AI-Powered Coatings: The Future of Paint Testing with Large Language Models (LLMs)

Introduction

In the coatings industry, evaluating paint performance has traditionally relied on human observation and verbal feedback. Researchers and technicians often describe paint performance using qualitative terms such as "smooth application," "great coverage," or "loud sound." While these descriptions are rich in detail, they are inherently subjective and are difficult to quantitatively compare different paint formulations in the evaluation process. And as newer generations of professionals enter the coatings industry, they bring fresh perspectives and innovative approaches. But not all researchers possess the same degree of experience and specialization. This diversity in expertise can influence how research is interpreted and applied across different sectors of the coatings industry. The lack of standardization slows down development cycles and complicates data-driven decision making.

To address these challenges, Dow is exploring the use of Large Language Models (LLMs) to analyze and quantify unstructured data. LLM is a machine learning model designed to generate and predict human-like language. These models' transformer architecture is trained to predict the next best token based on contextual patterns, allowing them to process and generate text and other media in natural language. The integration of LLMs to quantify unstructured data bridges gaps in understanding and streamline collaboration across diverse teams. LLMs can standardize data formats, categorize feedback by topic, and even translate content across languages while preserving technical nuances. These capabilities enhance clarity and foster more effective knowledge sharing across the value chain.

An example of quantifying subjective, unstructured data is paint contractor feedback on the coating appearance and ease of application of tested paints. These insights are rich in value but difficult to standardize or analyze. For example, one of the case studies explored later in this article will demonstrate how LLMs are utilized to extract actionable insights from open-ended feedback from a blind paint trial.

Transforming Feedback into Actionable Insights

Dow is uniquely positioned to unlock new insights from unstructured data by combining their deep expertise in coatings with advanced capabilities in LLMs. Incorporating coatings-specific knowledge is called in-context learning, and it enables LLMs to recognize patterns, terminology, and trends specific to the industry. These models can map qualitative descriptions to specific performance attributes, for example, linking "great coverage" to Hiding or "smooth flow" to Viscosity. Once mapped, the data can be organized into structured formats that support quantitative analysis and visualization. This synergy of domain knowledge and LLMs allows for more accurate analysis of unstructured data.

Painter	Paint Feedback
Painter 1	"Covers really nice, goes on very smooth. Spreads really well, coverage is well. Like this one a lot"
Painter 2	"Easy, some dripping, needs touch up. Not good coverage on the wood."
Painter 3	"Few more passes to cover; Good coverage; Looks like will dry without marks; Hit the wall paint while doing trim and noted 'flashing.'"
Painter 4	"Feels good with brush, better than A. Runs and flows a bit - be careful - it puddles in the corner. One coat coverage is good but def thinner. Flowability better than A but not 100% perfect."

Table 1: This table shows an example of unstructured data used for LLM analysis. The subjective nature of textual feedback makes it hard to quantify the paint performance. Additionally, jargons such as "flashing" and "flowability" used in the feedback requires domain knowledge for LLM to accurately analyze the data.

Additionally, LLMs can enable relative performance comparisons across formulations, helping researchers identify which products excel in particular areas, even when differences are subtle. By applying sentiment analysis, LLMs can also detect nuanced trends in user feedback—such as recurring concerns or emerging preferences—that might otherwise go unnoticed. These insights can be visualized through plots or data tables, offering a clear, data-driven view of how formulations perform over time and across trials. This capability not only enhances consistency and objectivity in evaluations but also accelerates development cycles and strengthens confidence in product decisions.

Case Studies

The case studies introduced in this section highlight the versatility of LLMs and how it has been utilized at Dow to enhance the efficiency and reliability of paint application testing.

A crucial aspect of working with LLMs is prompt engineering, the practice of writing an effective set of instructions for LLMs to generate the most desired responses. Prompt engineering is used in all case studies covered in this article to allow customization for the model to focus on a specific task. Crafting task-specific prompts allows users to guide the model's behavior and improve performance. Additionally, in-context learning allows users to embed domain-specific knowledge into the prompts, enabling the models to interpret technical language more accurately and deliver more relevant insights.

