

Supplementary Material for FoodSense: A Multisensory Food Dataset and Benchmark for Predicting Taste, Smell, Texture, and Sound from Images

Sabab Ishraq¹ Aarushi Aarushi² Juncai Jiang² Chen Chen³

¹College of Engineering and Computer Science, University of Central Florida, Orlando, FL, USA

²College of Business Administration, University of Central Florida, Orlando, FL, USA

³Institute of Artificial Intelligence, University of Central Florida, Orlando, FL, USA

A. FoodSense Annotation Protocol

A.1. Task Design

Participants were shown one food image at a time and asked to evaluate four sensory dimensions: **taste**, **smell**, **texture**, and **sound**. For each dimension, participants completed two sub-tasks sequentially.

Quantitative rating. The survey prompt read: “Based on the image above, how would you rate the likely [taste / smell / texture / sound] of this food?” Responses were recorded on the seven-point scale shown in Table A1. Participants could also select a *Can’t tell from picture* option (coded as 0) whenever the image provided insufficient visual cues.

Qualitative descriptor. After rating, participants were asked: “What do you think this food would sound like, taste like, smell like, and feel like (texture)? Please write *one or two words* for each sense.” Representative responses include *crispy*, *golden edges*, *smoky*, and *silent*.

This dual-format design captures both the magnitude of anticipated sensory experience and the natural language people use to ground those judgments in visual evidence. Valid numeric ratings were linearly rescaled from the original 1–7 range to a 1–5 scale via the transformation $r_k = 1 + [(r_{\text{orig}} - 1) \times 4] / 6$, preserving relative ordering. Responses marked as *Can’t tell from picture* were excluded from rescaling and retained as a separate binary CanInfer_k flag per dimension.

Table A1. Seven-point Likert scale used for sensory ratings.

Value	Meaning
0	Can’t tell from picture
1	Very bad
2	Bad
3	Slightly bad
4	Neutral
5	Slightly good
6	Good
7	Very good

A.2. Survey Interface

The annotation survey was administered via Qualtrics. Each survey page presented a single food image at the top, followed by four rating scale questions (one per sensory dimension) and four free-text response fields. Figure A1 shows a representative screenshot of the survey interface as presented to participants.

A.3. Participant Recruitment

Participants were recruited through two channels: (i) a professional online survey panel ($n = 7,734$; 63,741 annotations, 95.4%) and (ii) a university behavioral laboratory

In this study, you will be asked to **evaluate 5 food images**.

Please rate how you expect each one to taste, smell, feel, and sound when eaten.

You will also be asked to write a word or two about how you think it will taste, smell, feel, and sound when eaten.

There are no right or wrong answers.

Thank you for your willingness to take this survey.



Based on the image above, how would you rate the likely...

	Very bad 1	2	3	Neither good nor bad 4	5	6	Very good 7	Can't tell from picture
texture of the food	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
taste of the food	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
smell of the food	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sound of the food when eaten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When you see the image, what do you think it would sound like, taste like, smell like, and feel like (texture)?

Please write *one or two words* for each.

Taste

Texture

Sound

Smell

Figure A1. Qualtrics survey interface as presented to participants, showing the food image display, seven-point rating scales for each sensory dimension, and free-text descriptor entry fields.

($n = 648$; 3,101 annotations, 4.6%). In total, 8,382 participants contributed 66,842 assessments across 2,987 images (mean = 22.38 annotations per image, $SD = 2.02$). The dual-channel recruitment strategy strengthens annotation quality—laboratory participants provide controlled, distraction-free responses, while the large online panel ensures scale and demographic diversity. No demographic information was collected as part of the annotation protocol.

Following quality filtering—removing 72 images due to filename inconsistencies in the training pipeline and retaining only assessments where participants indicated visual infeasibility for all four dimensions ($\text{CanInfer}_k = 1$)—the final training dataset comprises 58,443 annotations across 2,915 images.

B. Additional FoodSense Dataset Statistics

B.1. Annotator Distribution

Table B2 summarizes the distribution of annotation counts across the full released dataset of 2,987 images. The distribution is approximately Gaussian, centered near 22 an-

notations per image (Figure B2), with a long right tail attributable to variability in the number of participants assigned to each image across survey sets.

Table B2. Summary statistics for annotation counts across 2,987 images.

Statistic	Value
Mean annotations per image	22.38
Standard deviation	2.02
Minimum	3
Maximum	42
Total annotations (full dataset)	66,842
Total annotations (training set)	58,443

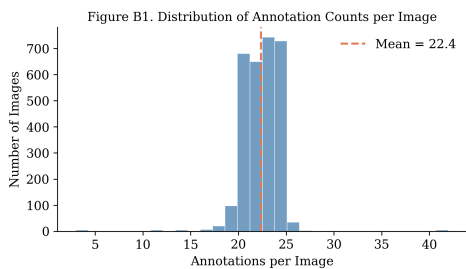


Figure B2. Distribution of annotation counts per image across the full dataset of 2,987 images. The dashed line indicates the mean (22.38 annotations per image).

B.2. Data Partitioning

To construct train/validation/test splits, we applied pseudo-random stratified shuffle splitting at the image level (75% / 10% / 15%) based on binned mean overall rating, ensuring all three partitions share equivalent sensory rating distributions. All splits are strictly image-level—no image appears in more than one partition. Table B3 summarizes the final partition sizes.

Table B3. Dataset partitions with image and annotation counts.

Split	Images	Annotations	% of Total
Train	2,185	43,758	74.9%
Validation	292	5,834	10.0%
Test	438	8,851	15.2%
Total	2,915	58,443	100%

B.3. Descriptor Vocabulary

Participants provided one to two free-text words per sensory dimension. The descriptor vocabulary spans 25,508

unique terms across 265,915 total entries (including repetitions across participants). Table B4 reports unique vocabulary sizes and representative high-frequency terms per dimension. Figure B3 shows the top-10 descriptor frequencies for each sense.

Table B4. Descriptor vocabulary statistics and top-10 terms by sensory dimension.

Dimension	Unique Terms	Top 10 Descriptors
Taste	9,217	sweet, savory, salty, good, spicy, delicious, bland, fishy, fresh, meaty
Smell	9,161	sweet, fishy, savory, fresh, good, spicy, meaty, cheesy, fried, delicious
Texture	6,807	soft, crunchy, chewy, smooth, crispy, mushy, slimy, creamy, rough, hard
Sound	5,323	crunchy, quiet, soft, crunch, none, squishy, chewy, slurp, wet, no sound

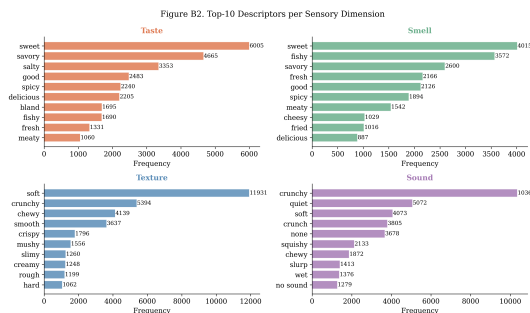


Figure B3. Top-10 descriptor frequencies per sensory dimension across all 2,987 images.

B.4. Cross-Sensory Correlations

Table B5 reports Pearson correlations between per-image mean ratings across all four sensory dimensions ($N = 2,987$ images, $\text{CanInfer}_k = 1$ only, all $p < .001$). Correlations are uniformly strong and positive, consistent with a shared visual appetitiveness signal driving ratings across dimensions. The weakest pairings involve sound ($r = 0.726$ – 0.840), reflecting the greater difficulty of auditory inference from static images—a pattern consistent with the lower ICC(1,k) for sound reported in the main paper.

Table B5. Pearson correlations between per-image mean sensory ratings ($N = 2,987$; all $p < .001$).

	Taste	Smell	Texture	Sound
Taste	—	0.893	0.931	0.759
Smell	0.893	—	0.856	0.726
Texture	0.931	0.856	—	0.840
Sound	0.759	0.726	0.840	—

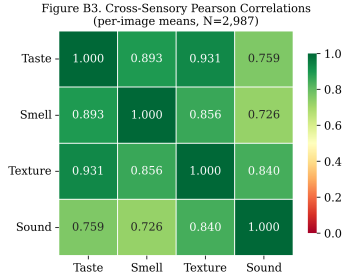


Figure B4. Heatmap of Pearson correlations between per-image mean sensory ratings ($N = 2,987$).

B.5. Descriptor Consistency and Uncertainty Structure

To assess annotation consistency in the free-text descriptors, we computed per-image lexical diversity and coverage statistics for each sensory dimension, excluding missing responses and uncertainty expressions (e.g., “not sure”, “can’t tell”). Table B6 reports the mean type-token ratio (TTR; unique descriptors divided by total valid descriptors per image), top-1 coverage (fraction of annotations using the most common descriptor), and top-3 coverage per sense.

Table B6. Per-image descriptor diversity and coverage statistics across 2,987 images. TTR = type-token ratio (higher = more diverse). Top- k coverage = fraction of annotations using one of the k most common descriptors for that image.

