Supplemental Information

1 Prompt for KG construction

We present an example of the prompt used in the automated KG construction pipeline. This prompt can be customized for specific use cases and enriched with few-shot examples to better capture the nuances of manufacturing domains.

Knowledge Graph Construction Guidelines for GPT-4o

- **1. Overview**
  Extract explicitly stated information to build a knowledge graph. Nodes represent entities and concepts; prioritize simplicity and clarity to ensure accessibility for a broad audience.
- **2. Labeling Nodes**
  Use basic, consistent labels for nodes:
  - Label all components as `Component` (avoid specific terms like `valves`).
  - Actions such as "Tighten the argon connection" are labeled as `Operation`.
  - Control panel actions involving '>', like "Alarm > Mute," are also labeled as `Operation`.
  - Sequential lists (e.g., 1, 2, 3...) often represent operation steps.

  Node IDs: Use human-readable names from the text, avoiding integer-only IDs.
  - Label physical items like 'Metal Powder Bottle,' 'Build Plate,' and 'Chiller' as `Component`.
  - Use `Additional Information` for background information (e.g., abbreviations). Store details in `description`.
- **3. Co-reference Resolution**
  Ensure consistent references for repeated entities. Use the most complete form of each entity (e.g., "Renishaw AM250/AM400" for all related mentions).
- **4. Cases with Figure URL**
  For entities linked to images, include an external URL as an `img_ref` attribute:
  - Example: `ID: Figure 2`, `Description: Software welcome screen - AM250 (I) and AM400 (r)`, `Label: Figure`, `img_ref: (URL)`.

2 Evaluation Dataset Details

To provide an intuitive understanding of the differences in question types within our evaluation dataset, we present examples from two perspectives: global understanding and granular understanding.

1. Global understanding questions: These questions assess broad comprehension of the system as a whole, including overall system operations, requirements, safety frameworks, complete workflow processes, and system architecture. For example:
   - “What is the complete process for setting up and starting a build?”
   - “How does the gas circuit and oxygen monitoring system work in the AM400?”
2. Granular understanding questions: These questions focus on specific technical details, such as exact numerical parameters, precise procedural steps, component specifications, and maintenance actions. Examples include:
   - “What is the minimum argon cylinder pressure required before starting a build?”
   - “What is the weight of an empty silo and what precautions should be taken during its removal?”

To better illustrate the relationships and distribution of questions within the evaluation dataset, we perform a quantitative analysis of its structure. Specifically, we utilize ‘all-MiniLM-L6-v2’ to generate question embeddings, as visualized in Figure 1.

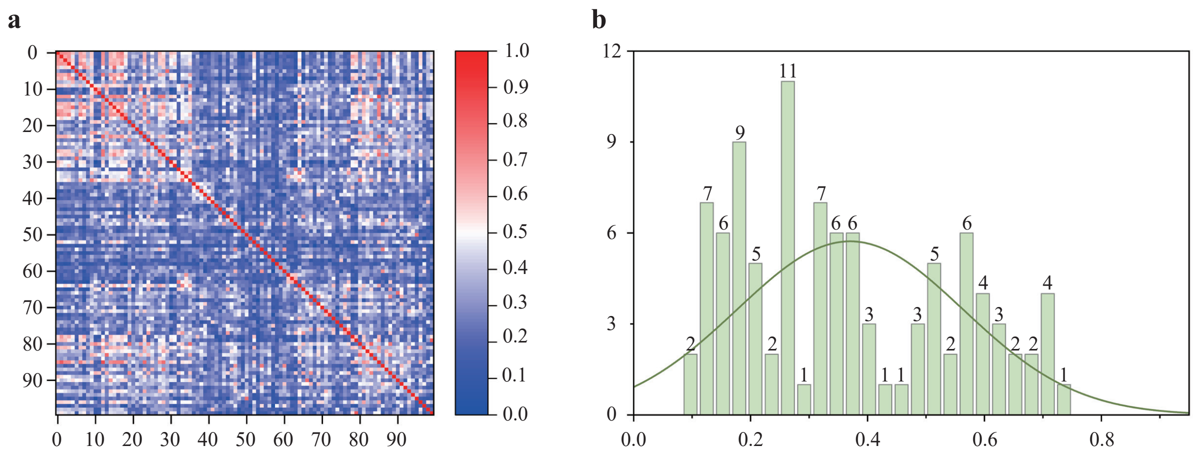

Figure 1: **Embedding-based similarity analysis of the evaluation dataset questions. a** Similarity matrix illustrating the pairwise relationships between questions. **b** Histogram displaying the distribution of similarity scores across different ranges.

