

# Using ChatGPT for storytelling from structured data

Ana Costa e Silva, Shreshtha Sarkar, Udaybhanu Gaddam, Pravin Shelke, Amit Channappagol, Benjamin Schannes, Giovana Gelatti, Ashwini Pandit, Amit K. Gupta, Nikhil Kolthankar, Saminda Dhillon, Kate Bravery

Mercer Inc., New York, USA  
kate.bravery@mercer.com

**Abstract.** This article presents a novel use of ChatGPT to gain insight over a quantitative data survey of 12000+ respondents. This is achieved by using a data pipeline that creates the data aggregations that deserve comment and generating text to “read them out loud”. We then add these summaries to a vector database and use ChatGPT 3.5, aka the cheaper GPT, to answer end user questions through the addition of a carefully iterated prompt that ensures the LLM’s answers adhere to a baseline of what makes useful insight. As a result, we created a path to saving hundreds of hours of laborious manual commentary of said aggregations. Finally, we used unsupervised learning techniques to highlight multivariate axes of divergence and personas, that informed the global report.

**Keywords:** Generative AI, survey analysis, prompt engineering, data storytelling.

## 1 Objectives

Mercer’s Global Talent Trends study is one of the flagship offerings from Marsh McLennan and brings in million + revenue globally in associated consulting work off the back of survey presentations. Every 2 years we conduct a comprehensive survey of more than 12000+ employees, HR professionals, investors, and C-Suite Execs across the globe. As such this is a robust database but the time to derive country and industry insights is labor and time intensive. The survey gathers insights across diverse topics including business priorities for the year ahead, employee sentiment regarding why workers plan to stay, leave, what drives their sense of thriving at work and trending skills. It also juxtaposes Executives’ views on priorities and HR’s practices – including topics such as ESG, Cyber risk and AI, along with other workforce initiatives. The results culminate in a focused report called Mercer’s Global Talent Trends.

### 1.1 The Global Talent Trends Report (GTT)

Report extract: “As we look ahead to 2024, the world of work is in full metamorphosis, forever changed by the seismic shifts of the past few years. Societal dynamics are reshaping work structures, technology and AI are advancing at breakneck speed, and we are still navigating the aftermath and recovery of a post-pandemic world.

We wanted to hear how business executives, HR leaders, employees and investors were navigating this evolving landscape. More than 12 200 voices contributed to this year’s Global Talent Trends study [...]. This year’s Global Talent Trends study draws on insights from 845 C-Suite executives, 1 920 HR leaders, 9 449 employees and 84 investors from across 17 geographies and 16 industries. Their input defined the four trends that are shaping the People agenda in 2024” [2]. Although there are 4 surveys that make up the 2024 study, for the purpose of this work, we took the different personas with valid data sets across 17 markets and 16 industries. The three core studies that are focused on insights from three personas are: C-Suite, HR and Employees, (see Table 1).

**Table 1.** Number of questions by survey/persona in 2024

Persona/Survey	Nr respondents	Nr of demographic questions	Nr of single/multi choice questions
C-Suite	845	12	33
Employee	9 449	24	33
HR	845	11	38