Dimension	Mean Unique	Mean TTR	Top-1 Coverage	Top-3 Coverage
Taste	16.04 (3.22)	0.725 (0.137)	0.218	0.396
Smell	16.75 (3.26)	0.760 (0.135)	0.200	0.367
Texture	14.60 (2.84)	0.658 (0.120)	0.245	0.457
Sound	14.43 (3.08)	0.663 (0.134)	0.232	0.446

TTR values ranging from 0.658 to 0.760 indicate substantial lexical diversity across all dimensions, consistent with the inherently subjective nature of sensory inference from images. Texture and sound show slightly lower TTR and higher top-3 coverage than taste and smell, suggesting modestly greater convergence on a smaller set of perceptual anchors (e.g., *soft*, *crunchy* for texture; *quiet*, *crunchy* for sound). However, top-3 coverage does not exceed 0.457 for any dimension, confirming that no small cluster of terms dominates annotations for a given image.

We further examined whether descriptor diversity and numeric rating disagreement capture the same underlying uncertainty signal. Table B7 reports Pearson correlations between per-image TTR and per-image rating SD across all four dimensions.

For taste and texture, descriptor diversity and rating variance covary positively and significantly, indicating that the two annotation modalities capture a shared uncertainty signal: images that elicit more varied numeric ratings

Table B7. Pearson correlations between per-image descriptor TTR and rating SD ($N = 2,987$). Higher TTR indicates more diverse descriptors; higher rating SD indicates more numeric disagreement.

Dimension	Pearson r	p -value
Taste	0.249	< .0001
Smell	-0.047	.011
Texture	0.141	< .0001
Sound	0.068	.0002

also attract more diverse textual descriptions. Sound and smell show weaker effects. For sound specifically, the CanInfer rate is also uncorrelated with rating SD ($r = 0.007$, $p = .71$), unlike the other three dimensions where higher inferability rates associate with lower rating variance (taste: $r = -0.160$; smell: $r = -0.137$; texture: $r = -0.112$; all $p < .0001$). Together, these patterns suggest that auditory inference from static images operates through a qualitatively different uncertainty mechanism than the other sensory dimensions, consistent with sound’s lower ICC(1,k) reported in the main paper.

C. Additional FoodSense Annotation Examples

This section presents two representative annotation examples drawn from the test set (sampled randomly, seed = 42). Each entry reports the per-image mean rescaled rating (1–5 scale) and the primary free-text descriptor for each sensory dimension.

Figure C1-C2. Annotation Examples (Random Sample)

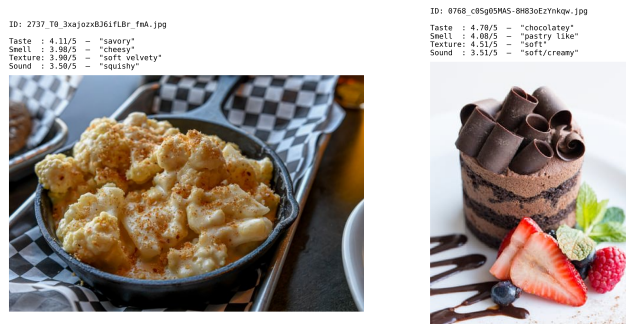


Figure C5. Two annotation examples randomly sampled from the test set (seed = 42). Left: Image 2737 (mac and cheese). Right: Image 0768 (chocolate dessert).

Example 1 — Image 2737. Taste: 4.11 (*savory*); Smell: 3.98 (*cheesy*); Texture: 3.90 (*soft velvety*); Sound: 3.50 (*squishy*).

Example 2 — Image 0768. Taste: 4.70 (*chocolatey*); Smell: 4.08 (*pastry like*); Texture: 4.51 (*soft*); Sound: 3.51

Key observations: (1) The single-stage ablation and base model have nearly identical distributions ($D = 0.036$, not significant), suggesting that without the staged curriculum, the model does not meaningfully reshape its output distribution. (2) FoodSense-VL’s distribution is significantly different from both the single-stage ($D = 0.148$) and the base ($D = 0.134$), confirming the curriculum effect. (3) Generalist VLMs (LLaVA, InternVL, Qwen) cluster together with small pairwise distances ($D \leq 0.133$) but diverge sharply from fine-tuned models ($D > 0.3$).

E. Full Sensory Inference Comparison

This supplementary material provides the full sensory ratings table and complete text inferences from all models across six representative food images. Sensory keywords are underlined. Human GT is shown on a 1–5 scale.

E.1. Model Prompts and Configuration

To ensure reproducibility, we provide the exact text prompts used for training and evaluating FoodSense-VL, as well as for all closed-source and open-source baseline inferences.

During the two-stage QLoRA fine-tuning, FoodSense-VL is conditioned on the following **System Prompt** to adopt the persona of a sensory analysis expert:

System Prompt

You are a sensory analysis expert trained to evaluate food images across multiple sensory dimensions. Your task is to analyze the food image and predict sensory experiences based on visual cues.

Rate each of these four sensory attributes on a 1.0–5.0 scale: Taste, Smell, Texture, and Sound (when eating). Each sense may have a different rating; do not default to identical ratings across all senses.

If reference reviews are provided, treat them as background about similar dishes; do not quote them verbatim and do not mention Yelp.

Respond in plain text only (no markdown).

For zero-shot baseline evaluation across generalist VLMs (InternVL, Qwen, LLaVA, Food-Llama, etc.), we utilize the **User Evaluation Prompt** to extract uniform Sensory Assessment formatting:

User Evaluation Prompt

Analyze this food image and estimate the likely sensory experience. Plain text only (no markdown; no bullets; no asterisks). Provide a Sensory Assessment with 4 sections, one each for Taste, Smell, Texture, and Sound, formatted as: Sense (X.X/5.0): 3–4 sentences of detailed visual justification.

E.2. Sandwich (0001)

Human GT: Taste=4.1, Smell=4.1, Texture=4.3, Sound=3.6



FoodSense-VL (Stage 2) — Taste:3.5 Smell:3.0 Texture:2.5 Sound:2.0

Taste: The sandwich appears savory and comforting, with meaty notes from the filling and richness from melted cheese. Toasted bread contributes a mild sweet-toasty balance.

Smell: Expected aroma is warm and inviting, driven by toasted bread, cooked egg, processed meat, and dairy notes.

Texture: The bite likely skews soft overall: yielding bread, melted cheese, and smooth egg with only limited crispness.

Sound: Predicted sound profile is quiet, with minimal crunch and mostly muted chew.

Base (Gemma 3 27B) — Taste:3.5 Smell:3.0 Texture:2.5 Sound:1.5

Taste: The sandwich appears to contain classic breakfast flavors with eggs, cheese, and processed meat, likely sausage. The toasted bread suggests slight caramelized sweetness that balances savory elements. Overall it likely tastes salty, umami-rich, and rich from cheese and egg.

Smell: The image suggests a comforting breakfast aroma. Expected notes include cooked egg, slightly smoky processed meat, and warm, wheaty toasted bread, with possible buttery hints.

Texture: Bread appears soft with some structure from toasting. Egg looks smooth, while sausage likely adds slight snap or chew. Melted cheese contributes a stretchy, yielding mouthfeel with limited contrast.

Sound: Eating this would be mostly quiet. The soft bread and egg produce little crunch; likely cues are slight squish and a subtle snap from processed meat.

InternVL-2.5 26B — Taste:4.5 Smell:4.0 Texture:4.0 Sound:3.5

Taste: The sandwich likely offers a rich, savory taste with the crispy bacon providing a salty, smoky flavor, complemented by the creamy, slightly tangy taste of melted cheese. The toasted bread adds a warm, slightly sweet and nutty flavor, enhancing the overall taste experience.

Smell: The aroma of the sandwich is likely to be enticing, with the scent of crispy bacon and melted cheese dominating. The toasted bread adds a subtle, comforting aroma that complements the other flavors, making it a delightful sensory experience.

Texture: The sandwich offers a delightful contrast in textures, with the crispy bacon providing a satisfying crunch, the melted cheese offering a smooth, gooey texture, and the toasted bread adding a firm, slightly chewy base. The combination creates a mouthfeel that is both satisfying and enjoyable.

Sound: While the visual appeal is strong, the sound of this sandwich is less pronounced. The crunch of the bacon and the slight sizzle of the cheese might be audible, but the overall sound is likely to be minimal, focusing more on the

visual and olfactory senses.

Food-Llama 11B — Taste:- Smell:- Texture:- Sound:-

Taste: The generated output describes a hearty savory profile from melted cheese, toasted bread, and bacon-like filling.

Smell: It suggests a warm cooked aroma associated with bacon, toasted bread, and rich dairy elements.

Texture: It emphasizes a comfort-food texture blend with crisp toast edges and creamy melted components.

Sound: The output references mild sizzling/crispness cues rather than loud crunch.

