Enrichment Performance Analysis Results

This document summarizes retrieval performance across different enrichment subtypes, query strategies (Last Turn, Rewrite, Questions), and retrieval methods (ELSER, BM25, BGE).

Enrichment Categories and Subtypes

1. Question Type (10 Categories)

- **Factoid** (25.4%) - Asking for a specific piece of information, such as a date, quantity, name, yes/no answer, or other singular fact. It can be answered directly and concisely, and does not require an explanation, opinion, interpretation, or subjective judgment to answer.
- **Summarization** (19.1%) - Asking to summarize a process or a policy.
- **Explanation** (14.7%) - Explain the reason behind something.
- **Opinion** (8.0%) - Asking model's opinion on something. The question could also be phrased as leading (i.e., suggesting a particular answer, such as "Don't you think X is better?").
- **How-To** (8.0%) - Instructions describing how to perform a task.
- **Non-Question** (6.9%) - Not asking a question but instead answering a question or providing information asked by the model.
- **Keyword** (7.3%) - Asking using keywords (not full sentence/phrase). This may be ambiguous.
- **Composite** (4.9%) - Comprises several questions. They may be related or dependent.
- **Comparative** (4.4%) - Asking for comparison. This can be comparison (a) of multiple entities/concepts, (b) of characteristics of a single entity, or (c) comparison with decision.
- **Troubleshooting** (1.5%) - Finding solutions to issues, problems, challenges.

---

2. Multi-Turn Type (3 Categories)

- **Follow-up** (73.9%) - Ask a question that requests more information or related information to continue the conversation.
- **N/A** (13.1%) - First turn of conversation (no previous context). Initial question in a conversation.
- **Clarification** (13.1%) - (a) Write a statement clarifying the user's intent (typically used when the agent misinterpreted one of the prior questions). (b) Ask a question to clarify the model's previous answer. Clarifications are typically asked when something is unclear or hard to understand.

---

3. Answerability (4 Categories)

- **Answerable** (91.2%) - The question can be fully answered from the passages.
- **Partially Answerable** (8.8%) - Only part of the question can be answered from the passages.
- **Unanswerable** (6.5%) - The question cannot be answered neither fully nor partially from the passages.
- **Conversational** (1.2%) - The user turn does not contain a question but is a conversational statement (e.g., "Hello", "Hi, I had a question", "Cool", "That's interesting", "That was all", "Thank you").

The analysis compares:

- **Query Strategies**:
  - **Last Turn**: Uses only the current question (no context)
  - **Rewrite**: Uses LLM-rewritten question with full context
  - **Questions**: Uses original questions (No Agent Response)
- **Retrieval Methods**:
  - **ELSER**: Learned Sparse
  - **BM25**: Lexical
  - **BGE**: Dense

Question Types Performance

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
Comparative	44	4.4%	ELSER	Lastturn	0.155	0.354	0.461	0.559	0.386	0.390	0.425	0.466
			ELSER	**Rewrite**	**0.193**	**0.389**	**0.527**	**0.626**	**0.500**	**0.429**	**0.487**	**0.535**
			ELSER	Questions	0.085	0.225	0.340	0.435	0.250	0.242	0.292	0.333
			BM25	Lastturn	0.045	0.155	0.195	0.313	0.114	0.152	0.170	0.220
			BM25	Rewrite	0.076	0.197	0.272	0.374	0.205	0.202	0.240	0.282
			BM25	Questions	0.089	0.178	0.231	0.334	0.205	0.184	0.212	0.256
			BGE	Lastturn	0.096	0.188	0.283	0.379	0.250	0.226	0.267	0.307
			BGE	Rewrite	0.107	0.269	0.403	0.520	0.295	0.306	0.360	0.408
			BGE	Questions	0.051	0.129	0.229	0.299	0.136	0.140	0.192	0.222
					—	—	—	—	—	—	—	—
Composite	49	4.9%	ELSER	Lastturn	0.189	0.407	0.505	0.646	0.490	0.464	0.486	0.543
			ELSER	**Rewrite**	**0.211**	**0.461**	**0.573**	**0.750**	**0.531**	**0.514**	**0.548**	**0.622**
			ELSER	Questions	0.082	0.241	0.314	0.458	0.224	0.264	0.289	0.346

