

Supplementary Material S3

Full per-image accuracy breakdown by model

Reference quality tiers: 1 = packet/label (highest), 2 = weighed/measured, 3 = portioned, 4 = visual estimate.

IMG-20260410-WA0016.jpg

Description: Breakfast burrito (two halves). User-measured carb count. **Reference:** 34 g (range [32, 36]), quality tier 2 (weighed/measured)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	40.1	38.6	2.6	6.6	7.7	6.1	+6.1	99.8	100.0
gemini-2.5-pro	500	86.7	85.0	20.2	23.3	107.1	52.7	+52.7	0.4	4.8
gemini-3.1-pro-preview	500	57.1	56.0	9.8	17.2	51.0	23.1	+23.1	10.8	39.0
gpt-5.4	560	67.9	70.4	4.3	6.3	18.4	33.9	+33.9	0.0	4.1

IMG-20260410-WA0017.jpg

Description: Neapolitan pizza capricciosa (whole pizza). Restaurant dish, no known carb count. Consistency test only. **Reference:** None (consistency analysis only)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	147.3	145.8	2.9	2.0	26.1	—	—	—	—
gemini-2.5-pro	500	129.0	127.8	13.0	10.1	67.5	—	—	—	—
gemini-3.1-pro-preview	500	118.6	117.0	9.8	8.3	64.0	—	—	—	—
gpt-5.4	560	139.9	145.6	11.8	8.4	40.8	—	—	—	—

IMG-20260410-WA0018.jpg

Description: Eggs benedict on two slices of sourdough with sweet potatoes and side salad. Restaurant dish, no known carb count. Consistency test only. **Reference:** None (consistency analysis only)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	76.7	76.2	2.5	3.2	12.3	–	–	–	–
gemini-2.5-pro	500	70.9	69.7	5.0	7.1	27.0	–	–	–	–
gemini-3.1-pro-preview	495	91.1	90.7	7.0	7.7	42.4	–	–	–	–
gpt-5.4	561	89.3	88.6	4.3	4.8	22.8	–	–	–	–

IMG-20260410-WA0019.jpg

Description: Crema catalana. Restaurant dessert, no known carb count. Consistency test only. **Reference:** None (consistency analysis only)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	29.9	29.2	0.8	2.8	4.5	–	–	–	–
gemini-2.5-pro	500	28.2	28.3	3.3	11.9	12.4	–	–	–	–
gemini-3.1-pro-preview	500	29.1	28.7	2.4	8.3	17.6	–	–	–	–
gpt-5.4	561	38.4	39.0	1.1	2.8	2.6	–	–	–	–

IMG-20260410-WA0020.jpg

Description: Paella with chicken and green beans (whole pan consumed). Restaurant dish, no known carb count. Consistency test only. **Reference:** None (consistency analysis only)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	117.1	101.8	34.2	29.2	135.6	–	–	–	–
gemini-2.5-pro	500	222.2	214.2	59.5	26.8	428.7	–	–	–	–

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
gemini-3.1-pro-preview	500	122.9	120.0	36.9	30.0	162.0	–	–	–	–
gpt-5.4	560	135.3	140.0	19.6	14.5	165.6	–	–	–	–

IMG_20200412_154249.jpg

Description: Roast beef dinner: 5 medium-small roast potatoes, tenderstem broccoli, roast beef slices. Shared peas/beans on the side. **Reference:** 47.5 g (range [45, 50]), quality tier 3 (portioned)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	58.4	61.2	8.8	15.1	38.0	11.5	+10.9	45.5	84.7
gemini-2.5-pro	500	48.7	47.5	10.2	21.0	61.6	8.2	+1.2	70.0	95.8
gemini-3.1-pro-preview	500	52.1	51.3	5.3	10.2	28.8	5.6	+4.6	88.0	99.6
gpt-5.4	560	54.9	56.2	4.7	8.6	25.9	7.8	+7.4	76.8	99.8

IMG_20250915_112359.jpg

Description: Bakewell tart slice with jam layer and white icing drizzle. **Reference:** 40 g (range [37, 43]), quality tier 2 (weighed/measured)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	50.9	51.1	0.9	1.8	3.3	10.9	+10.9	15.1	100.0
gemini-2.5-pro	500	49.1	49.3	3.5	7.0	29.7	9.1	+9.1	65.8	99.2
gemini-3.1-pro-preview	500	51.3	52.2	2.7	5.3	16.5	11.3	+11.3	30.0	100.0
gpt-5.4	559	32.6	31.9	2.6	7.9	14.7	7.6	-7.4	99.3	100.0

IMG_20260209_192318.jpg

Description: Tomato, chorizo and butter bean soup with 2 thick slices white bread.