## 1.2 The manual aspects of GTT

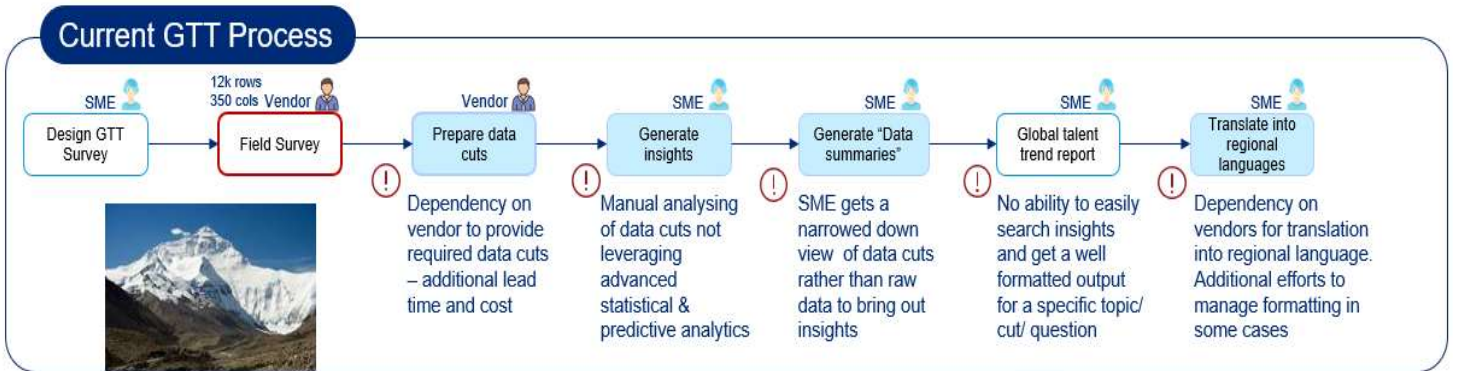
Fig. 1 depicts the typical manual process for creating GTT reports. Initially, the analysis process was Vendor dependent where Vendor conducted the survey and provided preliminary analyses. Specifically, a separate excel file was created for each of the three personas. Each file contained the percentage of respondents that answered each answer to each question globally. Further, the global analysis was then broken into “data-cuts”, i.e. separating between high versus low growth companies, agile firms versus not, different continents, countries, and industries, thriving employees versus not, and demographics such as age and gender to name a few.

After this data is created using Microsoft Excel, Mercer subject matter expert (SME) engaged in back-and-forth discussions, manually identifying the most surprising insights (typically top/bottom 5 responses to each question and top 5 largest absolute differences between a data cut and the global average), leading to the creation of data summaries. This is followed by a distillation of the 4-5 themes for the year and their respective sub-themes. Slides and report are then written up according to these themes and supplemental insights for each country and industry where completed (all manually by hand eyeballing differences to the global mean) (white box in Fig 1). Finally, the report is translated by regional SMEs into their respective languages and data cuts and data summaries may be re-analysed locally to highlight further topics of interest and spot regional differences in relation to global trends at the sub-trend level. Again, this is done as a largely manual process.

## 1.3 Automating the process of GTT with AI

Building on the success of using Generative AI, specifically ChatGPT to inform the creation of [3], whereby we used AI to validate that the ideas we intended the book to cover were recognizable by the AI and therefore came across clearly to readers. Our

intention for GTT was to use AI to automate manual steps, facilitate the extraction of a narrative from data summaries and eventually inform the write-up of the actual report as well as meaningful industry/country insights. Then a UI from which regional leaders and SMEs can explore the data further. Specifically, we identified that data architecture, engineering, data science, prompt engineering, as well as off-the-shelf translation software needed to automate the manual tasks in the blue boxes of Fig.1. Our goal was also to do this in a reusable pattern so that the same data pipeline can handle future surveys, so that SME can focus on the tip of the iceberg of insight creation, building upon the automation of the mind-numbing manual steps, and even cross compare data sets.