Qwen2.5-VL 32B — Taste:4.5 Smell:4.0 Texture:4.5 Sound:2.5

Taste: The sandwich appears to be filled with melted cheese, which suggests a creamy and slightly salty flavor. The presence of what looks like bacon or a similar ingredient indicates a savory, smoky taste that would complement the cheese well. The toasted bread adds a subtle sweetness and a hint of crunch, enhancing the overall richness of the sandwich. The combination of these elements likely results in a balanced, satisfying, and indulgent taste profile.

Smell: The image shows a freshly toasted sandwich, which implies a warm, inviting aroma of golden-brown bread. The melted cheese would contribute a rich, buttery scent, while the bacon or similar filling might add a savory, slightly smoky odor. The overall smell is likely to be comforting and appetizing, evoking a sense of warmth and homeliness.

Texture: The toasted bread has a crispy exterior and a soft interior, providing a pleasing contrast in texture. The melted cheese appears smooth and gooey, offering a creamy texture that would melt in the mouth. The bacon or similar filling seems to have a slightly chewy, crispy texture, adding another layer of variety. The combination of these textures creates a satisfying and diverse mouthfeel.

Sound: While the image itself does not provide auditory information, one can imagine the sound of biting into the sandwich. The crispy toast would likely produce a satisfying crunch, while the cheese might stretch slightly, creating a gentle tearing sound. The overall sound experience would be minimal but could include the soft squish of the cheese and the slight resistance from the filling, contributing to an engaging eating experience. However, without actual audio, this assessment is speculative.

GPT 5.4 Thinking — Taste:4.2 Smell:3.9 Texture:4.3 Sound:3.4

Taste: This looks strongly savory, rich, and slightly salty, with the toasted bread, melted cheese, and reddish meat filling suggesting a classic comfort-food profile. The glossy red sauce adds a likely sweet-tangy note that would brighten the heavier cheese and meat. Visually, the sandwich appears indulgent rather than subtle, with bold, familiar flavors layered closely together. Overall, it seems likely to taste warm, hearty, and satisfying, with mild sweetness balancing the fat and salt.

Smell: The image suggests a warm aroma led by toasted bread and melted cheese, which usually create a nutty, buttery scent when heated. The meat filling likely contributes a smoky, cured, or grilled smell that would make the sandwich feel more substantial. The red sauce probably adds a faint sweet-acidic note, similar to ketchup or barbecue sauce, giving the aroma a sharper edge. Altogether, it seems like a noticeable but not highly complex smell, more cozy and appetizing than intensely fragrant.

Texture: The bread looks well toasted on the outside, so the first bite would probably have a light crispness before giving way to a soft interior crumb. The cheese appears warm and pliable, adding a smooth, slightly stretchy layer between the firmer bread slices. The filling looks juicy and a bit bulky, which suggests a dense, meaty bite with some softness from the sauce. Overall, the texture likely combines crunch, softness, and gooeyness in a very tactile, layered way.

Sound: This sandwich looks like it would make a gentle crisp bite sound rather than a loud crackle, because the bread is toasted but still fairly thick and soft inside. The edges may give a light crunch when bitten, especially at the corners where the toast looks drier and firmer. The cheese and saucy filling would likely mute the sound quickly, turning the bite into more of a soft chew than a brittle snap. So the eating sound would probably be a brief crunch followed by a quiet, dense, slightly squishy chew.

Gemini 3.1 Pro — Taste:4.2 Smell:3.8 Texture:4.5 Sound:1.2

Taste: The vibrant green powder covering the entire top surface strongly indicates a rich, earthy, and slightly bitter matcha flavor. Beneath this top layer, the pale creamy section suggests a contrasting sweet and milky profile, likely balancing the bitterness of the tea. The overall taste profile would be a harmonious blend of sweet dairy and astringent, herbaceous green tea.

Smell: The thick dusting of matcha powder will provide a dominant, earthy, and

grassy aroma that is characteristic of high-quality green tea. The creamy layer underneath might offer subtle, sweet dairy or vanilla notes that soften the intense tea scent. Together, these visual cues suggest a fragrant, botanical, and comforting smell that is not overly sweet.

Texture: The fine, dry powder on top will provide a slightly velvety and dusty initial mouthfeel. This will quickly give way to the smooth, rich, and aerated texture of the cream or mousse layer visible just below the rim. The contrast between the dry, powdery topping and the lush, creamy base suggests a delicate and melting consistency on the palate.

Sound: The soft, yielding nature of the powdered topping and the creamy mousse indicates that eating this dessert would produce almost no audible sound. The wooden spoon resting on the saucer might make a dull, gentle clink when picking up a portion, but the food itself is entirely soft. Therefore, any sound would be limited to the quiet, muted noise of the spoon sinking into the delicate layers.

LLaVA — Taste:3.5 Smell:4.0 Texture:3.0 Sound:3.0

Taste: The sandwich is described as savory and rich, with toasted bread, melted cheese, and meaty filling as primary cues.

Smell: Aroma is inferred as warm and appetizing from toasted bread, dairy, and cooked protein.

Texture: Predicted texture combines soft interior components with modest toast crispness.

Sound: Expected sound remains moderate-to-quiet, with occasional soft crunch.

Claude 4.6 Opus — Taste:4.2 Smell:3.9 Texture:4.0 Sound:3.5

Taste: This appears to be a grilled cheese and bacon sandwich made with golden-toasted white bread and what looks like melted American or cheddar cheese layered with strips of crispy bacon. The taste profile would be dominated by rich, savory, salty, and umami notes from the rendered bacon fat and melted cheese, complemented by a mild buttery sweetness from the toasted bread. The combination of salty cured pork and creamy melted cheese creates a deeply satisfying, indulgent flavor that registers high on the palatability scale. A slight caramelization on the bread surface suggests a Maillard reaction that would add toasty, nutty depth to the overall flavor.

Smell: The aroma emanating from this sandwich would be warm and immediately appetizing, led by the smoky, fatty scent of cooked bacon which tends to carry strongly through the air. Melted cheese contributes a rich, slightly tangy dairy note that blends with the toasty, yeasty fragrance of the grilled bread.

The overall olfactory impression would be one of comfort food, with the rendered bacon grease and butter used for toasting creating a savory, slightly smoky haze. The scent would not be particularly complex or layered compared to dishes with herbs or spices, but it would be intensely appetizing and warm.

Texture: The exterior of the bread shows a deep golden-brown crust that indicates a firm, crunchy bite on the outside, while the interior crumb would remain soft and slightly compressed from the grilling process. The melted cheese visible between the slices appears gooey and stretchy, providing a smooth, creamy contrast to the crisp bread. The bacon strips look like they were cooked to a moderately crispy state, adding a snappy, slightly chewy element that breaks up the softness of the cheese and bread. The interplay between the three distinct textures, crunchy toast, molten cheese, and crisp-chewy bacon, creates a satisfying and varied mouthfeel.

Sound: Biting into this sandwich would produce an audible crunch from the well-toasted bread exterior, which shows enough browning and firmness to generate a satisfying initial snap. The bacon would contribute additional crackling sounds as teeth break through the rendered strips, adding a secondary layer of crispness. Once past the outer crust, the sound would diminish as the softer cheese and inner bread muffle the bite, making the acoustic experience front-loaded. Cutting the sandwich in half might produce a slight scraping sound against the plate, and pulling the halves apart would create a quiet stretching sound from the melted cheese strands.

E.3. Steak Rice (0002)

Human GT: Taste=4.3, Smell=4.3, Texture=4.4, Sound=4.1



FoodSense-VL (Stage 2) — Taste:4.0 Smell:3.5 Texture:3.0 Sound:2.5

Taste: The deep, dark color of the steak suggests a robust, savory flavor with a slightly charred tone. The red rice and visible garnish imply a spicy, tomato-forward base. Onion rings add pungency and contrast, and the plated components suggest a bold, well-seasoned profile.

Smell: Grill marks and sheen on the steak suggest a sizzling, savory aroma. The rice color cues warm spice notes, while lime and green onion suggest a bright citrus lift. Overall smell appears rich and layered.

Texture: The steak looks firm and chewy, with fluffy separated rice grains. The tortilla/chip element appears brittle and crunchy, while onion rings add slight bite resistance.

Sound: The dish likely stays relatively quiet overall, with most sound from a crisp snap in the tortilla/chip element and subtle rustling from rice.

Qwen2.5-VL 32B — Taste:4.5 Smell:4.0 Texture:4.0 Sound:2.5

Taste: The dish appears rich and flavorful, with savory, slightly smoky grilled meat, warm earthy orange rice, and a creamy side component. Pickled onions add a tangy, sharp contrast.

Smell: The profile is likely robust, led by the charred aroma of grilled meat and fragrant spices. Lime and fresh vegetables suggest a citrusy lift, while pickled onions add a pungent vinegary note.

Texture: Expected texture is varied: tender-juicy meat with a slightly crispy charred exterior, fluffy moist rice, a creamy-smooth side, and crisp crunchy onion accents.

Sound: Auditory cues are mostly subtle, with imagined sizzle from grilled meat, occasional crunch from onions, and soft chew from rice and side components.

