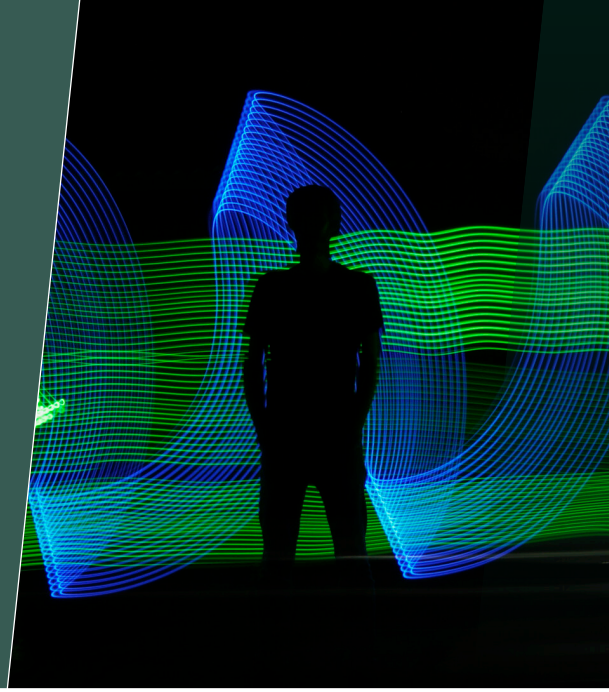
KORN FERRY

BE **MORE** THAN

TESTING AI IN ASSESSMENTS
CAN AI REPLACE CONSULTANTS? SIGNS POINT TO NO.



Thought Leadership

Introduction

Many leaders see the potential of artificial intelligence (AI) tools like ChatGPT and Microsoft Copilot can increase productivity and efficiency. Some even believe that up to half of jobs today could be fully automated or substantially changed due to this evolution of technology. Experts are working to understand AI's role in the future of work, its potential for disruption, and where the hype may be overblown.

Psychometric assessments have become an interesting testing ground for generative AI, challenging the assumption that testing humans requires a human. These assessments play a key role in talent management—from recruitment to development and promotion. Assessment reports are central to the larger process, turning data into actionable insights that support the growth of individuals and teams within organizations. However, the growing availability and capabilities of large language models (LLMs) have raised new curiosities and concerns about the extent to which LLMs might take over the role of the consultant.

In a traditional engagement, consultants use assessment results as one point of input for making recommendations to a client. Critical thinking and adaptability enable them to analyze complex data and adjust their approach as needed. They provide contextual insights, considering cultural and organizational factors to make recommendations more relevant. A consultant can enhance the assessment feedback process by delivering feedback in a supportive, constructive, and motivating way. It remains to be seen whether LLMs like ChatGPT can replicate these skills, or whether the role of a consultant in these processes is indeed uniquely human.

The Korn Ferry Institute partnered with students in New York University's Human Capital Analytics and Technology program to investigate how readily available GPT models can interpret and summarize group-level assessment results. We wanted to examine the relative capabilities of LLMs such as ChatGPT to automate or assist with producing reports related to Korn Ferry assessments. For the study, we prompted the GPT 4.o model to assume the role of an expert Korn Ferry consultant and provided sample group reports to serve as a template for a typical report structure. We used a mock data set of assessment results for 10 competencies and 7 traits across managers, directors, and VPs in addition to some general organizational information. Using this input, the students were tasked to define an efficient process using LLMs to produce a group level summary report and organizational insights. Then, we tested GenAI's ability to reproduce current approaches and invent new and different reporting solutions.

The project explored three separate stages of the group assessment report production process: graphics generation, company-specific insights, and development of interpretation material. Below, we share the strengths and weaknesses of the LLM for each task.

Our Key Findings

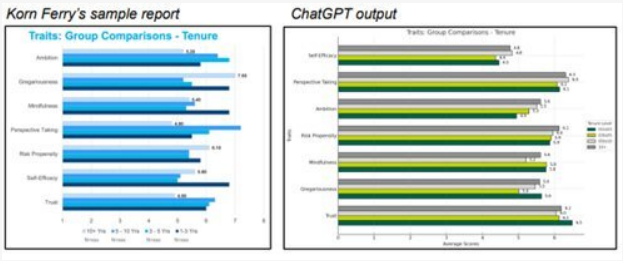
We conducted this investigation in three phases: the first focused on graphics generation capability, the second tested the production of company-specific insights, and the third explored the development of integrated user interpretation content. During each phase, we provided ChatGPT—our LLM of choice—with data and samples, guiding it through an iterative prompting process to recreate Korn Ferry's typical group assessment reports. The outputs from each stage were evaluated based on clarity, accuracy, and efficiency.