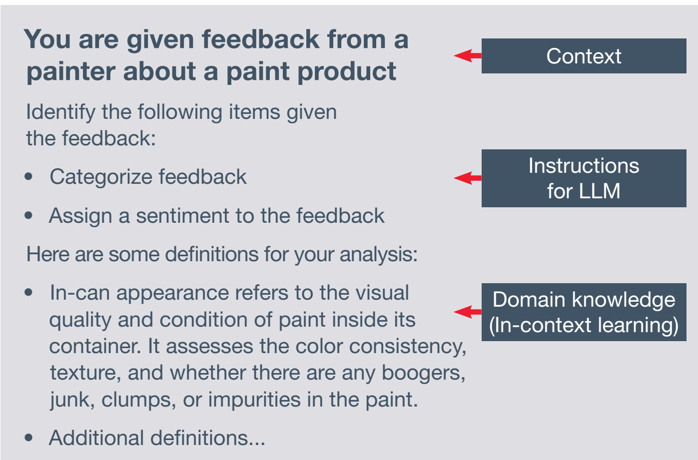

Figure 1. This figure shows an outline of a prompt used to analyze painter feedback. The prompt here uses in-context learning to further provide domain knowledge to the LLM to ensure accurate analysis.

Video-based Application Testing

The Additives team at Dow conducted a roller application study, where a painter applied a series of paints formulated with different rheology modifiers onto drywall while being filmed. During the application process, the painter provided real-time verbal feedback on key performance indicators such as paint viscosity, application noise, coating appearance, and overall ease of application. Historically, researchers manually reviewed each video to transcribe and interpret the painter's observations. These insights were entered into a spreadsheet for comparison across products to determine which paint showed the best application experience. This approach, while thorough, was time-consuming, labor-intensive, and required careful interpretation due to varying interpretations among researchers.

To address these challenges, the team leveraged LLMs to automate and standardize the evaluation process. The team filmed about 160 videos, all ranging from two to five minutes long. These videos were analyzed by an LLM using a customized prompt to eliminate the time needed to watch videos while reducing human subjectivity.

A key innovation in this approach was the inclusion of clearly defined performance categories within the prompt. This addition gives the model domain knowledge and provides customization so that it can better extract parts of the transcripts that meet the definition and properly categorize the transcript texts into the correct categories. This helps the model assign accurate ratings across the standardized categories. Once the model-generated ratings were compiled, averages were calculated across all categories to identify the top-performing paint. The results were validated by a subject matter expert (SME), whose manual assessment aligned with the model's conclusion.

This validation strengthened the researcher's conclusion that the best performing rheology package was the combination of ACRYSOL™ RM-725 or ACRYSOL™ RM-735BF and ACRYSOL™ RM-3030. This combination of rheology modifiers showed the best application performance across multiple formulations.

Video	In-can appearance	In-can appearance rating	Application noise	Application noise rating	Average rating
1	Drippy, needs to be shaken off before use	2	Noisy	2	2
2	Good consistency, some debris in the paint	4	It's quiet, very quiet	5	4.5
3	Not specified	3	Quiet	4	3.5
4	The paint has some lumps and impurities	2	The paint is quieter compared to the control	4	3

Table 2: This table shows an example of the results from the LLM analysis described above. Text specific to each category is extracted from the video and a rating is assigned by the LLM. The last column is the average of all ratings for all categories used to identify the best performing rheology package.

Blind Paint Performance Trial

Another innovative application of LLMs in paint testing at Dow involved the evaluation of test paints through a blind trial. In this study, four professional contractors were hired to assess the performance of experimental paints formulated with Dow binders against commercial benchmarks made with traditional binders. The contractors provided candid, qualitative feedback, which was recorded in writing. While this captured rich insights, the unstructured format made it difficult for researchers to quantify and compare performance across products in a consistent and scalable way.

To address these issues, the team leveraged an LLM to perform relative comparison analysis. The model was prompted to identify comparative language within the feedback, such as phrases like “better than,” “similar to,” or “not as smooth as,” and to interpret the directionality of each comparison and assign a similarity score. This allowed the model to understand how the test paints and commercial standard were performing relative to each other. The similarity scores also provided a structured, numerical basis for evaluating relative performance.