Base (Gemma 3 27B) — Taste:3.8 Smell:3.5 Texture:3.2 Sound:2.5

Taste: The steak likely has a savory, umami-rich profile with hints of char from grilling. The red rice suggests tomato-based seasoning with slight acidity and sweetness. Lime and pickled onions add a fresh, tangy counterpoint that balances the richer components.

Smell: Expected aroma combines smoky, meaty grilled steak with warm spice notes from rice. Lime adds bright citrus, while onions add a pungent, sharp top note.

Texture: Steak appears seared outside and tender, juicy inside. Rice looks fluffy, chips provide distinct crunch, and onions contribute crisp bite with occasional juicy bursts from lime.

Sound: The dish should be moderately audible: chips deliver loud crunch, steak adds soft chew, while rice and garnish remain mostly quiet.

InternVL-2.5 26B — Taste:4.5 Smell:4.0 Texture:4.0 Sound:3.5

Taste: The dish likely offers a rich and diverse flavor profile. The grilled steak would provide a savory, slightly charred taste, complemented by the tangy, creamy sauce. The rice, seasoned with spices, would add a warm, slightly spicy note, while the taco would offer a mix of savory meat and a hint of cheese. The lime would add a fresh, citrusy zing to balance the flavors.

Smell: The aroma would be inviting, with the smoky scent of the grilled steak mingling with the earthy, slightly spicy aroma of the rice. The fresh lime and herbs would add a bright, citrusy note, enhancing the overall olfactory experience.

Texture: The dish would offer a variety of textures. The steak would be tender and juicy, contrasting with the crispy, slightly crunchy taco shell. The rice would be fluffy and slightly sticky, while the sauce would add a smooth, creamy

element. The raw vegetables would provide a crisp, refreshing contrast.

Sound: The sound of this dish would primarily be the sizzle of the steak as it cooks, followed by the gentle clinking of utensils against the plate. The crunch of the taco shell and the soft rustle of the lime and herbs would add subtle auditory details, enhancing the overall sensory experience.

GPT 5.4 Thinking — Taste:4.6 Smell:4.4 Texture:4.4 Sound:3.7

Taste: This dish looks intensely savory and spiced, with grilled meat, seasoned rice, and a filled tortilla shell suggesting a bold, layered flavor profile. The char on the meat implies smoky, slightly bitter edges that would deepen the richness, while the onions and salsa-like garnish likely add brightness and a mild acidic lift. The reddish rice appears well seasoned, hinting at tomato, garlic, and spice, which would make the plate feel hearty and robust. Overall, it seems likely to taste rich, smoky, salty, and warmly spiced, with small bursts of freshness from the toppings.

Smell: The strongest visual cue is the grilled meat, which suggests a pronounced smoky, roasted aroma coming off the dark seared surface. The seasoned rice and stuffed shell likely contribute warm smells of spices, oil, and cooked starch, creating a full and comforting scent. Fresh onion, lime, and nearby vegetables imply a lighter sharpness that could cut through the heavier grilled notes. Altogether, the aroma seems vivid and appetizing, combining smoke, spice, and a fresh acidic edge.

Texture: The meat looks firm and juicy, with a charred exterior that would give some resistance before yielding to a tender interior. The rice appears fluffy yet oily enough to cling together slightly, creating a soft and granular mouthfeel. The stuffed tortilla shell seems crisp at the edges but softened by its filling, adding a contrast between crunch and softness. Overall, the plate likely offers a satisfying mix of chew, fluffiness, crisp edges, and moist, juicy bites.

Sound: This meal seems moderately audible, mostly through the slight scrape and bite resistance of the grilled meat rather than any loud crunch. The charred exterior of the meat and the crisp shell edges could produce a faint crackle or crisp snap when cut or bitten. The rice and toppings would be much quieter, contributing soft, muted eating sounds. So the overall sound profile would likely be a combination of light crispness and steady, substantial chewing rather than sharp crunches.

Gemini 3.1 Pro — Taste:4.5 Smell:4.6 Texture:4.3 Sound:2.8

Taste: The dark, glossy glaze coating the skewers suggests a rich, deeply savory, and slightly sweet flavor profile characteristic of teriyaki or tare sauce.

Prominent charred edges on the meat pieces indicate a robust, bitter, and smoky undertone that balances the sweetness. The presence of a fresh lemon wedge implies an optional burst of sharp acidity to cut through the heavy umami flavors.

Smell: The heavily blistered and blackened edges of the meat point to a strong, intense smoky aroma from high-heat grilling. The thick, caramelized coating will contribute a fragrant, savory scent of roasted soy and caramelized sugars. An underlying hint of bright, fresh citrus will mingle in the air if the adjacent lemon is utilized.

Texture: The sticky, glistening sauce promises a tacky and clinging mouthfeel upon the first bite. This is immediately contrasted by the crispy, rigid texture of the charred exterior crust gripping the meat. The distinct, thick chunks threaded onto the wooden sticks appear plump and moist, suggesting a tender, yielding interior beneath the slight exterior crunch.

Sound: Removing the sticky chunks of meat from the wooden skewers will likely produce a subtle, damp sliding noise. The first bite will yield a faint but distinct crunching sound due to the crystallized glaze and charred edges. Subsequent chewing will be relatively quiet and soft, dominated by the tender and juicy interior of the grilled meat.

LLaVA — Taste:3.5 Smell:4.0 Texture:3.0 Sound:3.0

Taste: The dish appears Mexican-inspired, with savory meat, sauce, and rice as the core flavor base. Onion and tomato elements suggest fresh tang and contrast.

Smell: Likely aromatic and savory, with grilled-meat notes and spiced sauce. Rice adds a mild earthy background smell.

Texture: Expected texture mix is tender meat, soft rice, and some chew from tortilla components, with occasional crispness from garnish.

Sound: Predominantly gentle eating sounds, with occasional crunch from tortilla elements and soft utensil contact.

Claude 4.6 Opus — Taste:4.5 Smell:4.3 Texture:4.2 Sound:3.2

Taste: This is a Mexican-style plate featuring a large grilled or pan-seared steak topped with sliced red onion rings and what appears to be a warm salsa or ranchero sauce, alongside a portion of Mexican rice and a taco or tostada with crumbled cheese. The steak shows deep char marks and dark browning across its surface, indicating intense Maillard caramelization that would deliver bold,

smoky, beefy umami flavors. The rice has a vivid orange-red hue suggesting it was cooked with tomato, cumin, and possibly chili, contributing a savory, mildly spicy, and slightly tangy base note. The layered composition of rich grilled meat, seasoned rice, and a cheese-topped taco creates a deeply satisfying flavor profile with interplay between smoky, salty, tangy, and mildly spicy elements.

Smell: The dominant aroma would be the smoky, charred scent of the grilled steak, which carries powerfully and is one of the most universally appetizing smells in cooking. Underneath that, the Mexican rice would contribute a warm, toasty fragrance with notes of cumin, garlic, and tomato that are characteristic of well-seasoned arroz rojo. The raw red onion rings and the salsa draped over the meat would add a sharp, slightly pungent, and bright acidic note that cuts through the heavier meat and fat aromas. Fresh ingredients visible nearby on the table, including cherry tomatoes, green onions, and a lime, suggest garnishing elements whose bright, herbaceous, and citrus scents would round out the olfactory experience with freshness.

Texture: The steak appears to have a firm, well-seared crust on the exterior from the grill marks, which would give way to a denser, chewier interior consistent with a thicker cut cooked to at least medium doneness based on the degree of surface char. The Mexican rice looks fluffy yet slightly sticky, with individual grains holding together in soft clumps that would provide a pillowy contrast to the meat. The taco or tostada shell has a visible corn tortilla base that would contribute a light crunch or soft pliability depending on whether it was fried or simply warmed, and the crumbled cheese on top adds a crumbly, slightly grainy texture. The sliced onion rings on the steak remain raw and crisp, providing a snappy, juicy bite that contrasts sharply with the cooked elements.

Sound: Cutting into the steak with a knife would produce a firm resistance and a light scraping sound against the plate, and the seared crust might yield a faint crackling as the blade breaks through the charred exterior. Chewing the rice would be relatively quiet, producing soft, muffled sounds typical of cooked grains, contributing minimal acoustic presence to the eating experience. The raw onion rings would offer a mild snap and a juicy crunch when bitten, adding a brief audible moment to each bite that includes them. If the taco shell was lightly fried or toasted, biting into it would generate a moderate crunch, though this would be somewhat dampened by the moist fillings and melted cheese on top.

Food-Llama 11B — Taste:4.0 Smell:4.0 Texture:4.0 Sound:3.0

Taste: The plate suggests a rich savory profile from grilled steak and seasoned rice, with tang from onion/lime garnish and creamy contrast from cheese-covered tortilla.

Smell: Likely smoky and warm from the grilled meat, with a spiced rice aroma and fresh citrus-herb top notes.

Texture: Expected texture is mixed: tender-chewy steak, fluffy rice, crunchy tortilla/chip components, and creamy-soft topping elements.