**Fig. 1.** The manual process from survey to analysis for creating a Global Talent Trends report.

## 2 Project implementation

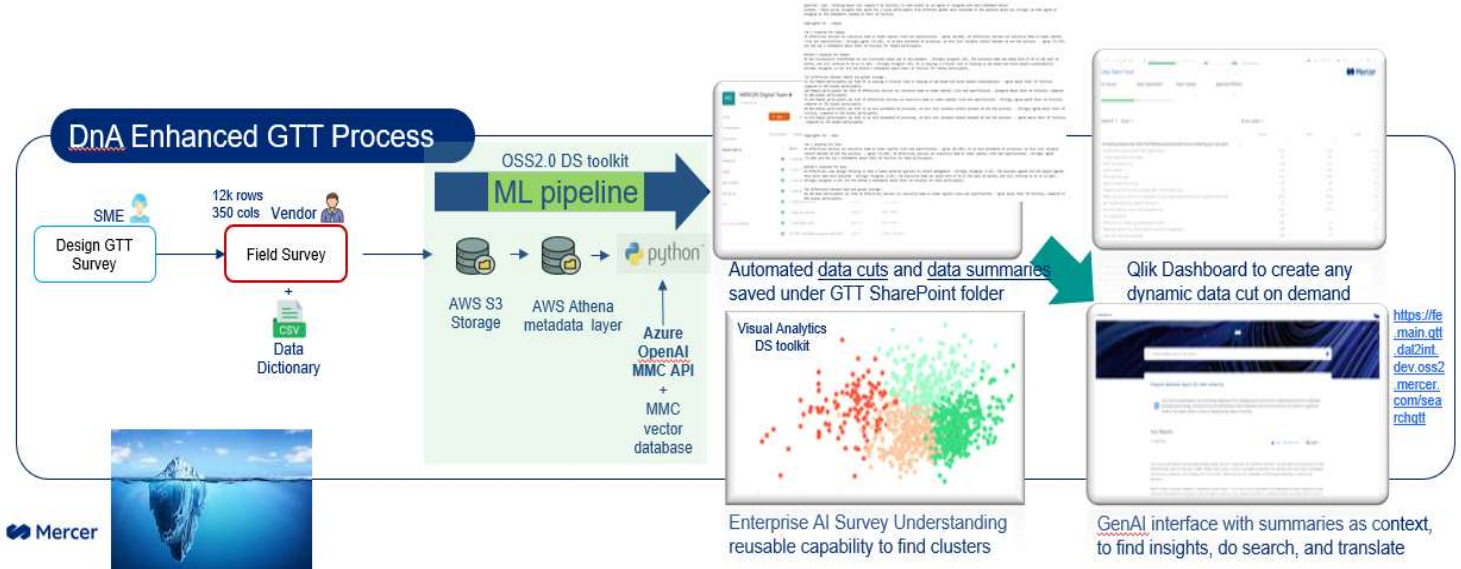
Fig. 2. shows a representation of the revamped process.

### 2.1 Structuring the input data to facilitate unsupervised learning and automation of data cuts

The first step was to request the vendor who runs the surveys to send us three structured data files, one per survey, where each row is a respondent and each column is the answer they gave to each of the single or multi-choice questions as well as each demographic question that might be used for subsequent cuts. We loaded these files into AWS S3.

We then worked with our data architect to come up with a desirable structure for the data that would facilitate our downstream analyses needs as well as future reusability of the pipeline for future surveys. They also harmonized the naming convention of the demographic data. Our data engineer then implemented the changes by pivoting each of the survey files so that each row is now a person question-answer pair with the following columns: <survey, year, date, question, answer, personID, demographic 1...k>, and saved the files partitioned by persona. If this were a database, it would look like the three surveys were unioned vertically into one.

Because all the answers arrived from vendor as categorical string data but most of the data was actually ordinal, in order to facilitate unsupervised learning storytelling (which we will explain in section 2.3), we added a data dictionary that, where possible, translated all ordinal data into a numeric representation. For every question, to facilitate interpretation, we ensured that positive answers had high values and negative answers had low values. A section of the data dictionary can be seen in Table 2. The data dictionary was also added to S3 and joined with the survey data.



**Fig. 2.** The improved process from survey to analysis for creating the GTT report.

**Table 2.** Example of ordinal response and its translation in the dictionary

Ordinal multi-choice	Numerical representation
Strong disagree 1	1
2	2
3	3
4	4
Strongly agree 5	5

We then created metadata layers for the two structures in AWS Athena, so that we can run SQL language on top of the unioned surveys joined with the data dictionary. With this, automating the first blue box of Fig.1 was a matter of creating a Python script that iterates over each data-cut to export CSV files out for manual consumption. The goal here is to enable SMEs to continue analysing information the old way if they wish to manually validate the results of this new approach.

Furthermore, we created a BI dashboard connected directly to the unioned survey data, so that as users navigate the respective filters, they are effectively consuming the

data that used to be spread over hundreds of CSV files. The dashboard contained more informative visuals than the simple table-like CSV files, thus being more informative than the dull and numerous excel sheets. Now SMEs can slice and dice the data as required, thus replacing the need for multiple excel sheets that the vendor used to send.