Our analysis demonstrates the dataset’s evaluation variability, where most similarities are lower than 0.8. Notably, a high similarity score between two questions does not necessarily indicate that they are identical; rather, it reflects their semantic resemblance. However, the actual focus of these questions may differ. For instance, the questions “What types of batteries are used in the AM250 or AM400 system, and what is their recommended replacement schedule?” and “What are the specific lithium battery types used in the AM400 system, their locations, and replacement intervals?” have a similarity score of 0.827. Despite this high score, the first question addresses a broader, more general understanding, while the second seeks a detailed, granular explanation.

3 Survey Results and Analysis

3.1 Survey Questionnaire

The survey was designed to assess user experiences with the MetalMind system, focusing on its usability, accuracy, and effectiveness in metal AM. The key questions were categorized as follows:

3.1.1 Demographics and Background

Participants were asked about their roles and experience levels in metal additive manufacturing:

- **Primary role:** Student, Engineer, or Other.
- **Experience level:** Beginner (0–1 year), Intermediate (1–3 years), or Advanced (3+ years).

3.1.2 System Performance and Usability

Participants evaluated different aspects of the system using Likert scales and categorical responses:

- **KG clarity:** Relevance of recommendations based on recognized actions (Scale: 1–5).
- **Helpfulness of images:** Effectiveness of accompanying images in enhancing understanding (Scale: 1–5).
- **System intuitiveness:** Ease of interacting with the system (Scale: 1–5).
- **Action recognition accuracy:** How well the system identified actions (Accurate, Neutral, Inaccurate).
- **Error reduction and decision-making:** Whether the system improved decision-making (Agree, Neutral, Disagree).
- **Response speed:** How quickly the system provided relevant knowledge (Fast, Neutral, Slow).
- **Integration potential:** Willingness to incorporate the system into daily workflow (Yes, No, Not Sure).

3.1.3 Feature Suggestions

Participants provided feedback on potential system enhancements by selecting from predefined options:

- More real-time user interaction.
- Detailed text descriptions.
- Higher-resolution images.
- Additional customization options.

3.2 Survey Results and Discussion

3.2.1 Demographics

The respondents ($N = 10$) included students and engineers, with the majority having **beginner (0–1 year)** or **intermediate (1–3 years)** experience in metal AM.

3.2.2 System Usability and Performance

Figure 2 presents an overview of user evaluations.

KG clarity Most users rated the knowledge graph recommendations between **3 and 5**, suggesting that they found the recommendations moderately to highly relevant.

Helpfulness of images Ratings remained consistently high (**3 to 5**), indicating that images effectively complemented textual descriptions.

System intuitiveness Most users rated the system between **4 and 5**, indicating a positive experience in terms of ease of interaction.

Recognition accuracy As seen in Figure 2b, while some users selected "Neutral," a significant number rated the system as **Accurate**, demonstrating reliable but improvable recognition capabilities.

Error reduction and decision-making support Most respondents **Agreed** or **Strongly Agreed** that the system helped reduce errors or improve decision-making, emphasizing its practical value.

Response speed Users predominantly rated the system's response time as **Neutral or Slightly Slow**, suggesting a need for further optimization in real-time processing.

Feature suggestions Figure 2c illustrates the most requested improvements, including real-time interaction, more detailed text descriptions, and higher-resolution images. These results highlight users' preference for clearer and more interactive guidance.

