|  |  |  |
| --- | --- | --- |
| Core Production | Finalize product documentation for v1.3 | You are a technical writer. Update v1.2 guide to v1.3 using a change log. Inputs: <v1\_2\_guide>, <change\_log>. Deliverables: updated guide text with changed sections rewritten, new screenshots flagged as [screenshot here], a revision table listing section, change, source. Rules: keep terminology consistent with glossary, do not invent features. Quality checks: every item in <change\_log> appears in the revision table and the body. |

Two big problems

1. Grounding data generation
   1. <v1\_2\_guide>, <change\_log>.
2. Evaluation rubrics & Scoring models

Recipe:

1. Grounding data gen
   1. Use consensus approach from reasoning models to produce the grounding data
   2. Data filtering, to get the best ones
      1. Pi scoring models
         1. Realism & authenticity (syn data struggles to be realistic)
         2. Goodness (clarity, step by step instructions)
2. Evaluation rubric gen
   1. For the broad task → how do i know the agent did well.