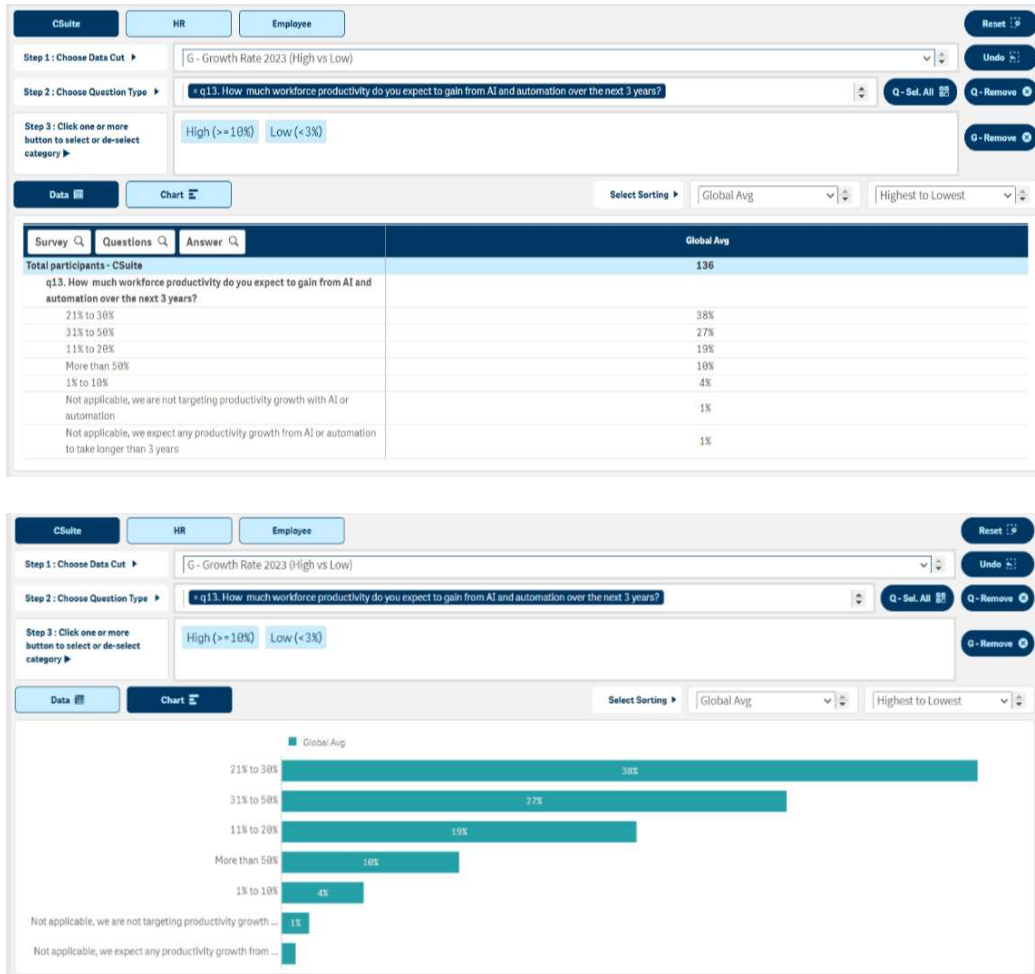


Fig. 3. SMEs can now slice and dice the data behind GTT.

## 2.2 Enabling Q&A with Generative AI on top of structure data

Now we needed to focus on how to get all this data into GPT.

To begin with, we chose not to use OpenAI's models directly. Our parent company has used Microsoft hosted OpenAI models, specifically ADA and ChatGPT, and made them available through internal APIs, which we use in this work [4]. This way, we are ensured that the data we enquire GPT with stays internal.

**Teaching GPT to get context from structured data.** The question of how to get GTT to understand CSV data is not straight forward. During extensive experimentation, we discovered that GPT / LLM work better with text than CSV files or tabular data. In fact, our data scientists observed that:

- using the CSV agent of langchain, we got ChatGPT to answer “pandas” like questions like ‘how many rows exist in the data?’. This may be very useful but uninteresting in terms of insight.
- taking each of the CSV files as textual context to ChatGPT and asking questions such as ‘give me the most interesting findings on this CSV’?, also gave us flaky results. Not surprising, given we know GPT3.5 is not a mathematician.