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
			BM25	Lastturn	0.105	0.180	0.265	0.376	0.286	0.226	0.258	0.301
			BM25	Rewrite	0.083	0.184	0.265	0.410	0.245	0.215	0.244	0.303
			BM25	Questions	0.070	0.134	0.196	0.318	0.204	0.170	0.191	0.242
			BGE	Lastturn	0.187	0.337	0.415	0.547	0.490	0.402	0.420	0.475
			BGE	Rewrite	0.201	0.369	0.456	0.595	0.510	0.428	0.455	0.515
			BGE	Questions	0.078	0.189	0.264	0.379	0.204	0.213	0.245	0.295
					—	—	—	—	—	—	—	—
Explanation	148	14.7%	ELSER	Lastturn	0.186	0.368	0.451	0.575	0.392	0.377	0.411	0.466
			ELSER	**Rewrite**	**0.187**	**0.394**	**0.492**	**0.616**	**0.439**	**0.411**	**0.450**	**0.505**
			ELSER	Questions	0.102	0.195	0.236	0.341	0.223	0.206	0.220	0.265
			BM25	Lastturn	0.045	0.161	0.206	0.317	0.135	0.157	0.175	0.221
			BM25	Rewrite	0.046	0.169	0.236	0.356	0.135	0.165	0.191	0.240
			BM25	Questions	0.045	0.142	0.189	0.246	0.142	0.144	0.161	0.185
			BGE	Lastturn	0.136	0.277	0.330	0.450	0.304	0.287	0.307	0.358
			BGE	Rewrite	0.160	0.308	0.386	0.508	0.358	0.325	0.356	0.410
			BGE	Questions	0.073	0.141	0.201	0.265	0.182	0.152	0.178	0.207
					—	—	—	—	—	—	—	—
Factoid	255	25.4%	ELSER	Lastturn	0.170	0.320	0.414	0.515	0.388	0.352	0.387	0.430
			ELSER	**Rewrite**	**0.182**	**0.360**	**0.465**	**0.599**	**0.427**	**0.390**	**0.430**	**0.486**
			ELSER	Questions	0.088	0.212	0.280	0.369	0.200	0.213	0.242	0.279
			BM25	Lastturn	0.081	0.163	0.212	0.276	0.184	0.176	0.193	0.220
			BM25	Rewrite	0.085	0.191	0.253	0.334	0.200	0.200	0.222	0.255
			BM25	Questions	0.065	0.152	0.194	0.255	0.153	0.159	0.173	0.199
			BGE	Lastturn	0.127	0.261	0.316	0.401	0.294	0.279	0.297	0.332
			BGE	Rewrite	0.163	0.321	0.385	0.482	0.384	0.347	0.368	0.407
			BGE	Questions	0.066	0.185	0.229	0.317	0.157	0.183	0.201	0.236
					—	—	—	—	—	—	—	—
How-To	80	8.0%	ELSER	Lastturn	0.175	0.382	0.442	0.567	0.400	0.401	0.415	0.469
			ELSER	**Rewrite**	**0.191**	**0.407**	**0.468**	**0.619**	**0.425**	**0.424**	**0.441**	**0.506**
			ELSER	Questions	0.105	0.240	0.311	0.394	0.237	0.247	0.275	0.309
			BM25	Lastturn	0.074	0.171	0.226	0.271	0.150	0.174	0.192	0.210
			BM25	Rewrite	0.081	0.192	0.242	0.333	0.163	0.191	0.209	0.244
			BM25	Questions	0.055	0.128	0.173	0.231	0.138	0.140	0.151	0.175
			BGE	Lastturn	0.130	0.274	0.357	0.473	0.325	0.293	0.320	0.369
			BGE	Rewrite	0.170	0.301	0.379	0.499	0.375	0.322	0.352	0.405
			BGE	Questions	0.063	0.169	0.221	0.281	0.150	0.168	0.186	0.213
					—	—	—	—	—	—	—	—