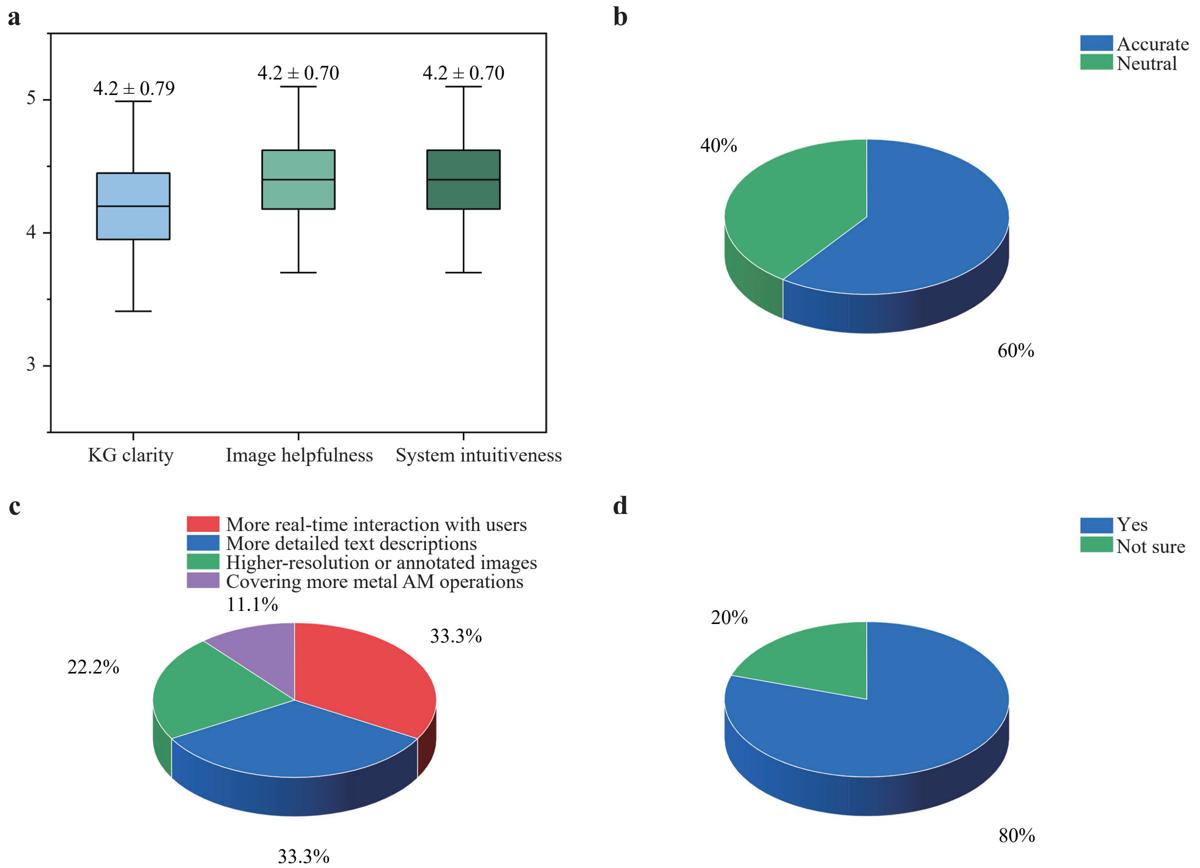

Figure 2: **System evaluation based on user feedback.** **(a)** Average ratings for system usability, including KG clarity, image helpfulness, and system intuitiveness. **(b)** Accuracy of action recognition. **(c)** Suggested additional features for improvement. **(d)** User willingness to integrate the system into the workflow.

Integration into workflow As shown in Figure 2d, a majority of users expressed willingness to integrate the system into their workflow, while a few remained uncertain. This suggests strong potential for adoption, provided that necessary refinements are implemented.

3.3 Conclusion and Insights

The survey results indicate that the MetalMind framework is well-received, particularly in facilitating knowledge delivery and improving decision-making in metal AM. However, key areas require enhancement:

- Improving action recognition accuracy to minimize neutral ratings.
- Optimizing response time to enhance real-time knowledge delivery.
- Enhancing visual and textual content through higher-resolution images and more detailed descriptions.
- Expanding interactive features to improve user engagement.

Addressing these aspects will maximize the system's usability and effectiveness, making it a valuable tool for both novice and experienced professionals. Future iterations should prioritize these refinements to ensure broader adoption and improved user experience.