Sound: Audible cues are moderate, mainly from crisp tortilla crunch and light utensil sounds, with softer chew from meat and rice.

E.4. Southern Scampi (0010)

Human GT: Taste=3.7, Smell=3.5, Texture=3.5, Sound=2.7



FoodSense-VL (Stage 2) — Taste:3.5 Smell:3.0 Texture:2.5 Sound:2.0

Taste: The baked scampi pasta appears savory and creamy, with seafood notes from shrimp and mild acidity from tomato pieces.

Smell: Predicted aroma is warm and dairy-forward, with cooked seafood and herb accents.

Texture: The dish looks mostly soft and yielding: creamy sauce, tender pasta, and small shrimp pieces with limited crunch.

Sound: Overall sound is expected to be quiet, with soft utensil and chew cues rather than sharp crackle.

Base (Gemma 3 27B) — Taste:3.8 Smell:3.5 Texture:3.2 Sound:2.0

Taste: Visible shrimp, pasta, and creamy, cheesy sauce suggest a rich, savory profile. Herbs and tomato pieces likely add mild acidity that cuts richness. The scampi label implies buttery, garlic-forward flavor with possible mild spice.

Smell: The dish likely smells of cooked seafood and dairy with herb notes. Browned cheese suggests a slightly caramelized, nutty scent. A buttery and subtle garlic aroma is expected.

Texture: Likely layered: soft pasta, shrimp with slight bite, and creamy, smooth sauce. The top may have mild crisp edges, but overall the dish appears soft and yielding.

Sound: Expected minimal sound during eating, mostly utensil contact with container and soft chew; little crunch or snap.

InternVL-2.5 26B — Taste:4.5 Smell:4.0 Texture:4.0 Sound:3.5

Taste: The Southern Scampi pasta likely offers a rich, savory flavor profile with a hint of garlic and lemon, complemented by the brininess of the shrimp. The creamy sauce would add a smooth, indulgent quality, while the pasta provides a satisfying chew.

Smell: The aroma of the dish would be enticing, with the scent of sauteed garlic and herbs mingling with the buttery, cheesy sauce. The shrimp would add a subtle seafood aroma, making the dish smell hearty and comforting.

Texture: The texture of the pasta would be al dente, providing a pleasant resistance to the bite. The creamy sauce would coat the pasta evenly, while the shrimp would offer a tender, slightly firm bite. The cheese on top would be melted and gooey, adding a contrasting softness.

Sound: While the dish itself wouldn't make much noise, the act of eating it would likely involve the satisfying sounds of pasta being twirled and the occasional crunch of shrimp. The creamy sauce might create a slight slurping sound as it coats the pasta.

Qwen2.5-VL 32B — Taste:4.0 Smell:4.0 Texture:4.5 Sound:2.5

Taste: The image shows a dish labeled "Southern Scampi Pasta," which suggests a rich and flavorful meal. The visible creamy sauce indicates a likely buttery and savory taste profile, enhanced by the presence of what appears to be shrimp or scampi, adding a briny and slightly sweet element. The golden-brown crust on top hints at a crispy, toasted texture that could contribute a nutty flavor when combined with the pasta. The overall appearance suggests a well-balanced dish with a mix of umami, salty, and slightly sweet notes.

Smell: While the image doesn't convey smell directly, the visual cues suggest an aromatic experience. The creamy sauce and the golden-brown crust imply a scent of melted butter, garlic, and possibly herbs like parsley or basil, which are common in scampi dishes. The shrimp or scampi would add a fresh seafood aroma, while the toasted elements might introduce a warm, nutty fragrance. The combination of these elements suggests a comforting and inviting smell.

Texture: The dish presents a variety of textures. The pasta appears soft and al dente, coated in a creamy sauce that looks smooth and velvety. The golden-brown crust on top suggests a crunchy, slightly crispy layer, providing contrast to the softer elements. The shrimp or scampi pieces likely have a tender, slightly chewy texture, complementing the smoothness of the sauce and the firmness of the pasta. The overall texture seems well-balanced, offering a satisfying mouthfeel.

Sound: The image itself does not provide auditory information, but based on the visual cues, one can infer potential sounds associated with eating this dish. The creamy sauce might make a slight slurping sound as it coats the pasta, while the crispy crust could produce a satisfying crunch when bitten into. The shrimp or scampi might offer a gentle sizzle if freshly cooked, though this is speculative without actual audio. Overall, the sound experience would likely be subtle but enjoyable, focusing on the satisfaction of textures rather than loud noises.

GPT 5.4 Thinking — Taste:4.4 Smell:4.5 Texture:4.2 Sound:2.6

Taste: This dish looks rich, savory, and buttery, with the browned cheese and creamy surface suggesting a deeply comforting flavor. The label and appearance imply a shrimp-based preparation, so it likely has a mild seafood sweetness underneath garlic, salt, and herb-heavy seasoning. The golden spots on top suggest oven-browned edges that would add a slightly roasted, almost nutty note. Overall, it seems likely to taste warm, creamy, and indulgent, with a strong

savory depth and a gentle briny lift.

Smell: The visual cues point to a pronounced aroma of butter, garlic, and baked cheese rising from the hot surface. If the dish is indeed shrimp-based, there would probably also be a light oceanic smell, but softened by the creamy sauce and herbs rather than sharp or fishy. The browned top suggests a toasty baked scent that makes the whole dish feel heavier and more fragrant. Altogether, it likely smells robust, cozy, and highly appetizing, with garlic and butter leading the profile.

Texture: The top looks soft and molten, with a creamy, slightly thick consistency under a lightly browned surface. The shrimp or filling pieces would probably add tender, springy bites that contrast with the smoother sauce and soft base underneath. The baked finish suggests a faint crust or resistance at the top edges, but most of the dish appears lush and spoonable rather than crisp. Overall, the texture likely feels creamy, tender, and substantial, with only light textural variation from the browned surface.

Sound: This looks like a fairly quiet food to eat, since creamy baked dishes rarely produce much crunch or snap. The only likely audible element would be a faint scrape of the utensil against the container and a soft pull through the thicker baked top. Any seafood pieces might offer a slight firm bite, but not enough to create a strong eating sound. So the sound profile would probably be muted and soft, dominated more by gentle scooping than by crispness.

Gemini 3.1 Pro — Taste:4.4 Smell:4.5 Texture:4.3 Sound:3.8

Taste: The heavily blistered and charred edges of the crust suggest a smoky, slightly bitter flavor that balances the pie's richness. The thick layer of bright green, oily sauce indicates a vibrant, herbaceous, and garlicky taste profile typical of fresh pesto. The extensive coverage of melted white cheese will provide a milky, salty, and savory umami base that ties the strong flavors together.

Smell: The heavily bubbled and blackened dough points to a strong aroma of toasted yeast and wood-fired smoke. The generous application of the green sauce likely emits a pungent, fragrant scent of fresh basil, crushed garlic, and rich olive oil. The baked, gooey cheese will contribute a warm, comforting, and distinctly rich dairy fragrance to the overall bouquet.

Texture: The deeply charred and prominent air bubbles on the outer crust promise a rigid, brittle, and highly crispy exterior upon the first bite. This dry crunch will quickly give way to a soft, chewy, and airy interior dough network beneath the toppings. The heavy layer of melted cheese and oily pesto will coat the palate with a slick, gooey mouthfeel that contrasts sharply with the crust.

Sound: Biting into the thick, heavily blistered crust will produce a loud, sharp crackling and crunching noise. As chewing continues, the thick and chewy interior dough will create a dense, muffled sound profile. The wet, oily surface of the melted cheese and pesto might generate a faint squelching or smacking audio alongside the dominant crust crunch.

LLaVA — Taste:3.5 Smell:4.0 Texture:3.0 Sound:3.0

Taste: The model describes a rich creamy pasta profile with seafood and herb components.

Smell: Aroma is inferred as buttery and inviting, with seafood and cooked herb notes.

Texture: Predicted texture is a soft pasta base with some chew from seafood and sauce thickness.

Sound: Expected sound is modest, mainly soft clinking/slurping with limited crispness.

Claude 4.6 Opus — Taste:3.6 Smell:3.3 Texture:3.4 Sound:1.8

Taste: This is a packaged takeout container labeled "Southern Scampi Pasta" from what appears to be a brand called Slim and Hungry, featuring a baked pasta dish with visible shrimp, a creamy cheese sauce, and what looks like diced tomato or red pepper mixed throughout. The flavor would center on a rich, buttery, garlicky scampi sauce coating the pasta, with the shrimp providing a sweet, briny seafood note that anchors the dish. The visible layer of melted and lightly browned cheese across the top suggests a gratinated finish that would add a salty, savory depth and a slight nuttiness from the browning. Being a pre-packaged, ready-to-eat meal, the flavors may be somewhat muted compared to freshly prepared scampi, as reheating and holding tend to dull the brightness of garlic and butter, though the overall combination of creamy pasta, shrimp, and cheese remains an inherently satisfying comfort food profile.

