

gpt-4.1-mini

compliance	26	4	8	12	-8	-12	-15	9
cost	-6	-6	-7	-4	-10	-1	-20	-32
customer	-4	8	11	-20	21	-20	-1	-19
effort	-9	14	3	-29	-35	-26	-59	38
maintainability	-18	-1	13	-1	-24	-32	-16	-39
performance	-1	-2	-12	-15	7	-26	-8	-19
quality	-3	2	-11	-4	-18	-1	5	4
reliability	-15	5	5	-11	-2	45	-40	-15
scalability	-9	9	5	-20	-9	2	-52	-22
security	5	11	-9	-10	-11	-11	-26	-27
team	2	8	27	-6	-9	11	-2	31
timeline	0	8	-9	-3	-4	-10	47	-13
trendiness	9	2	44	-11	4	16	3	11
usability	-21	3	1	6	-19	-17	75	-2

gpt-4.1-nano

1	-8	40	7	-14	3	-9	-13
-8	3	-12	14	-3	0	4	-19
2	25	29	-24	46	-21	44	-23
-13	-5	19	9	-25	-11	-64	0
-10	-19	27	-15	-2	-6	-38	-19
3	-14	1	-14	14	-13	96	-20
3	-5	-1	3	-21	12	-7	12
3	-32	3	-11	-45	35	-65	-47
-5	13	-8	6	-32	-24	-100	-49
-13	-28	-10	12	-8	-2	-44	-28
5	1	27	4	7	-10	-39	4
5	28	12	-15	7	24	173	-5
10	7	53	-10	0	6	-17	14
-5	-13	-18	46	83	11	281	1

gpt-4o-mini

compliance	4	-8	-4	11	-17	47	-9	-8
cost	4	22	-20	-11	-20	3	-10	-20
customer	7	-13	6	-15	26	-67	37	-7
effort	-6	5	4	-10	-46	3	-14	21
maintainability	-8	-10	-27	-1	31	-130	2	-6
performance	3	-29	17	-6	6	-13	13	-18
quality	7	6	0	-1	-5	56	2	-3
reliability	11	-5	1	-2	-19	218	-26	-20
scalability	0	-1	-4	-8	-20	-31	-37	-9
security	-12	-12	4	-7	-36	-38	-3	-16
team	-3	1	30	4	23	-70	-32	34
timeline	2	18	3	-14	3	30	45	-28
trendiness	9	3	35	0	6	-3	-5	-7
usability	-6	-2	3	4	-7	-41	92	5

llama-3.3-70b-versatile

-7	-7	11	-10	-16	-26	-56	-10
-2	7	12	-8	0	-4	-77	-24
6	2	8	-29	10	-27	-56	1
-11	1	6	-2	-53	15	-199	-14
0	-16	2	10	-16	-29	-160	-12
-6	-14	4	-27	-16	-35	37	-19
0	7	-6	-4	-3	10	102	4
-1	-10	2	-19	-12	33	-116	-21
-5	-5	5	-13	-12	29	-176	-18
6	2	6	-15	-5	-17	-62	7
3	-2	17	-5	1	-10	-86	19
-5	-1	-6	-23	-11	56	121	-18
7	4	35	-5	-2	-1	-29	-5
-5	-1	-17	-15	-24	-26	155	-8

llama-3.1-8b-instant

compliance	2	-5	2	8	-11	-85	-26	-2
cost	4	-3	19	15	6	1	-21	-22
customer	-1	3	21	-16	-12	-116	14	4
effort	-6	21	-39	15	-112	-78	-63	-3
maintainability	17	17	-8	-8	-33	-2	-22	13
performance	6	-18	-5	-13	5	-23	28	5
quality	4	-1	6	2	-11	-70	7	2
reliability	8	-12	-11	0	-43	43	-6	-6
scalability	2	5	-24	29	-23	23	-36	-7
security	-11	4	2	-7	-52	-73	-22	4
team	-4	-2	52	-19	10	-95	-35	34
timeline	20	14	-12	-3	-7	15	93	-15
trendiness	14	-3	39	-8	4	-7	-10	0
usability	-1	28	-8	0	10	-55	70	-11

deepseek-r1-distill-llama-70b

-13	-2	3	2	-16	-29	-27	-7
-2	0	-17	2	-23	1	-64	-21
-2	4	37	-19	28	-24	56	-24
-19	-14	-21	-16	-114	-2	-102	4
-8	-13	-22	-18	9	-24	-77	-13
-1	-12	-2	-13	-36	-23	47	-34
5	-2	-6	7	-7	-32	-6	-1
-10	-33	-18	-37	-78	21	-84	-44
-15	-4	-15	-18	-52	-12	-106	-36
-6	-15	2	-12	9	-3	-14	1
-10	-15	35	1	3	-30	-83	18
-11	12	-3	-8	-10	36	82	-17
8	14	51	0	2	10	-10	-3
-13	-2	7	13	-13	-36	154	3

anchoring bias  
availability bias  
bandwagon effect  
confirmation bias  
framing effect  
hindsight bias  
hyperbolic discounting  
overconfidence bias

anchoring bias  
availability bias  
bandwagon effect  
confirmation bias  
framing effect  
hindsight bias  
hyperbolic discounting  
overconfidence bias