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
Keyword	73	7.3%	**ELSER**	**Lastturn**	**0.173**	**0.345**	**0.432**	**0.545**	**0.411**	**0.376**	**0.401**	**0.454**
			ELSER	Rewrite	0.159	0.315	0.395	0.546	0.342	0.333	0.359	0.425
			ELSER	Questions	0.052	0.106	0.161	0.206	0.110	0.109	0.132	0.152
			BM25	Lastturn	0.111	0.181	0.224	0.291	0.247	0.211	0.223	0.251
			BM25	Rewrite	0.096	0.171	0.237	0.346	0.192	0.184	0.208	0.252
			BM25	Questions	0.042	0.103	0.151	0.172	0.096	0.101	0.122	0.131
			BGE	Lastturn	0.091	0.172	0.226	0.315	0.192	0.181	0.201	0.243
			BGE	Rewrite	0.115	0.198	0.282	0.395	0.274	0.221	0.255	0.303
			BGE	Questions	0.042	0.067	0.115	0.176	0.082	0.071	0.092	0.119
					—	—	—	—	—	—	—	—
Non-Question	69	6.9%	ELSER	Lastturn	0.136	0.289	0.386	0.466	0.275	0.295	0.338	0.374
			ELSER	**Rewrite**	**0.162**	**0.335**	**0.425**	**0.548**	**0.348**	**0.343**	**0.383**	**0.440**
			ELSER	Questions	0.046	0.153	0.210	0.289	0.101	0.144	0.171	0.204
			BM25	Lastturn	0.098	0.161	0.190	0.249	0.203	0.172	0.183	0.208
			BM25	Rewrite	0.120	0.217	0.256	0.316	0.261	0.228	0.243	0.268
			BM25	Questions	0.041	0.118	0.163	0.243	0.101	0.110	0.130	0.163
			BGE	Lastturn	0.115	0.232	0.286	0.379	0.246	0.238	0.258	0.300
			BGE	Rewrite	0.129	0.289	0.359	0.487	0.319	0.296	0.322	0.376
			BGE	Questions	0.046	0.128	0.167	0.210	0.130	0.122	0.141	0.158
					—	—	—	—	—	—	—	—
Opinion	80	8.0%	ELSER	Lastturn	0.156	0.313	0.408	0.486	0.362	0.347	0.381	0.417
			ELSER	**Rewrite**	**0.187**	**0.346**	**0.462**	**0.578**	**0.450**	**0.395**	**0.439**	**0.489**
			ELSER	Questions	0.099	0.190	0.253	0.312	0.225	0.198	0.228	0.253
			BM25	Lastturn	0.097	0.144	0.204	0.252	0.212	0.164	0.190	0.210
			BM25	Rewrite	0.122	0.189	0.242	0.295	0.263	0.209	0.231	0.253
			BM25	Questions	0.053	0.114	0.157	0.215	0.138	0.120	0.143	0.166
			BGE	Lastturn	0.130	0.246	0.307	0.410	0.312	0.273	0.294	0.340
			BGE	Rewrite	0.140	0.302	0.358	0.456	0.338	0.321	0.338	0.383
			BGE	Questions	0.084	0.151	0.221	0.297	0.188	0.158	0.196	0.226
					—	—	—	—	—	—	—	—
Summarization	192	19.1%	ELSER	Lastturn	0.147	0.325	0.416	0.530	0.365	0.363	0.388	0.437
			ELSER	**Rewrite**	**0.175**	**0.349**	**0.455**	**0.587**	**0.438**	**0.398**	**0.429**	**0.487**
			ELSER	Questions	0.102	0.191	0.259	0.325	0.250	0.215	0.240	0.270
			BM25	Lastturn	0.075	0.141	0.187	0.306	0.177	0.162	0.176	0.224
			BM25	Rewrite	0.092	0.166	0.216	0.334	0.229	0.192	0.208	0.257
			BM25	Questions	0.062	0.126	0.168	0.235	0.167	0.141	0.155	0.184
			BGE	Lastturn	0.124	0.241	0.316	0.433	0.307	0.273	0.299	0.347