Smell: Upon opening the clear plastic clamshell lid, the primary aroma would be a warm, creamy, dairy-heavy scent from the melted cheese and butter-based scampi sauce that has been baked over the pasta. Garlic would be present as a background note, though in a pre-packaged product like this the garlic aroma tends to be mellowed and integrated rather than sharp and pungent as it would be in a freshly sauteed scampi. The shrimp would contribute a subtle, slightly sweet, oceanic undertone that mingles with the cheese but does not dominate the

aromatic profile. The overall smell impression would be comforting and familiar, reminiscent of a baked mac and cheese with seafood additions, though lacking the aromatic intensity and freshness of a dish plated moments after cooking.

Texture: The pasta appears to be a short tubular shape, likely penne or rigatoni, that looks well-cooked and coated in a thick, clinging cream sauce, suggesting a soft, tender bite with a smooth, almost velvety mouthfeel from the sauce. The cheese layer on top shows some light browning and appears to have formed a thin, slightly firmer skin across the surface, which would provide a mild contrast to the softer pasta beneath when first broken through. The shrimp visible near the edges appear to be medium-sized and cooked through, and in a baked preparation like this they would likely be somewhat firm and slightly rubbery rather than the tender, snappy texture of freshly sauteed shrimp. The overall textural experience would be predominantly soft, smooth, and homogeneous, which is characteristic of baked pasta dishes where extended heat and a creamy sauce unify the components into a cohesive, comfort-food consistency.

Sound: This baked pasta dish would produce very little audible sound during eating, as the soft, sauce-coated pasta, tender shrimp, and melted cheese are all quiet foods that compress easily under the teeth without cracking, snapping, or crunching. The most noticeable sound would be the initial snap and crinkle of opening the plastic takeout container and peeling back the lid, which is unrelated to the food itself. A fork pressing into the baked cheese top might produce a very faint crackling if the gratinated layer has dried slightly, but this would be barely perceptible. The eating experience would be essentially silent, dominated by soft chewing and the quiet squish of cream-coated pasta, making this one of the quieter dishes on the sensory sound spectrum.

E.5. Ice Cream (0015)

Human GT: Taste=4.1, Smell=3.3, Texture=3.7, Sound=1.8



FoodSense-VL (Stage 2) — Taste:4.5 Smell:4.0 Texture:4.5 Sound:3.5

Taste: Visual cues suggest a vanilla-forward sweet profile with caramel drizzle adding buttery depth.

Smell: The dessert likely carries a sweet dairy aroma with caramelized sugar notes.

Texture: The scoop appears creamy and smooth, with sticky syrup contrast from the topping.

Sound: Eating sound is likely subtle, mostly spoon contact and soft melt cues.

Base (Gemma 3 27B) — Taste:4.2 Smell:3.8 Texture:4.5 Sound:2.5

Taste: The ice cream appears vanilla-based with caramel drizzle, suggesting sweet, buttery, and slightly burnt-sugar notes. The base likely tastes creamy and dairy-rich, with a generally indulgent dessert profile.

Smell: Visual cues suggest a sweet, milky aroma with prominent caramel notes. A fresh dairy scent is likely, with possible subtle vanilla nuances.

Texture: The scoop appears smooth, dense, and creamy, with melt-in-mouth behavior. Caramel drizzle looks viscous and sticky, adding textural contrast.

Sound: Sound profile is minimal, mainly spoon scraping and soft contact with the cup. Minor crunch might occur only with optional toppings.

InternVL-2.5 26B — Taste:4.5 Smell:4.0 Texture:4.5 Sound:3.5

Taste: The ice cream likely offers a rich, creamy vanilla base, complemented by a sweet, caramel drizzle that adds a deeper, buttery flavor. The combination of these elements suggests a delightful balance of sweetness and smoothness.

Smell: The aroma of the ice cream is likely to be enticing, with the vanilla providing a warm, comforting scent and the caramel adding a hint of toasted sugar, creating an inviting and appetizing fragrance.

Texture: The ice cream itself is probably smooth and velvety, with a soft,

creamy consistency that melts slightly in the mouth. The caramel drizzle adds a contrasting texture, being slightly thicker and more viscous, enhancing the overall eating experience.

Sound: The sound of scooping the ice cream might be satisfying, with a soft, creamy crunch as the spoon cuts through the frozen treat. The drizzle of caramel might add a slight glistening sound as it flows over the ice cream, contributing to the sensory experience.

Qwen2.5-VL 32B — Taste:4.5 Smell:4.0 Texture:4.5 Sound:2.5

Taste: The image shows a scoop of ice cream in a striped cup, topped with what appears to be a drizzle of caramel sauce. The creamy texture of the ice cream suggests a smooth and rich flavor profile, likely sweet and slightly tangy from the vanilla base. The caramel drizzle adds a warm, buttery sweetness that complements the coolness of the ice cream, creating a balanced and indulgent taste experience.

Smell: While the image does not directly convey smell, the visual cues suggest a delightful aroma. The vanilla ice cream is likely to have a classic, sweet, and slightly floral scent, while the caramel drizzle would add a warm, nutty, and slightly toasted fragrance. These scents combined would evoke a comforting and inviting olfactory experience, enhancing the overall sensory appeal.

Texture: The ice cream appears soft and creamy, indicating a smooth and velvety texture that melts gently in the mouth. The caramel drizzle on top likely has a glossy, slightly sticky texture that contrasts with the cool, fluffy consistency of the ice cream. This combination offers a satisfying interplay between the soft, melting ice cream and the slightly firmer caramel topping.

Sound: The image itself does not provide auditory information, but one can imagine the sound of the ice cream being scooped into the cup or the slight crunch of any potential toppings like nuts or sprinkles. However, the primary focus is on the visual elements, leaving the auditory experience to the imagination. The lack of direct audio cues limits the assessment in this area.

GPT 5.4 Thinking — Taste:4.3 Smell:3.8 Texture:4.6 Sound:2.1

Taste: This dessert looks distinctly sweet, creamy, and smooth, with the pale ice cream and caramel drizzle suggesting a rich, vanilla-forward flavor. The glossy sauce likely adds buttery sweetness and a slightly deeper toasted-sugar note that makes the cup feel more indulgent. Visually, there are no sharp or bitter elements, so the flavor impression is soft, rounded, and dessert-like rather than complex or contrasting. Overall, it seems likely to taste mellow, sugary, and comforting, with a clean dairy richness.

Smell: The image suggests a gentle but pleasant aroma led by cold dairy, vanilla, and sweet caramel. Ice cream usually has a softer scent than hot foods, so the smell would likely be noticeable only at close range rather than strongly filling the air. The caramel topping implies a faint buttery, cooked-sugar fragrance that adds warmth to the otherwise cool profile. Altogether, it likely smells sweet and inviting, but relatively delicate.

Texture: The scoop looks exceptionally smooth and dense, suggesting a creamy, velvety mouthfeel with very little iciness. The caramel drizzle appears slightly syrupey, which would add a glossy, sticky layer over the colder, firmer ice cream surface. As it softens, the dessert would probably become even silkier and more luscious, coating the mouth evenly. Overall, the texture seems dominated by softness, creamininess, and a slow melt.

Sound: This looks like a very quiet food, since smooth ice cream and syrup produce almost no natural crunch or crackle. The main sounds would likely come from the spoon scraping lightly against the paper cup and the soft cut through the dense frozen scoop. Without visible toppings like nuts or crisp mix-ins, there is little reason to expect sharp biting sounds. So the sound profile would probably be minimal, muted, and gentle.

Gemini 3.1 Pro — Taste:4.5 Smell:4.6 Texture:4.3 Sound:2.5

Taste: The dense, dark brown gravy enveloping the meat suggests a deeply rich, savory, and heavily seasoned flavor profile typical of a robust curry or stew. The large chunks of slow-cooked meat indicate a meaty, concentrated umami taste that permeates the thick sauce. An underlying earthy, pungent taste is likely due to the heavily caramelized base, providing a complex and possibly spicy finish.

Smell: The dark, glossy, and heavily simmered sauce promises an intense, fragrant aroma dominated by toasted spices, caramelized onions, and roasted meat. The thick consistency suggests the smell is heavy and lingering, characteristic of a slow-cooked, complex curry. A robust, earthy, and pungent scent will strongly emanate from the hot, deeply colored gravy.

Texture: The meat, clearly broken down and coated entirely in the thick sauce, promises an extremely tender, yielding, and melt-in-the-mouth consistency. The dense, glossy gravy will provide a heavy, clinging, and velvety mouthfeel that coats the palate completely. The lack of distinct, separate vegetables suggests a mostly uniform, soft, and unctuous experience, heavily weighted toward the rich sauce.

Sound: The dense and wet nature of the dish implies that interacting with it will

produce heavy, dull sloshing and soft squelching noises. Because the meat and sauce are both soft and slow-cooked, there will be no sharp crunch or crackle during eating. Instead, the sound profile will consist of muted, wet slaps as the thick gravy moves, and silent, tender yielding during chewing.

