

Explore OpenAI for Hyper-specific Scent attribute exploration

OpenAI gpt3.5 and gpt4 was used to extract hyper-specific scent names from different CFIT departments

Both gpt3.5 and gpt4 is helpful in extracting hyper-specific scent from the product description. It can extract hyper-specific without explicitly mentioning. Below is the screenshot of the prompting used.

```
PROMPT += ". For example, give the response in a JSON format like this - {'\" + attribute + "\": '\"Apple Cinnamon Cheerios\\\", \"Confidence\": 100} in one line with key-value pair enclosed within double quotes, for the given input - '\"Apple Cinnamon Cheerios Cereal\\\". Enlist multiple \" + attribute + \" items as well if found separated with a \";'. Both the keys should be included in the response as shown in the example above. If you find no \" + attribute + \"s, mention the attribute value as OTHER and confidence as 100"
```

Percentage of "Non-Other" results obtained from gpt3.5 and 4 for a sample of 100 records is below

Gpt 3.5	41%
Gpt 4	45%
Rapidfuzz	44%

Gpt4 has done good job in extracting Scent names for few of the products which were marked as OTHER by gpt3.5. This will ensure in increasing more coverage of attributes. Screenshot below.

GTIN	Item Name	Scent_gpt3.5	Scent_gpt4
47400005242	Proglide Sensitive Sport Shave Gel Gillette Fusion ProGlide Sensitive 2 in 1 Shave Gel, Active Sport, 6oz GL1	OTHER	Active Sport
37000947820	Arctic Fresh Toothpaste CREST 3DW ARCTIC FRESH TP 3.5OZ CREST 3D WHT ARTC FRSH TP	OTHER	Arctic Fresh
37000976783	Whitening Gel Toothpaste Crest Pro-Health Whitening Gel Toothpaste (4.6oz) CREST P-HLTH WHT CLNMT	OTHER	Clean Mint
11111458727	Dove Men+Care Soap Bar Deep Clean 40 OZ DOVE MEN CARE BODY FACE D	OTHER	Deep Clean
70501113448	Hydro Boost Body Gel Cream Fragrance Free Neutrogena Hydro Boost Fragrance-Free Body Gel Cream NEOTHER	OTHER	Fragrance Free
640986035223	Rose All Day Moisturizing Lotion PRIMAL ELM ROSE ALL	OTHER	Rose
640986103663	PSL Bar Soap PRIMAL PSL SOAP BAR	OTHER	Pumpkin Spice Latte
603084244157	Nourishing Color Creme Garnier Nutrisse 43 Dark Golden Brown Cocoa Bean Nourishing Hair Color Creme	OTHER	Cocoa Bean

Gpt4 has the knowledge and learning to understand and extract attributes from abbreviations which is difficult for human eyes and other algorithms.

- For example: **PSL Bar Soap PRIMAL PSL SOAP BAR**

It is not possible for an algorithm or for human eyes to decode PSL as "PUMPKIN SPICE LATTE" but gpt4 could do it

640986103663	PSL Bar Soap PRIMAL PSL SOAP BAR	Pumpkin Spice Latte
--------------	----------------------------------	---------------------

Hyper-specific Scents extracted from gpt3.4 and gpt4 looks promising and doesn't need special prompting to extract hyper-specific attribute.

GTIN	Item Name	Scent_gpt3.5	Scent_gpt4
37000947714	Radiant Mint Toothpaste CREST 3DW RADIANT MNT TP 3.5OZ CREST 3D WHT RADNT MNT TP	Radiant Mint	Radiant Mint
37000944799	Pure Sport Plus Anti-Perspirant Deodorant Old Spice Men's Antiperspirant & Deodorant Sweat Defense P	Pure Sport Plus	Pure Sport Plus
17000027951	Pomegranate & Tangerine Hand Soap Dial Liquid Hand Soap Pomegranate & Tangerine 12/7.5 oz DIAL PN	Pomegranate & Tangerine	Pomegranate & Tangerine

- gpt4 takes more time to execute and provide results as compared to gpt3.5. The initial analysis reveals it is 5x slower.
- Adding department name into the prompt marginally increased the "Non-Other" count.

Conclusion:

From all the analysis and extractions made from both gpt3.5 and gpt4 it is observed that gpt3.5 runs much faster than gpt4 and both have capabilities to extract hyper-specific scent names without specifically mentioning it. Gpt4 is more capable and can extract marginally more attributes from product description. It also has the capability to understand abbreviations and make sense out of it. Examples are outlined above.