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
			BGE	Rewrite	0.149	0.280	0.357	0.508	0.385	0.322	0.346	0.411
			BGE	Questions	0.055	0.141	0.183	0.265	0.146	0.147	0.165	0.199
					—	—	—	—	—	—	—	—
Troubleshooting	15	1.5%	**ELSER**	**Lastturn**	**0.383**	**0.639**	**0.706**	**0.706**	**0.600**	**0.627**	**0.654**	**0.654**
			ELSER	Rewrite	0.278	0.572	0.611	0.739	0.400	0.520	0.540	0.595
			ELSER	Questions	0.144	0.289	0.333	0.467	0.267	0.273	0.300	0.345
			BM25	Lastturn	0.239	0.356	0.411	0.433	0.333	0.341	0.368	0.378
			BM25	Rewrite	0.367	0.478	0.578	0.678	0.533	0.475	0.530	0.572
			BM25	Questions	0.111	0.178	0.267	0.300	0.200	0.171	0.209	0.221
			BGE	Lastturn	0.239	0.428	0.472	0.494	0.333	0.406	0.419	0.430
			BGE	Rewrite	0.194	0.483	0.517	0.539	0.333	0.434	0.438	0.450
			BGE	Questions	0.111	0.211	0.256	0.256	0.200	0.192	0.217	0.217

Multi Turn Performance

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
Clarification	101	13.0%	ELSER	Lastturn	0.102	0.223	0.308	0.410	0.248	0.241	0.272	0.316
			ELSER	**Rewrite**	**0.117**	**0.277**	**0.377**	**0.507**	**0.297**	**0.298**	**0.335**	**0.391**
			ELSER	Questions	0.045	0.137	0.197	0.278	0.109	0.129	0.160	0.194
			BM25	Lastturn	0.037	0.096	0.116	0.165	0.099	0.107	0.110	0.130
			BM25	Rewrite	0.066	0.156	0.211	0.315	0.149	0.166	0.185	0.226
			BM25	Questions	0.035	0.107	0.136	0.190	0.099	0.106	0.118	0.140
			BGE	Lastturn	0.089	0.167	0.225	0.298	0.188	0.176	0.202	0.233
			BGE	Rewrite	0.121	0.259	0.318	0.417	0.287	0.272	0.292	0.334
			BGE	Questions	0.054	0.119	0.165	0.203	0.129	0.117	0.141	0.157
					—	—	—	—	—	—	—	—
Follow-up	574	73.9%	ELSER	Lastturn	0.161	0.328	0.409	0.514	0.368	0.354	0.381	0.427
			ELSER	**Rewrite**	**0.178**	**0.350**	**0.446**	**0.582**	**0.413**	**0.379**	**0.413**	**0.473**
			ELSER	Questions	0.062	0.143	0.198	0.283	0.148	0.150	0.172	0.208
			BM25	Lastturn	0.079	0.154	0.203	0.287	0.183	0.169	0.187	0.222
			BM25	Rewrite	0.086	0.174	0.233	0.328	0.207	0.188	0.210	0.249
			BM25	Questions	0.054	0.115	0.158	0.217	0.141	0.126	0.142	0.168
			BGE	Lastturn	0.122	0.240	0.305	0.411	0.289	0.262	0.285	0.330
			BGE	Rewrite	0.148	0.285	0.363	0.485	0.352	0.312	0.340	0.392

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
			BGE	Questions	0.039	0.105	0.160	0.232	0.101	0.108	0.133	0.163
					—	—	—	—	—	—	—	—
N/A	102	13.1%	**ELSER**	**Lastturn**	**0.307**	**0.591**	**0.744**	**0.851**	**0.647**	**0.612**	**0.677**	**0.723**
			ELSER	Rewrite	0.307	0.591	0.744	0.851	0.647	0.612	0.677	0.723
			ELSER	Questions	0.307	0.591	0.744	0.851	0.647	0.612	0.677	0.723
			BM25	Lastturn	0.142	0.308	0.384	0.498	0.294	0.301	0.335	0.382
			BM25	Rewrite	0.142	0.308	0.384	0.498	0.294	0.301	0.335	0.382
			BM25	Questions	0.142	0.308	0.384	0.498	0.294	0.301	0.335	0.382
			BGE	Lastturn	0.213	0.461	0.547	0.651	0.461	0.457	0.494	0.537
			BGE	Rewrite	0.213	0.461	0.547	0.651	0.461	0.457	0.494	0.537
			BGE	Questions	0.213	0.461	0.547	0.651	0.461	0.457	0.494	0.537