LLaVA — Taste:4.5 Smell:4.0 Texture:4.0 Sound:3.5

Taste: The output suggests a rich sweet-cream base with caramel-driven buttery sweetness.

Smell: Aroma is inferred as warm and dessert-like from vanilla-dairy and caramel notes.

Texture: Predicted texture is soft and creamy with sticky syrup contrast.

Sound: Sound is low-intensity, mostly spoon and soft mouthfeel-related cues.

Claude 4.6 Opus — Taste:4.0 Smell:3.5 Texture:4.6 Sound:1.5

Taste: This is a serving of ice cream from Creamistry, a chain known for making ice cream to order using liquid nitrogen, presented as a generous scoop of what appears to be vanilla or sweet cream ice cream drizzled with a golden caramel or salted caramel sauce in a branded paper cup. The base flavor would be a clean, rich, creamy vanilla with a high dairy fat content, as liquid nitrogen freezing produces an exceptionally smooth and dense product with minimal ice crystal formation. The caramel drizzle adds a warm, buttery sweetness with deep toasted sugar notes and possibly a hint of sea salt if it is the salted variety, creating a pleasant sweet-on-sweet layering that avoids monotony through the contrast of cool cream and warm caramel flavor tones. Jars of toppings visible in the background, including what appear to be granola, jellybeans, and candies, suggest an array of mix-in options, though this particular serving looks elegantly simple and lets the quality of the base ice cream and caramel shine.

Smell: The aroma of this ice cream would be delicate and subdued due to the extremely cold temperature, which suppresses volatile compound release, but as the surface begins to temper slightly a gentle, sweet, dairy-rich vanilla scent would emerge. The caramel sauce, being warmer and more viscous than the frozen base, would be the more aromatically active component, releasing a toasty, buttery, cooked-sugar fragrance with hints of brown butter and possibly a faint salted note. Because Creamistry uses liquid nitrogen flash-freezing, the ice cream tends to be served at a colder-than-average temperature, which further dampens the olfactory experience in the initial moments compared to traditionally churned ice cream. As the scoop sits and the surface layer begins to soften, the combined vanilla-caramel aroma would gradually intensify into a warm, dessert-shop sweetness.

Texture: Liquid nitrogen ice cream is renowned for its exceptionally smooth, dense, and velvety texture, and this scoop displays the hallmark characteristics of that process with a tight, glossy surface that shows no visible ice crystals or air pockets. The first spoonful would glide across the tongue with an almost silken quality, melting evenly and coating the palate with a rich, heavy cream sensation that feels noticeably denser than conventional soft serve or store-bought ice cream. The caramel drizzle adds a contrasting textural element, introducing a slightly sticky, viscous ribbon that clings to the spoon and stretches as it is pulled, creating a pleasant interplay between the smooth frozen cream and the syrupy sauce. The paper cup itself provides a slight tactile warmth in contrast to the frozen contents, and the rounded scoop shape suggests it was portioned with care, maintaining structural integrity that indicates a well-balanced fat and sugar ratio in the base.

Sound: Ice cream is among the quietest foods to eat, and this serving would produce almost no audible sound during consumption, as the dense, creamy base simply compresses and melts under the pressure of a spoon or against the warmth of the mouth. The most prominent sound would be the spoon scraping lightly against the bottom and sides of the paper cup as the serving is eaten, producing a soft, muffled scratching that is characteristic of disposable dessert containers. If any of the topping options from the background jars were added, such as granola or candy pieces, they would introduce crunching and crackling sounds, but as presented this plain scoop with caramel is essentially silent. The only other incidental sound might be a faint sticky separation noise as the caramel sauce pulls apart between the spoon and the ice cream surface.

Food-Llama 11B — Taste:3.0 Smell:2.0 Texture:4.0 Sound:1.0

Taste: The generated output indicates sweet vanilla-like ice cream with caramel richness and an overall comfort-dessert profile.

Smell: Aroma is described as mild sweet dairy with caramel/toffee-like notes.

Texture: Texture is described as smooth and creamy with a sticky syrup contrast.

Sound: Predicted sound is very soft, mostly spoon scraping and gentle melt/slurp cues.

E.6. Taco (0005)

Human GT: Taste=4.5, Smell=4.4, Texture=4.5, Sound=3.7



FoodSense-VL (Stage 2) — Taste:4.2 Smell:4.0 Texture:4.5 Sound:3.8

Taste: The tacos show fresh avocado, cabbage, pico de gallo, and grilled protein, suggesting a balanced savory profile with acidity and heat. The tortilla color implies lightly toasted nutty notes.

Smell: Steam and fresh garnish indicate a warm, appetizing aroma combining herbaceous freshness, smoky grilled notes, and spicy jalapeño character.

Texture: Strong texture contrast is expected: soft avocado, crisp cabbage/jalapeños, juicy pico, and tender grilled filling in pliable tortillas.

Sound: Likely moderate and not loud, with gentle crunch from vegetables and subtle chew from tortilla and protein.

Base (Gemma 3 27B) — Taste:4.2 Smell:3.8 Texture:4.5 Sound:2.5

Taste: The tacos appear to contain grilled fish, likely mild white fish, with slightly sweet seafood notes. Avocado and pico suggest fresh acidity and creamy richness. Jalapeños add moderate heat, while slaw likely contributes tangy balance.

Smell: Expected aroma combines smoky, savory grilled fish with fresh, herbaceous cilantro and zesty pico. Avocado adds subtle buttery notes and jalapeños add a pungent, spicy edge.

Texture: The dish shows strong texture variety: soft tortilla, flaky fish, creamy avocado, crisp slaw, and firm jalapeño slices. Pico appears juicy and slightly chunky, giving a complex mouthfeel.

Sound: Eating is likely relatively quiet; soft tortilla and fish produce little noise, while slaw and jalapeños add occasional light crunch. Overall auditory intensity remains low.

InternVL-2.5 26B — Taste:4.5 Smell:4.0 Texture:4.5 Sound:3.5

Taste: The tacos likely offer a delightful mix of flavors, with the creamy coleslaw providing a tangy and slightly sweet taste, complemented by the fresh avocado's rich, buttery flavor. The jalapeños add a spicy kick, while the grilled meat offers a savory, smoky element. The salsa adds a burst of fresh, slightly acidic flavor, enhancing the overall taste experience.

Smell: The aroma of the grilled meat and tortillas would be prominent, with a smoky, slightly charred scent. The fresh cilantro and avocado would add a bright, herbaceous note, while the jalapeños would contribute a mild, spicy aroma. The salsa would add a hint of fresh, ripe tomatoes and onions.

Texture: The tacos would offer a variety of textures, from the soft, slightly chewy tortillas to the crisp, fresh jalapeños and avocado slices. The coleslaw would provide a crunchy contrast, while the grilled meat would be tender yet slightly firm. The salsa would add a juicy, chunky texture, enhancing the overall mouthfeel.

Sound: The sound of biting into the tacos would likely be satisfying, with the crunch of the jalapeños and avocado, the slight snap of the coleslaw, and the soft yield of the grilled meat. The salsa would add a juicy, slightly messy element, contributing to the auditory experience of eating.

Qwen2.5-VL 32B — Taste:4.5 Smell:4.0 Texture:4.5 Sound:2.5

Taste: The tacos appear to combine grilled protein, avocado, jalapeños, and fresh salsa for a savory base with creamy and spicy contrast.

Smell: Visual cues suggest a fresh aromatic profile with herbs/vegetables, mild smoke from grilled filling, and zesty salsa character.

Texture: Soft tortilla and avocado contrast with crunch from jalapeños and chunky salsa; overall mouthfeel appears varied and engaging.

Sound: The expected sound profile is moderate, with occasional crunch from crisp components and otherwise soft bite sounds.

GPT 5.4 Thinking — Taste:4.5 Smell:4.1 Texture:4.7 Sound:3.8

Taste: These tacos look bright, savory, and fresh, with the avocado, cabbage slaw, jalapeños, and salsa suggesting a balance of richness, acidity, and heat. The grilled tortillas and likely seasoned filling underneath imply a warm, smoky base flavor that would support the cooler toppings. The green jalapeño slices indicate a noticeable spicy kick, while the pico de gallo adds juicy sweetness and tang from tomato and citrus. Overall, the flavor seems layered and lively, combining fresh vegetal notes with creamy, spicy, and grilled elements.

Smell: The strongest likely aromas come from the warmed corn tortillas and the cooked filling, which suggest a toasty, savory scent rising from the plate. Fresh jalapeño, herbs, and salsa would add sharp green notes and a slightly citrusy brightness that makes the dish smell vivid and clean. The avocado contributes less aroma visually, but it reinforces the impression of freshness and richness. Altogether, the smell likely feels inviting and energetic, with a mix of grilled warmth and crisp produce.