Reference: 60 g (range [57, 63]), quality tier 1 (packet/label)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	47.0	47.8	0.9	1.9	3.9	13.0	-13.0	0.0	100.0
gemini-2.5-pro	500	57.7	57.0	5.1	8.8	35.8	4.3	-2.3	93.8	99.8
gemini-3.1-pro-preview	500	50.1	48.0	5.2	10.3	22.9	10.0	-9.9	42.0	100.0
gpt-5.4	559	61.3	67.2	6.2	10.0	16.1	6.2	+1.3	99.5	100.0

IMG_20260210_142258.jpg

Description: Cheese sandwich on 2 slices thick-cut white bread. **Reference:** 40 g

(range [38, 42]), quality tier 1 (packet/label)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	28.2	28.2	0.1	0.3	1.3	11.8	-11.8	0.0	100.0
gemini-2.5-pro	500	27.9	27.9	1.1	3.9	11.3	12.1	-12.1	1.8	100.0
gemini-3.1-pro-preview	500	28.3	27.7	2.0	6.9	12.0	11.7	-11.7	11.0	100.0
gpt-5.4	559	73.9	80.2	14.2	19.2	55.7	35.2	+33.9	7.2	12.0

MVIMG_20260222_204918.jpg

Description: Matcha churros with pistachio ice cream. Restaurant dessert.

Reference: 80 g (range [70, 90]), quality tier 4 (visual estimate)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	61.7	62.3	1.5	2.4	14.1	18.3	-18.3	0.0	91.6
gemini-2.5-pro	500	86.2	87.9	11.9	13.9	67.9	11.2	+6.2	45.8	91.4
	500	93.7	94.8	11.6	12.4	61.5	15.0	+13.7	33.8	65.6

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
gemini-3.1-pro-preview										
gpt-5.4	559	61.6	63.1	6.4	10.3	39.2	18.4	-18.4	6.4	62.1

MVIMG_20260303_200132.jpg

Description: Chilli con carne with mixed beans on white rice. **Reference:** 58 g (range [54, 62]), quality tier 3 (portioned)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	83.9	83.0	1.8	2.1	11.8	25.9	+25.9	0.0	1.8
gemini-2.5-pro	500	92.1	89.0	7.0	7.6	46.6	34.1	+34.1	0.0	0.2
gemini-3.1-pro-preview	500	86.0	85.3	9.1	10.6	47.7	28.0	+28.0	5.2	11.4
gpt-5.4	560	89.2	89.3	0.4	0.5	2.6	31.2	+31.2	0.0	0.0

MVIMG_20260304_120022.jpg

Description: Large bakery cookie (Battle Bakes and Cakes) with Mini Egg pieces and white chocolate. **Reference:** 47.5 g (range [45, 50]), quality tier 2 (weighed/measured)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	49.4	49.3	0.4	0.8	2.9	1.9	+1.9	100.0	100.0
gemini-2.5-pro	500	47.9	48.0	5.2	11.0	22.8	4.8	+0.4	99.2	100.0
gemini-3.1-pro-preview	500	44.0	45.5	7.7	17.6	27.4	6.3	-3.5	81.2	100.0
gpt-5.4	560	44.1	42.8	3.6	8.2	9.5	4.0	-3.4	100.0	100.0

MVIMG_20260308_142023.jpg

Description: Stuffed pork loin with extra bread stuffing and fried courgette slices.

Reference: 40 g (range [35, 45]), quality tier 3 (portioned)

Model	n	Mean (g)	Median (g)	SD (g)	CV (%)	Range (g)	MAE (g)	Bias (g)	Within 10g (%)	Within 20g (%)
claude-sonnet-4-6	510	44.8	44.8	1.3	2.8	8.6	4.8	+4.8	100.0	100.0
gemini-2.5-pro	500	33.7	20.0	23.0	68.3	118.6	22.8	-6.3	2.8	27.8
gemini-3.1-pro-preview	500	27.3	28.6	6.2	22.7	27.1	12.7	-12.7	39.6	80.8
gpt-5.4	561	77.4	82.5	14.7	19.0	61.7	37.5	+37.4	13.0	15.2