Explore OpenAl for Hyper-specific Scent attribute exploration

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

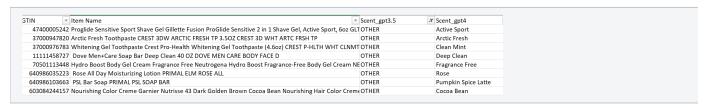
Both gpt3.5 and gpt4 is helpful in extracting hyper-specific scent from the product description. It can extract hyper-specific without explicitly mentioning. Below is the screenshot of the prompting used.

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 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.



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	▼ Scool	cent_gpt3.5	Scent_gpt4	~
370009	947714	Radiant Mint Toothpaste CREST 3DW RADIANT MNT TP 3.5OZ CREST 3D WHT RADNT MNT TP	Ra	adiant Mint	Radiant Mint	
370009	944799	Pure Sport Plus Anti-Perspirant Deodorant Old Spice Men's Antiperspirant & Deodorant Sweat Defense	PiPu	ure Sport Plus	Pure Sport Plus	
170000	027951	Pomegranate & Tangerine Hand Soap Dial Liquid Hand Soap Pomegranate & Tangerine 12/7.5 oz DIAL P	PN Po	omegranate & Tangerine	Pomegranate & Tangerin	1e

- 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.