Stage #1: Generating Graphics

ChatGPT demonstrates highly variable performance in producing meaningful data visualizations, limiting their overall utility at this stage.

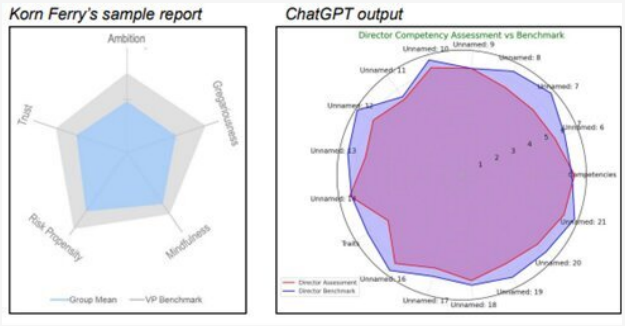
ChatGPT demonstrated highly mixed performance in reproducing and innovating with graphics generation. With proper guidance, it created decent replications of existing visualizations, such as line graphs and bar charts, occasionally reconfiguring the data to offer alternative visualizations that improved readability and clarity. In these cases, using ChatGPT can be helpful for generating ideas for different scenarios. Figure 1 shows how ChatGPT compares in producing some of the more basic visualizations.

Figure 1.
A comparison of a bar chart produced by assessment consultants in Korn Ferry's sample group report and one based on the same data produced by ChatGPT.



In other cases, ChatGPT seemed entirely incapable of creating other, more intricate graphic types, such as scatter plots and spider charts (see Figure 2). In these situations, we determined that the complexity of both the data input and the requested output did not align with the current capabilities of the technology, and the human generated visualizations were more effective.

Figure 2.
A comparison of a traits spider chart from Korn Ferry's consultants that the students used as a model for ChatGPT to create a similar chart for competencies. The chart generated by ChatGPT was highly illegible and inaccurate.



In all cases, we found that ChatGPT struggled with interpreting certain formatting conventions that humans tend to understand without issue, such as merging cells for legibility. Basic conventions can confuse the LLM as it tries to interpret the data set, and thus, produce inaccurate results. This suggests that using ChatGPT or other LLMs requires careful configuration of a data set, which must be taken into consideration ahead of any prompt.

Stage 1: Graphics Generation

The goal was to replicate Korn Ferry's standard group report graphics, such as spider charts, bar charts, heatmaps, and line charts. We checked the data set for accuracy and completeness prior to introducing it to ChatGPT. We gave prompts to ChatGPT to use the provided data set to generate a series of data visualizations. Early in the process of prompting ChatGPT at this stage, it became clear that certain formatting conventions "confused" ChatGPT and led to errors. As a result, we had to reformat the data for GPT interpretation. The time ChatGPT took to produce these graphics varied from 30 minutes for a basic bar chart to over an hour for a more complex heatmap. This includes time spent on revisions and necessary adjustments to ensure high-quality visualizations. We evaluated the graphics on criteria such as data accuracy, data integrity, clarity, and readability.

Stage #2: Creating Interpretation Guides

As a tool in the consultant's toolbox, ChatGPT may enhance the efficiency of the data interpretation processes.

Based on Stage 1 data and analysis, ChatGPT quickly identified several areas of strong leadership competencies and key development needs within the participant sample. This analysis provided a nuanced view of business unit leadership and valuable insights into the group's strengths and development areas relative to standard benchmark scores.

Despite ChatGPT's effectiveness demonstrated at this stage, the need to ensure the relevance and accuracy of the data interpretations remained a priority. Some AI-generated reports lacked nuanced language and tended to be choppy, fragmented, and disorganized compared to standard human-produced reports. As a result, the outputs felt more mechanical and less engaging. What's more, ChatGPT occasionally overlooked subtle contextual nuances that human analysts could capture, especially in sections requiring a deep understanding of organizational culture or specific industry dynamics. Ultimately, the additional supervision and iteration necessary to ensure readability, clarity, and flow reduced the overall efficiency of the process.

Stage 2: Creating Interpretation Guides

The goal was to evaluate ChatGPT's capability to identify key strengths, areas for development, and actionable talent management insights for the group based on the analysis of their collective scores. Using the same Stage 1 data set and preparation steps, we adopted a systematic and iterative prompt approach. We based initial prompts on key objectives identified in the preliminary analysis to extract insights and generate the interpretation report. We used subsequent prompts to further refine the language and structure, improving clarity and precision of the interpretations and action steps. To ensure the quality of ChatGPT's interpretation output, we focused the project on accuracy, relevance, and readability. Accuracy was paramount, requiring verification of factual content to align with source data. Relevance guaranteed the information addressed critical data points and analytical objectives. Readability aimed for clear, logical, and coherent presentation.

