5 (Very easy)

2. Which visualization helped you better understand the distribution of Nobel laureates' degree and categories over time? (Forced-choice question)

- A. Human Made
 - B. AI- Generated
 - C. Both equally
 - D. Neither
3. Did the interactive visualization help you explore the data (categories, years, degree vs no degree) more effectively than the static versions?
- Scale: 1 (Not helpful at all) – 5 (Extremely helpful)
4. How visually appealing do you find each visualization?
- Scale: 1 (Not appealing at all) – 5 (Extremely appealing)
5. Which visualization felt most engaging and memorable in presenting the journey of Nobel laureates? (Forced-choice question)
- A. Human Made
 - B. AI- Generated
 - C. Both equally
 - D. Neither
6. Which visualization made you feel more confident about the accuracy and reliability of the data shown? (Forced-choice question)
- A. Human Made
 - B. AI- Generated
 - C. Both equally
 - D. Neither
7. Which visualization would you prefer to use for learning or presenting this dataset, and why? (Open-ended question)
8. Overall, which visualization felt more engaging or memorable in representing the journey of Nobel laureates? Why? (Open-ended question)

Comparison 2: " European Banks and Government Debt"

This section compares two visualizations that depict the relationship between European banks and government debt in 2011–2012.

Both visualizations use the same dataset and display the sovereign debt exposure of major banks and public debt as a percentage of GDP, but with different visual styles and narrative approaches.

Please examine both before answering the following questions

A. Human made (*Figure 5.3: European Banks and Government Debt (Giorgia Lupi / Accurat, La Lettura, Corriere della Sera, 2013).*

B. AI generated (*Figure 5.4: AI-Generated reinterpretation of European Banks and Government Debt*)

click here to interact: <https://em-rg.github.io/3visual2/>

1. How easy was it to understand the content and core message of each visualization?
Scale: 1 (Not clear at all) – 5 (Extremely clear)
2. Which visualization helped you better understand the relationship between European banks and government debt? (Forced-choice question)
 - ☐ A. Human Made
 - ☐ B. AI- Generated
 - ☐ C. Both equally
 - ☐ D. Neither
3. Did the interactive visualization help you explore the data (countries, banks, debt levels) more effectively than the static version?
Scale: 1 (Not at all) – 5 (Very much)
How visually appealing did you find each visualization?
Scale: 1 (Not appealing at all) – 5 (Extremely appealing)
4. Which visualization felt most engaging and memorable in showing the banks' debt exposure and the broader economic context? (Forced-choice question)
 - ☐ A. Human Made
 - ☐ B. AI- Generated
 - ☐ C. Both equally
 - ☐ D. Neither
5. Which visualization made you feel more confident about the accuracy and reliability of the data shown? (Force-choice question)
 - ☐ A. Human Made
 - ☐ B. AI- Generated
 - ☐ C. Both equally
 - ☐ D. Neither
6. Which visualization would you prefer to use for learning or presenting this dataset, and why? (Open-ended question)
7. Overall, which visualization felt more engaging or memorable in representing the relationship between European banks and government debt? Why? (Open-ended question)

Section 3: "Geniuses, visualized"

This section compares two visualizations based on Harold Bloom's *Geniuses* (2002). Both display one hundred "geniuses of language" following the Kabbalistic Sephirot structure, showing their historical period, field of contribution, and relative prominence.

Please review both visualizations before answering the following questions.

A. Human made (*Figure 5.5: Geniuses, visualized* (Giorgia Lupi / Accurat, *La Lettura, Corriere della Sera*, 2013).

B. AI generated (*Figure 5.6: AI-Generated reinterpretation of Geniuses, visualized*)

click here to interact: <https://em-rg.github.io/geniusesd3try3/>

1. How easy was it to understand the content and core message of each visualization ?

Scale: 1 (Not clear at all) – 5 (Extremely clear)

2. Which visualization helped you better understand the connection between each genius and their literary–philosophical category (Sephirot)? (Forced-choice question)

- ☐ A. Human Made
- ☐ B. AI- Generated
- ☐ C. Both equally
- ☐ D. Neither

3. Did the visual structure (colors, shapes, positions) help you recognize patterns among the literary geniuses and their categories?

Scale: 1 (Not clear at all) – 5 (Extremely clear)

4. How visually appealing did you find each visualization?

Scale: 1 (Not appealing at all) – 5 (Extremely appealing)

5. Which visualization felt more engaging or memorable in presenting the "Geniuses of Language"? (Forced-choice question)

- ☐ A. Human Made
- ☐ B. AI- Generated
- ☐ C. Both equally
- ☐ D. Neither

6. Which visualization gave you more confidence that the data was represented accurately? (Forced-choice question)

- ☐ A. Human Made
- ☐ B. AI- Generated
- ☐ C. Both equally
- ☐ D. Neither

7. Which visualization would you prefer to use for teaching or presenting this literary dataset, and why? (Open-ended question)

8. Overall, which visualization made you reflect more on the cultural or emotional meaning of these literary geniuses? Why? (Open-ended question)

Thank you for your time!

Your answers will help our research on visual storytelling and data perception.