Answerability Performance

Subtype	Count	%	Retrieval Method	Strategy	R@1	R@3	R@5	R@10	nDCG@1	nDCG@3	nDCG@5	nDCG@10
Answerable	709	91.2%	ELSER	Lastturn	0.177	0.358	0.447	0.553	0.401	0.384	0.415	0.462
			ELSER	**Rewrite**	**0.189**	**0.378**	**0.480**	**0.613**	**0.433**	**0.406**	**0.443**	**0.502**
			ELSER	Questions	0.088	0.195	0.263	0.349	0.205	0.203	0.232	0.269
			BM25	Lastturn	0.081	0.169	0.219	0.305	0.189	0.181	0.199	0.235
			BM25	Rewrite	0.088	0.187	0.245	0.349	0.209	0.198	0.219	0.262
			BM25	Questions	0.061	0.136	0.182	0.249	0.154	0.144	0.162	0.191
			BGE	Lastturn	0.135	0.268	0.337	0.442	0.310	0.287	0.312	0.357
			BGE	Rewrite	0.162	0.316	0.392	0.512	0.375	0.339	0.367	0.418
			BGE	Questions	0.066	0.154	0.212	0.282	0.155	0.157	0.184	0.214
					—	—	—	—	—	—	—	—
Partial	68	8.8%	ELSER	Lastturn	0.130	0.251	0.362	0.452	0.265	0.255	0.303	0.340
			ELSER	**Rewrite**	**0.169**	**0.308**	**0.436**	**0.556**	**0.382**	**0.330**	**0.379**	**0.430**
			ELSER	Questions	0.127	0.265	0.334	0.436	0.250	0.257	0.286	0.326
			BM25	Lastturn	0.083	0.141	0.179	0.242	0.162	0.149	0.164	0.189
			BM25	Rewrite	0.113	0.214	0.302	0.349	0.235	0.221	0.260	0.278
			BM25	Questions	0.082	0.170	0.213	0.264	0.176	0.171	0.187	0.208
			BGE	Lastturn	0.074	0.171	0.216	0.274	0.176	0.170	0.187	0.209
			BGE	Rewrite	0.061	0.190	0.273	0.355	0.176	0.188	0.220	0.253
			BGE	Questions	0.045	0.150	0.203	0.292	0.118	0.134	0.158	0.191

Summary of Each Enrichment Subtype

Question Types

Comparative (44 queries)

Questions that compare two or more entities, concepts, or options. Best performance with ELSER-Rewrite (R@5: 0.527). Shows significant improvement with context rewriting (+14.3% over Last Turn), suggesting these queries benefit from explicit comparison framing.

Composite (49 queries)

Multi-part questions requiring synthesis of multiple pieces of information. Strong performer (R@5: 0.573 with ELSER-Rewrite). High scores indicate the retrieval system handles complex, multi-faceted queries well.

Explanation (148 queries)

Questions seeking detailed explanations of concepts, processes, or phenomena. Moderate performance (R@5: 0.492 with ELSER-Rewrite). Largest category in question types, showing consistent but not exceptional retrieval accuracy.

Factoid (255 queries)

Simple fact-seeking questions with specific, concrete answers. Largest category overall. Moderate performance (R@5: 0.465 with ELSER-Rewrite), suggesting room for improvement in pinpointing specific factual information.

How-To (80 queries)

Procedural questions asking for step-by-step instructions or methods. Good performance (R@5: 0.468 with ELSER-Rewrite). Benefits moderately from context rewriting.

Keyword (73 queries)

Queries consisting primarily of keywords rather than full questions. **Worst overall performer** (R@5: 0.395 with ELSER-Rewrite). All retrieval methods struggle, with Questions strategy particularly poor (R@5: 0.161). Indicates keyword-based queries need different handling strategies.

Non-Question (69 queries)

Statements or phrases that aren't formatted as questions but imply information needs. Moderate performance (R@5: 0.425 with ELSER-Rewrite). Shows strong benefit from rewriting (+10.1%), suggesting context helps convert implicit queries into retrievable formats.

Opinion (80 queries)

Questions seeking subjective views, recommendations, or perspectives. Moderate performance (R@5: 0.462 with ELSER-Rewrite). Shows notable improvement with rewriting (+13.2%), indicating opinion queries benefit from contextual clarification.

Summarization (192 queries)

Requests for summaries or overviews of topics. Second-largest category. Moderate performance (R@5: 0.455 with ELSER-Rewrite). Consistent across strategies, suggesting summarization requests are relatively well-handled.