We found that the most reliable way to get insightful answers downstream was to automate the read-out in English of the numbers we expect that human SMEs need to typically pay more attention to. We automated the creation of these read-outs, added them to a vector database, and used this as context for the LLM.

Human SMEs typically comment on the top five answers to each question as part of global trend detection, then on the top 5 most different data cuts when compared to the global trend. Therefore, we used Python to iterate over each question and data cut and send SQL to Athena to obtain the respective noteworthy data, then used python to paraphrase it into simple Text. We did this for all questions, all surveys, all data cuts. And created a data summary txt file for each of the survey/persona, question, and data cut. Table 2 shows us the number of resulting summary files.

---

*Which of the following statements are true about artificial intelligence (AI) and automation at your organization? 53% of Employees from Australia say, in GTT 2024 survey, that I trust that my organization will teach me the skills I will need if my job changes as a result of AI and/or automation compared to the global average of 45% Employee Global participants.*

---

**Table 2.** Number of data summaries by survey/persona in 2024

Persona/Survey	Nr of single/multi choice questions (A)	Nr of data-cuts (B)	Nr of summary files (A*B)
C-Suite	33	29	957
Employee	33	49	1 617
HR	38	32	1 216
Total			3 790

Obviously, this approach generates a large volume of textual information that cannot be consumed within a single prompt of ChatGPT. Therefore, we fed all the text files into a vector database (we used Qdrant but in future will use Mongo Atlas DB),

using ADA embeddings model. We then integrated the vector database with the LLM using langchain's textloader agent. The vector database enables filtering from the database all the content that best matches the user query and use it as context for the LLM to compile its answer. We finally created a UI for users to interact with the data, which we show in Fig.5. For every question entered in the UI, the vector database will filter to the content that best matches the user query and feed this to the LLM as the context for the LLM needs to answer the user prompt. The UI includes some suggested prompts, e.g. "what are the most interesting findings for a data cut and persona". The UI also enables filtering by survey, when the user wants to ensure the answer refers to just that persona. This filter is passed to the vector database.

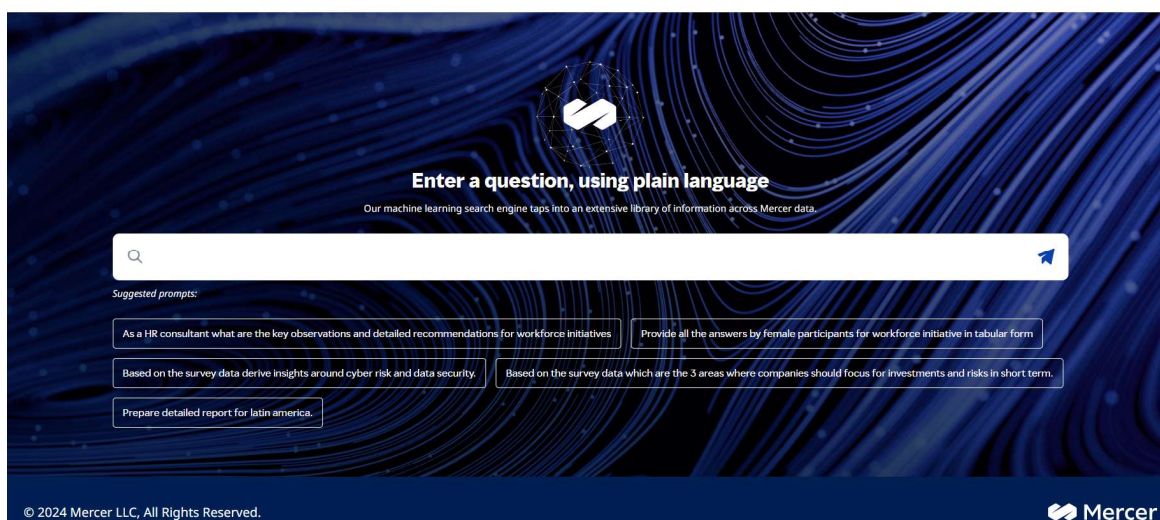
---

*Top differences between Mexico and global average:  
If you had unlimited budget, which areas of investment would make the biggest difference to your workers' day-to-day experience in 2024? 22% HR Mexico participants say, in GTT 2024 survey, that, Work design, processes, technologies: people feel efficient and have equitable digital access compared to global average of 31% HR Global participants.*

---

**Fig. 4.** Extract of textual summary created from HR survey, Mexico data cut. This extract is part of a txt file called "Employee 2024 – Focus Geos – q5 – Investment areas for workers' day to day experience – Summary.txt", with one text file per survey/persona, data cut, question.