Stage #3: Producing Company-Specific Insights

With the right guidance and structure, ChatGPT can be a valuable tool for gathering information about a specific company to contextualize assessment results.

We tasked ChatGPT with producing research-based insights on an organization to gather critical context for interpreting the data it generated. This material included a summary of recent challenges, leadership changes, strategic priorities, and financial information. We found this task to be the easiest and most beneficial use of ChatGPT. It adeptly scanned available online resources, gathering relevant information from financial reports, public news, and other sources. This made it a valuable tool for conducting company-specific research and generating initial drafts of executive summaries and recommendations, which could be particularly helpful when collaborating with new clients.

We also asked ChatGPT to pinpoint mission-critical traits and competencies based on the context it identified. ChatGPT's analysis met the criteria for data accuracy, relevance, and clarity; it provided precise and relevant information tailored to the company, addressed specific strategic challenges and goals, and presented in an easy-to-understand format. Still, producing this quality of output required considerable front-end structure, including context about the exercise, examples from other engagements, and highly specific prompt inputs.

Stage 3: Company Specific Insights

The final step involved testing and validation, incorporating expert feedback to ensure high-quality, consistent, and actionable outputs. This stage focused on ChatGPT's ability to create company-specific interpretation reports. We had ChatGPT focus on two tasks. The first was to conduct company-specific research on annual 10k reports, official websites, press releases, news articles, and interviews to identify organizational challenges, leadership changes, and company goals. The second was to identify mission-critical leadership traits and competencies specific to the organization's business challenges and goals.

Where Intention Meets Action

Overall, ChatGPT had mixed results, performing some tasks well while struggling with others. It also produced inconsistent outcomes across similar trials. What's more, the platform experienced unexpected "downtimes," delaying project progress by several hours at a time. Therefore, it is critical to assess when using LLMs is indeed beneficial in these processes. To this end, here are three recommendations to consider:

1. Be strategic about the role LLMs will play in your process. As we have seen, LLMs aren't perfect for every task. When integrating GenAI into your workflow, identify specific applications where it excels, such as:

- brainstorming new ideas
- generating basic graphs with simple data
- paraphrasing and restructuring text
- summarizing information from various sources
- drafting initial content as a starting point for further iteration

2. Develop a structured, iterative approach to optimize usage. Create prompt guides and break problems down into their parts to solve them step by step. Remember, even with a carefully developed approach, LLMs can still produce incorrect outputs. Establish a review process to evaluate the results, making sure your criteria address potential pitfalls, including accuracy, completeness, and ease of interpretation.

3. Consider the tradeoffs involved. Sometimes, using ChatGPT or other LLMs for increased efficiency may compromise the overall effectiveness of the output. Creating highly effective materials using LLMs may require a more iterative, structured approach, which could extend the time needed for both input and output. Always prioritize what is most important, and recognize that in some cases, an LLM may not be the best tool for the job.

Hold On to Your Consultants

The results of our investigation into ChatGPT's ability to produce group assessment reports were predictably mixed, emphasizing the need for ongoing collaboration between humans and technology. ChatGPT's capacity to draw from a diverse array of sources allows it to offer novel insights that can catalyze further exploration and idea generation. This initial creativity proves invaluable when consultants seek fresh perspectives or need to jump-start their thinking on a problem.

Our exploration revealed that while ChatGPT excelled at tasks such as gathering contextual information about a company, summarizing results, and producing basic visualizations, it struggled with more complex tasks. Specifically, its ability to produce intricate visualizations and provide nuanced interpretations in an engaging, human-like tone did not live up to traditional methods of production.

This project underscores Korn Ferry's commitment to staying at the forefront of evolving technologies. Although current technology offers unique advantages that we continue to integrate into our processes to maximize value for clients, there remains a clear role for consultants in supporting individuals on their assessment journey. No existing technology can replace the unique human factor that consultants bring to the table.

Authors

Amelia Haynes
Manager, Research and Partnership Development
Korn Ferry Institute

David Mayfield
Senior Principal
Korn Ferry

Contributors

Bryan Ackermann
Managing Partner, Assessment & Succession, Leadership & Professional Development
Head of AI Strategy & Transformation
Korn Ferry

Veronica Ge
Sr. Director, Product Management
Korn Ferry

Mary Starke
Senior Client Partner
Korn Ferry

Andrea Deege
Senior Director, IP Development
Assessment Science and Scoring
Korn Ferry Institute

About Korn Ferry

Korn Ferry is a global organizational consulting firm. We work with our clients to design optimal organizational structures, roles, and responsibilities. We help them hire the right people and advise them on how to reward and motivate their workforce while developing professionals as they navigate and advance their careers.