Texture: The tacos appear especially textural, with soft avocado, crunchy shredded cabbage, tender filling, and lightly charred tortillas all stacked together. The jalapeño slices would add a crisp snap, while the salsa looks juicy and chunky enough to bring bursts of moisture. The tortillas seem pliable but slightly blistered, so they would probably fold softly while still offering a faint toasted resistance. Overall, the eating experience looks highly varied and satisfying, moving between creamy, crisp, tender, and juicy textures in each bite.

Sound: These tacos would likely make a modest but noticeable eating sound, mostly from the crunch of cabbage and the snap of fresh jalapeño. The tortillas seem soft enough to avoid a brittle crack, though their charred edges may give a slight dry bite sound. The juicy filling and avocado would dampen louder sounds quickly, keeping the overall profile softer than crunchy fried food. So the sound would probably be a mix of light crispness and moist chewing rather than anything loud or sharp.

Gemini 3.1 Pro — Taste:4.6 Smell:4.4 Texture:4.8 Sound:4.5

Taste: The heavily browned, craggy exterior of the fried chicken suggests a deeply savory and rich flavor profile, enhanced by the fat of the frying oil. The shiny, clinging sauce coating the pieces points toward a robust, salty, and likely tangy or sweet glaze that dominates the palate. A generous dusting of grated hard cheese and fresh green herbs will provide an additional layer of sharp, salty umami and a fleeting, bright herbal finish to balance the richness.

Smell: The visible thick, crispy breading and slick sauce strongly imply an intense, fragrant aroma of hot frying oil and savory roasted meat. The glossy glaze likely emits pungent, mouth-watering scents of garlic, vinegar, or caramelized sugars depending on the specific sauce profile. Meanwhile, the white dusting of grated cheese on top will introduce a distinctly sharp, pungent dairy note that cuts through the heavy fried aromas.

Texture: The extremely jagged and blistered crust enveloping the chicken pieces promises a rigid, brittle, and highly crunchy initial bite. This substantial crunch is immediately contrasted by the slick, sticky mouthfeel of the wet sauce heavily painted across the surface. Beneath this thick exterior armor, the plump pieces of poultry imply a tender, fibrous, and juicy interior that easily yields to chewing.

Sound: Piercing the deeply rigid and uneven fried coating will produce a loud, sharp cracking and shattering sound upon the first bite. Continuous chewing will yield a sustained, noisy crunching audio profile as the thick crust breaks down in the mouth. Additionally, the heavy, wet glaze might generate faint smacking or squelching noises that complement the dominant, aggressive crunch of the breading.

Claude 4.6 Opus — Taste:4.4 Smell:4.1 Texture:4.5 Sound:3.0

Taste: These appear to be fish or chicken tacos in charred corn tortillas, generously topped with sliced avocado, fresh jalapeño rounds, shredded cabbage slaw, and green onion, with a side of fresh pico de gallo containing diced tomato, onion, and what looks like mango or yellow pepper. The flavor profile would be remarkably layered, starting with a mild, savory protein base and building through the creamy richness of the avocado, the bright heat of the raw jalapeño slices, and the tangy crunch of the cabbage slaw which likely has a lime or vinegar dressing. The pico de gallo would deliver a burst of fresh acidity and sweetness that cuts through the richer elements, while the lightly charred corn

Table E12. Sensory ratings on six representative food images (1–5 scale). Human GT (mean of ~25 annotators) is shown for comparison. Best model prediction (closest to GT) per cell in **bold**. “–” = unavailable.

	Sandwich				Steak Rice				Taco				Scampi				Ice Cream			
	T	S	X	U	T	S	X	U	T	S	X	U	T	S	X	U	T	S	X	U
Human GT	4.1	4.1	4.3	3.6	4.3	4.3	4.4	4.1	4.5	4.4	4.5	3.7	3.7	3.5	3.5	2.7	4.1	3.3	3.7	1.8
<i>FoodSense-VL</i>																				
FoodSense-VL (Stage 2)	3.5	3.0	2.5	2.0	4.0	3.5	3.0	2.5	4.2	4.0	4.5	3.8	3.5	3.0	2.5	2.0	4.5	4.0	4.5	3.5
Base (Gemma 3 27B)	3.5	3.0	2.5	1.5	3.8	3.5	3.2	2.5	4.2	3.8	4.5	2.5	3.8	3.5	3.2	2.0	4.2	3.8	4.5	2.5
<i>Open-source baselines</i>																				
InternVL-2.5 26B	4.5	4.0	4.0	3.5	4.5	4.0	4.0	3.5	4.5	4.0	4.5	3.5	4.5	4.0	4.0	3.5	4.5	4.0	4.5	3.5
Qwen2.5-VL 32B	4.5	4.0	4.5	2.5	–	–	–	–	4.5	4.0	4.5	2.5	4.0	4.0	4.5	2.5	4.5	4.0	4.5	2.5
LLaVA	3.5	4.0	3.0	3.0	3.5	4.0	3.0	3.0	3.5	3.0	4.0	3.0	3.5	4.0	3.0	3.0	4.5	4.0	4.0	3.5
Food-Llama 11B	–	–	–	–	4.0	4.0	4.0	3.0	3.0	2.5	4.5	2.0	–	–	–	–	3.0	2.0	4.0	1.0
<i>Commercial models</i>																				
GPT 5.4 Thinking	4.2	3.9	4.3	3.4	4.6	4.4	4.4	3.7	4.5	4.1	4.7	3.8	4.4	4.5	4.2	2.6	4.3	3.8	4.6	2.1
Gemini 3.1 Pro	4.2	3.8	4.5	1.2	4.5	4.6	4.3	2.8	4.6	4.4	4.8	4.5	4.4	4.5	4.3	3.8	4.5	4.6	4.3	2.5
Claude 4.6 Opus	4.2	3.9	4.0	3.5	4.5	4.3	4.2	3.2	4.4	4.1	4.5	3.0	3.6	3.3	3.4	1.8	4.0	3.5	4.6	1.5

T=Taste, S=Smell, X=Texture, U=Sound.

tortillas contribute a subtle smoky, earthy corn flavor. Blue corn tortilla chips visible in the background along with what appears to be refried black beans would add a salty, starchy complement to round out this vibrant, well-balanced meal.

Smell: The aroma would be led by the fresh, green, herbaceous scent of the raw jalapeño slices and green onions piled on top, which release sharp, peppery volatile compounds even before being eaten. The charred corn tortillas would contribute a warm, toasty, slightly smoky fragrance that serves as an earthy backdrop to the brighter toppings. The pico de gallo in the small white ramekin would emit a clean, garden-fresh scent of ripe tomato, raw onion, and likely cilantro and lime juice, creating an invigorating citrus-herbal note. The avocado itself is relatively mild in aroma but adds a subtle, fatty, slightly grassy undertone that blends with the overall fresh and lively olfactory character of the dish.

Texture: The textural complexity of these tacos is exceptional, beginning with the soft yet slightly chewy corn tortillas that show visible char blistering, giving them a pliable structure with occasional crisp spots along the scorched edges. The shredded cabbage slaw provides a crunchy, juicy, and springy bite that is the primary textural counterpoint to the softer ingredients, while the avocado slices contribute a buttery, smooth, almost silky mouthfeel that melts against the palate. The jalapeño rounds would add a firm, crisp, and watery snap with each bite, and the protein underneath appears shredded or pulled, suggesting a tender, flaky consistency. The pico de gallo on the side offers a chunky, juicy, and slightly pulpy texture that would add further variety when spooned onto each bite.

Sound: Biting into these tacos would produce a moderate initial crunch from the charred edges of the corn tortilla and the raw cabbage slaw, though the overall sound profile is relatively subdued given the predominance of soft, moist ingredients. The fresh jalapeño slices would contribute a clean, crisp snap when bitten through, briefly punctuating the quieter chewing of avocado and shredded protein beneath. Scooping pico de gallo with a blue corn chip from the side would generate a satisfying, sharp cracking sound as the chip breaks, which would be the loudest acoustic element of this meal. The overall eating experience leans toward quiet and soft, with intermittent crunchy moments from the slaw, jalapeños, and chips providing textural and auditory contrast.

LLaVA — Taste:3.5 Smell:3.0 Texture:4.0 Sound:3.0

Taste: The taco image suggests savory meat, creamy avocado, and tangy salsa with mild spice and freshness from green toppings.

Smell: Aroma is inferred as a mix of savory cooked filling, fresh salsa, and mild spicy notes.

Texture: Visual cues indicate soft tortilla and creamy avocado with crisp vegetables and tender filling for mixed bite texture.

Sound: Likely a moderate combination of soft chew and small crunch events from crispy elements.

Food-Llama 11B — Taste:3.0 Smell:2.5 Texture:4.5 Sound:2.0

Taste: The dish appears to offer familiar taco flavor balance: savory filling, creamy avocado, and zesty salsa in a soft tortilla.

Smell: Likely aromatic but moderate, combining fresh citrus-herb notes and a mild cooked-protein smell.

Texture: The output emphasizes strong texture diversity with soft, creamy, and crisp components in one bite.

Sound: Sound is expected to be relatively soft with occasional crunch from fresh/crisp toppings.