#### Global Talent Trends 2024 (BETA)



**Fig. 5.** SMEs and regional leaders can interact with the read-outs of the data behind GTT.

**Prompt engineering. Teaching GPT to answer questions about survey data.**

Apart from the question the user provides in the UI in Fig.5, we had to carefully craft a prompt that teaches the LLM to make sense of the data, especially because the data is multi-layered, so the LLM has to be taught the meaning of data cuts and personas and how to craft a response that would roughly approximate how an SME would look at the data for each user question. Our approach involved asking our SMEs for an example of a well-written data cut / global analysis and to describe the expectation they had in detail. We iterated multiple times with the business SMEs, until we got a prompt that largely produced insights that were factually correct and made useful sense. Our experience taught us that involving SMEs in prompt design is critical to obtaining business relevant outcomes. Also, we cannot expect end users to make prompts that get consistent meaningful results in one iteration, hence the importance of adding backend prompt to the question the user enters into the UI.

Here is the back-end prompt that worked best, after several trials. Note that we concatenate this prompt with whatever the user chooses in the UI. We sent it as one prompt, but here we break it here into portions to clarify the logic underlying each portion of the prompt.

*Explain the context:* “We created a survey that we sent to 3 personas globally: C-Suite, HR teams and Employees. The document I am about to feed you is the readout of aggregating those responses. You will get data on global trends and then main differences between specific groups of people against a global trend. Groups, or as we also call them data cuts, can be HR high versus low growth companies, continents, countries, agile firms versus not agile, industries, thriving employees versus not thriving, gender”.

*Explain the persona:* “Consider yourself a data analyst with a keen interest in understanding the HR trends coming out of this global survey and a keen desire to answer the end-user question that you were given”.

*Teach what to answer:* “If the question contains one or more groups or personas, cover just those groups or personas. If the end-user did not mention a persona, assume he cares for all 3 personas equally and give similar importance to each of those personas in your answer. If asked to compare different groups, please give similar importance to each of those groups in your answer”.

*Teach it how to answer:* “Please do this step by step. Split the document into broad text sections based on broader topics that can describe global talent trends. Examples of global trends can be productivity, work redesign, agility, talent assessment, skills, artificial intelligence (AI), ESG / sustainability, trust, Rewards/Compensation, pay equity, diversity and inclusion (DEI), Thriving, Health and well-being, flexible working. Please write one paragraph per section, starting with the name of the section. Each paragraph should contain relevant information from the text to support the trend. Make sure you give non-aggregated supporting numbers, for example show the specific questions with percentages and groups that most deserve comment. Please write in formal but fluid style, in expository writing style.

*Addition that differentiates the third from the fourth blue box of Fig.1.:* “Please remind the reader of the four top talent trends I care most about: ‘Drive human centric productivity’, ‘anchor to trust and equity’, ‘boost the corporate immune system’, and ‘cultivate digital first culture’. Attempt to weave this into the text elegantly”.

*Avoid Hallucinations and tell it you care:* “This is very important for my career. Include all relevant pieces of information, but do not make up false or misleading information. Please be aware of the questions our end-user wants answered.

*User question:* “ ”, with whatever the user puts in the UI.

### 2.3 Unsupervised learning to tell stories from data

GenAI is cool but predictive AI can also add value to highlight the main characteristics of all the respondents at once, in a way that is not possible when we analyse each question one by one.

Separately for each survey, we used the numerical version of the question-answer pairs, per table 2, and applied Principal Components Analysis (PCA), which is a factorial analysis that allows identifying the main axes of divergence of the respondents and getting rid of random noise. Onto this representation of the data, we applied k-means clustering, enabling us to get respondent personas, as in groups of people with similar answer in bulks of questions at once. For a description of these personas, please read [2].

