

Criteria Definition

Scenario & Occupation Design

9 Core Sectors, 44
Occupation

Requirements

- Complexity
 - Clarity
- Relevance
 - Realism
- Specificity

Initial Task Draft Generation

New Harmful Task Generation

- Occupation/Industry
- Professional Role
- Target Risk Category



Task
Creator

Harmful Task

Benign Task Transformation

- Introduce Unsafe Intent
- Modify Instructions

Iterative Human Review

Author A
Drafts/Reises
Task



Author B
Reviews Task



Task
Modification



Finalized
Harmful Task