Troubleshooting (15 queries)

Problem-solving queries about issues or errors. **Best overall performer** (R@5: 0.706 with ELSER-Lastturn). Surprisingly, Last Turn outperforms Rewrite, suggesting the immediate context is most relevant. Small sample size (15 queries) may affect reliability.

Multi-Turn Types

Clarification (101 queries)

Follow-up questions that seek clarification on previous responses. **Weakest multi-turn performer** (R@5: 0.377 with ELSER-Rewrite). Shows the challenge of maintaining context across turns when users need additional explanation.

Follow-up (574 queries)

Continuation questions building on previous conversation context. Largest multi-turn category. Moderate performance (R@5: 0.446 with ELSER-Rewrite). Shows consistent improvement with rewriting (+9%), indicating context enrichment helps continuation queries.

N/A (102 queries)

Queries that don't require multi-turn context (essentially single-turn within multi-turn conversations). **Highest performance** (R@5: 0.744, identical across all strategies). Perfect scores across strategies indicate these are independent queries that don't benefit from conversation history.

Answerability Types

Answerable (709 queries)

Questions that can be fully answered with available information. Largest answerability category. Solid performance (R@5: 0.480 with ELSER-Rewrite), representing the baseline for well-formed, answerable queries.

Partial (68 queries)

Questions that can only be partially answered with available information. Lower performance (R@5: 0.436 with ELSER-Rewrite) compared to Answerable. Shows larger improvement with Rewrite strategy (+20.4% vs +7.4%), suggesting context helps bridge information gaps.

High-Level Takeaways

1. **ELSER Dominance:** Across almost every category, **ELSER** (Elastic Learned Sparse Encoder) consistently outperforms both BM25 (lexical) and BGE (dense embedding). This suggests that learned sparse semantic encoding is currently the most effective retrieval method for this dataset.
2. **The "Rewrite" Strategy is Critical:** For most question types, using an LLM to **rewrite** the query with conversation context yields the best results.
   - *Why:* It explicitly resolves coreferences (e.g., changing "it" to "the server") and adds necessary context that is missing from the raw "Last Turn" query.
   - *Magnitude:* You see significant gains in types like **Comparative** (+14.3% improvement over Last Turn) and **Opinion** (+13.2%), which likely require more nuance than a simple keyword match.
3. **The "Keyword" & "Clarification" Struggle:**
   - **Keyword** queries are the worst performers (R@5 0.395). This indicates that when users search with just 1-2 words, the system struggles to infer intent, regardless of the retrieval method.
     - *Example:* In task `1065ea5ad1ae2b90e6fce67d851a7a66<::>2` (Query: "carthage?"), Last Turn achieved a relevance score of 12.92, while the Rewrite relevance score was 12.57. Both were significantly lower than the score achieved with a well-formed natural question (18.77), demonstrating that short keyword queries are often ambiguous even when rewritten. (Note: ELSER scores are unbounded; typical strong matches are >20).
   - **Clarification** questions (in multi-turn) are the weakest multi-turn type (R@5 0.377). These are short queries like "Why?" or "What about X?" that are hard to rewrite perfectly without retrieving the wrong context.
     - *Example:* In task `e52ab8d5f61ccdfc3712a2608d8c2aba<::>10`, the Rewrite strategy achieved a score of 11.62, which was only marginally better than Last Turn (9.86) and far below the score of the original self-contained question (23.84). This suggests that for clarification, the rewritten query often fails to capture the full semantic intent needed for high-precision retrieval.
4. **Troubleshooting prefers "Last Turn":**
   - For **Troubleshooting** queries, the "Last Turn" strategy actually *outperforms* the "Rewrite" strategy (R@5 0.706 vs 0.611).
   - *Hypothesis:* Troubleshooting queries often contain specific error codes or log messages. Rewriting might "hallucinate" or dilute these specific tokens, making exact match (or near-exact match) less effective. However, the sample size (15) is very small, so this might be noise.
     - *Example:* In task `1065ea5ad1ae2b90e6fce67d851a7a66<::>1` (Query: "Should we switch to a different method…"), Last Turn and Rewrite performed identically (Score ~17.10), suggesting that when the user is already specific, rewriting adds no value and might introduce risk.