PCA analysis brought to light the greatest disparities in the data in a multi-variate way:

- For Employees, the single most contributing question to the overall differences in the data was whether employees feel that they are thriving versus not, which closely correlates to how they answer questions about *trust* in their organization: to do the right thing for them, to design work experiences that bring out the best in them, to keep their skills up to date, to empower them to make decisions independently with minimum oversight, etc.

- For HR functions, the biggest difference was between those companies that are data-savvy versus not; less data-savvy companies tending to report lower target growth rate.

- For C-Suite, the biggest difference found was whether organizations place the right amount of trust in employees and balance empathy and economics, where favourable responses tend to report higher target growth rate.

These patterns have influenced the final report [2].

## 3 Results

Through our use of Generative AI, we effectively automated the third and fourth blue boxes of Fig.1. As an example, please find below the answer to an example user ques-

tion as put into the UI in Fig.5: “What are the most interesting findings for HR participants in Mexico?”

*“Global Talent Trends: HR Participants in Mexico*

*In the GTT 2024 survey, HR participants in Mexico provided valuable insights into global talent trends. One interesting finding is that 20% of HR participants in Mexico stated that better pension/retirement options are a reason why people leave their organizations, which is slightly lower than the global average of 23% among HR participants worldwide. This indicates that retirement benefits are a significant concern for employees in Mexico, but it is not the primary reason for leaving compared to the global average.*

*When it comes to priorities for the 2024 People agenda, HR participants in Mexico have a different focus compared to their global counterparts. Only 15% of HR participants in Mexico mentioned improving people managers' skills as a priority, while the global average stands at 38%. However, 49% of HR participants in Mexico emphasized enhancing the employee experience/EVP to attract and retain top talent, which is slightly higher than the global average of 55%. This suggests that HR professionals in Mexico recognize the importance of creating a positive employee experience to retain top talent.*

*In terms of data-driven insights, HR participants in Mexico are more focused on certain areas compared to the global average. For example, 57% of HR participants in Mexico prioritize understanding the prevalent health risks/issues in their workforce, which is higher than the global average of 50%. Additionally, 38% of HR participants in Mexico prioritize identifying groups with unmet needs related to pay, benefits, or retirement, compared to the global average of 48%. These findings indicate that HR professionals in Mexico are particularly concerned with employee well-being and ensuring equitable compensation.*

*Culture and Experience emerged as a significant focus for HR participants in Mexico. While 20% of HR participants in Mexico mentioned rolling out new HR technology or optimizing existing platforms as a priority, the global average stands at 27%. Moreover, 24% of HR participants in Mexico mentioned redesigning the total rewards package for different groups, compared to the global average of 34%. This suggests that HR professionals in Mexico are placing emphasis on creating a tailored employee experience and leveraging technology to enhance HR operations.*

*When it comes to reasons why people stay at organizations, HR participants in Mexico highlighted different factors compared to the global average. For instance, 18% of HR participants in Mexico mentioned "Our people" as a reason for employee retention, while the global average is only 10%. Additionally, 6% of HR participants in Mexico emphasized generous vacation/time-off policies, compared to the global average of 10%. These findings indicate that HR professionals in Mexico recognize the importance of a positive work culture and work-life balance in retaining employees.*

*In terms of investments contributing to improved workforce productivity, HR participants in Mexico have different priorities compared to the global average. While 22% of HR participants in Mexico mentioned employee training and up/reskilling as*

*a significant contributor to increased productivity, the global average is 33%. Similarly, 18% of HR participants in Mexico highlighted simplified organizational structure as a factor, while the global average is 29%. This suggests that HR professionals in Mexico prioritize investing in employee development and streamlining organizational processes to drive productivity.*

*Overall, the insights from HR participants in Mexico highlight their unique perspective on global talent trends. They prioritize enhancing the employee experience, understanding health risks/issues, and addressing unmet needs related to pay and benefits. Additionally, factors such as positive work culture and tailored rewards packages play a significant role in employee retention. These findings provide valuable insights for HR professionals in Mexico and can guide their strategies to attract, retain, and support future-fit talent.”*

Several end-user question types were passed to the LLM and consistently we obtained an analysis that highlights specific questions displaying significant differences to global trends organized by themes and summarized in a human-friendly way.