The prompt also included technical keywords that defined the qualities of a high-performing paint, giving the LLM domain knowledge needed to identify whether the test paints performed better, worse, or similarly to the commercial standard. This domain knowledge is a crucial part of the success of this process, because it gave the LLM a deeper understanding of the metrics for paint performance evaluation. Additionally, instead of manually reading and separating the information into a spreadsheet to evaluate, LLM quickly organized the information into a structured data frame that is used for visualizing the data.

Figure 2 shows the results of the analysis. It highlights how two test paints performed across various categories compared to the standard paint. Bars pointing right indicate better performance, while bars pointing left show worse. Based on expert judgment and empirical patterns, scores below 0.5 consistently indicated no meaningful deviation from the benchmark. All scores for both test paints fell within this range, suggesting performance comparable to the standard paint. Paint 1 showed slightly more positive deviations, but none exceeded the threshold, reinforcing that its performance was similar to the benchmark.

Benefits to R&D and Product Development

The integration of LLMs into Dow’s paint testing workflow marks a significant advancement in how qualitative data can be analyzed and transformed into structured and actionable insights. This allows Dow to reduce manual effort and minimize subjectivity, while accelerating decision making.

The use cases described here demonstrate how AI is a powerful and versatile tool that can enhance the traditional R&D process, making them more consistent and data driven. As this technology continues to evolve, there is strong potential to expand its application across other areas of product development and customer feedback analysis. Dow’s early adoption of LLMs in this space not only improves current workflows but also sets a foundation for smarter and faster innovation in the coatings industry.

References

Barth, A., Fregly, C., Eigenbrode, S., & Chambers, M. (n.d.). Generative AI with Large Language Models. Retrieved from DeepLearning.AI: https://www.deeplearning.ai/courses/generative-ai-with-llms/

Fulford, I., & Ng, A. (n.d.). ChatGPT Prompt Engineering for Developers. Retrieved from DeepLearning.AI: https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/

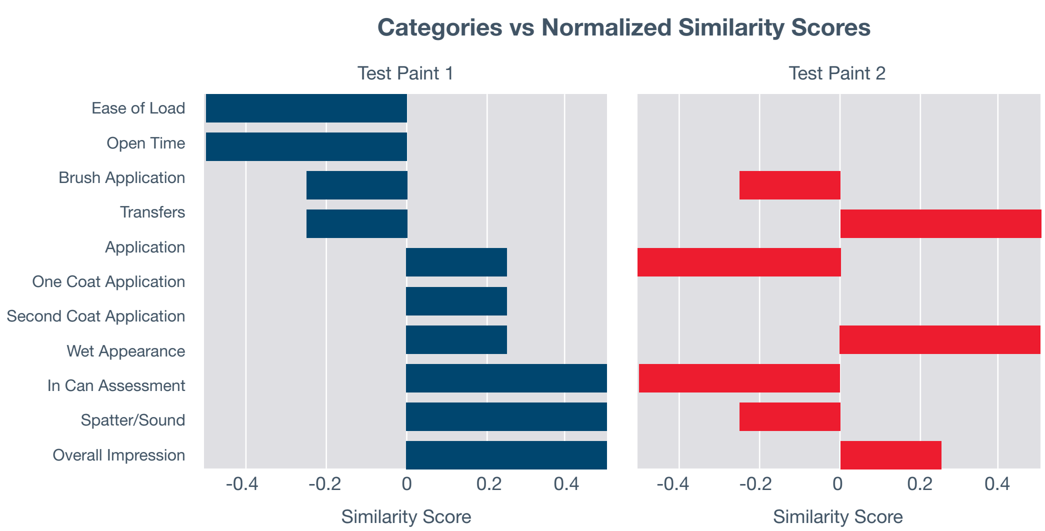

Figure 2. Comparison of test paints across performance categories relative to the standard paint

200002502-570206

Form No. 926-02460-01-1125 S2D