## **4 Innovative aspects of the project**

Data storytelling is not only a toolset but an art [5]. We propose and implement an approach that enables effective data storytelling using an LLM that was not created with structured data in mind, and yet we input CSVs and obtain relevant commentary.

Passing a CSV through OpenAI’s Code Interpreter Plugin and asking it to “generate graphs and formulate hypotheses and use visualisation to underline the findings, in no time offers insightful outputs” [6], it is said. Code Interpreter is however considered unsafe to use in our corporate environment until further notice. Multi-modal LLMs may one day enable us to consult data in various formats, but we are not quite there yet.

Until then, we have created a reusable data pipeline that breaks the large, structured survey data into consumable textual read-outs that get serviced up to the text-only LLM for end user prompting and provide a templated output. We have applied it to a realistic usecase and it is getting used on a trial basis by our SMSs. Note that this trial has run alongside the actual manual GTT write-up this year. All GTT of 2024 was humanly created.

The solution is designed in such a way that it can be reused in new surveys with their own data dictionary. The end-to-end pipeline is automated and creates data cuts, data summaries, dashboard, paraphrasing, and auto refreshed the UI points to the new folder location. Therefore, it has potential to be reused across multiple offerings, specifically several future surveys beyond GTT, and augment human intelligence in reaching insight from the survey data.

As a result, in the short run, we increase the efficiency in how we develop GTT insights by industry and issue, taking pressure of our team. We will also be able to respond to requests from colleagues / journalists in real time.

In the longer term, as we improve our use of this technology, this will enable us to bring our data sets together across our consultants and will fuel our industry and issue-led selling. This is the future of thought leadership - increasing the speed between advisory and insight and ensuring that our insights drive our advisory.

## 5 Future work

Future versions of this work will include further innovations. In fact, we want to perfect templates of which report types get written for every type of end-user question, so we can ensure the required summaries are all present in the vector database and the LLM has a detailed prompt for how to combine them into a report. We then want to use LAM (large action models) to choose the version of report-creator agent that best suits the users' needs [7]. We also want to experiment with using ontologies to inform the LLM of the main entities involved – personas, data cuts, question topics – to make it easier for the LLM to interpret user questions without having to describe ontology terms vaguely as part of the prompt.

Finally, as smaller tweaks, our summaries today are done question by question for global and then each data cut. In future, we intend to supply the vector database with data summaries by global or data cut, where questions are sorted within by more significant to less (in terms of their position in the global trend or the magnitude of the absolute difference between the data cut and the global trend). That way, we expect to make it easier for the LLM to comment on mathematically higher differences first. We also intend to include the topic of each question in the summary itself, so the spotting of the final trend and the writing of each section becomes cleaner. Finally, we will use a higher threshold for summaries to even get created, e.g. those which difference to the global average are statistically significant.

## References

1. Mercer.com Homepage, <https://www.mercer.com/insights/people-strategy/future-of-work/global-talent-trends/>, last accessed 2024/03/20.
2. Mercer.com Homepage, <https://www.mercer.com/assets/global/en/shared-assets/local/attachments/pdf-2024-global-talent-trends-report-en.pdf>, last accessed 2024/03/20.
3. Kate Bravery, Ilya Bonic, Kai Anderson: *Work Different: 10 Truths for Winning in the People Age*, 1<sup>st</sup> edn, Wiley, UK (2023.)
4. Mercer.com Homepage, <https://www.marshmcclennan.com/news-events/2023/november/marsh-mcclennan-develops-new-generative-ai-tool.html>, last accessed 2024/03/20.
5. Harvard Business School Online, <https://online.hbs.edu/blog/post/data-storytelling>, last accessed on 2024/04/22.
6. Mahak Jhamb, A few days ago. LinkedIn post, June 2023.
7. OpenAI.com HomePage, <https://community.openai.com/t/create-a-large-action-model-lam/719353>, last accessed on 2024/04/19.