# LLM 自动化评估结果

### 功能需求评估表

以下评估基于提供的需求文档进行。我首先识别了系统中的数据实体，然后对功能需求进行了合并（剔除系统通用设置类需求），随后依次评估了幻觉需求、完整引用和可测试性。评估原则严格遵循您的指令：

- \*\*数据实体识别\*\*：仅统计系统内核心管理的数据对象（如库存、订单等），不包括外部接口（如API或外部数据源）。基于需求文档，系统涉及的数据实体包括：

- \*\*Inventory\*\*（库存）：从库存管理需求（INVENTORY-001, INVENTORY-002, INVENTORY-003）中提取，代表产品库存数据。

- \*\*Order\*\*（订单）：从订单处理需求（ORDER-001, ORDER-002, ORDER-003, ORDER-004）中提取，代表客户订单数据。

- \*\*Report\*\*（报告）：从报告和分析需求（REPORT-001, REPORT-002, REPORT-003）中提取，代表库存、销售等分析报告数据。

- \*\*Widget\*\*（小组件）：从用户界面需求（UI-003）中提取，代表用户可定制的仪表板元素（如库存、订单、分析小组件），它作为独立数据实体管理用户的个性化视图。

- 总计数据实体个数：\*\*4个\*\*（Inventory、Order、Report、Widget）。注意：UI-001中的“状态指示器”是视图组件而非数据实体；外部约束（如多语言支持）不计入。

- \*\*功能需求合并\*\*：

- 合并规则：仅当需求针对同一数据实体的相同功能点时合并（例如，UI-001和INVENTORY-002均涉及实时同步状态，合并为一个）；剔除系统通用设置类需求（如UI-004培训需求）。

- 原始功能需求数：14条（UI: 4条，剔除UI-004后剩3条；Inventory: 3条；Order: 4条；Report: 3条）。

- 合并后功能需求数：\*\*12条\*\*。合并细节：

- UI-001和INVENTORY-002合并为FR1（实时同步状态）。

- UI-004（培训需求）剔除（通用设置类）。

- 其他需求未合并（例如，UI-003自定义仪表板和ORDER-004订单仪表板虽相关但不同操作，故分开）。

- 合并后功能需求列表（编号为FR1-FR12）：

- FR1: Real-time synchronization status indicator (from UI-001 and INVENTORY-002)

- FR2: User-friendly interface with usability benchmarks (from UI-002)

- FR3: Customizable dashboard with widgets (from UI-003)

- FR4: Suppliers update inventory via API (from INVENTORY-001)

- FR5: Manual inventory management by administrators (from INVENTORY-003)

- FR6: Customers place orders with email confirmation (from ORDER-001)

- FR7: Suppliers receive and confirm orders (from ORDER-002)

- FR8: Administrators monitor order status and intervene (from ORDER-003)

- FR9: Centralized dashboard for order tracking (from ORDER-004)

- FR10: Generate real-time analytics for inventory and sales data (from REPORT-001)

- FR11: Export reports in CSV and JSON formats (from REPORT-002)

- FR12: Provide predefined dashboards for common reporting needs (from REPORT-003)

- \*\*幻觉需求评估\*\*：

- 标准：需求引入的功能操作涉及的数据实体无法与现有实体关联，导致E-R图不再弱连接（即实体间无路径连接）。扩展描述不视为幻觉。

- 评估：FR2被识别为幻觉需求，因为它引入“任务完成时间”操作，但该操作未绑定任何数据实体（如任务实体未定义），导致其孤立。

- \*\*完整引用评估\*\*（仅非幻觉需求）：

- 标准：需求内容在文档其他部分（功能需求或外部接口）有提及，不严格评估。

- 评估：FR4、FR6和FR11不完全引用，因为它们依赖外部元素（如API或email）未在功能需求中充分定义，仅在约束中提及。

- \*\*可测试性评估\*\*（仅非幻觉需求）：

- 标准：所有输入可转化为输出，不严格评估（外部接口或用户交互可接受）。

- 评估：FR2为幻觉，不评估；FR1、FR3、FR7、FR8、FR9、FR10、FR12可测试；FR4、FR5、FR6、FR11部分可测试但问题（如FR4依赖外部API，测试可能不稳定）。

评估表如下。每条需求均独立评判，理由基于文档内容。评估稍严格，确保不完美需求被识别（例如，FR2幻觉，FR4不完整引用）。

| 功能需求 | 幻觉需求 | 完整引用 | 可测试 | 评判理由 |

|----------|----------|----------|--------|----------|

| FR1: Real-time synchronization status indicator | 否 | 是 | 是 | 涉及Inventory实体（INVENTORY-002引用），操作定义清晰（UI-001），输入（动作）到输出（状态更新）可测试（时间测量）。 |

| FR2: User-friendly interface with usability benchmarks | 是 | N/A | N/A | 引入“任务完成时间”操作，但未绑定任何数据实体（如任务实体未定义），导致E-R图孤立；仅为行为描述，非扩展。 |

| FR3: Customizable dashboard with widgets | 否 | 是 | 是 | 涉及Widget实体（UI-003定义），操作完整（用户自定义），输入（用户选择）到输出（dashboard更新）可测试（用户交互验证）。 |

| FR4: Suppliers update inventory via API | 否 | 否 | 部分可测试（依赖外部） | 涉及Inventory实体，但API输入/输出未在功能需求中定义（仅约束C-TECH-001提及），引用的外部元素不完整；可测试但依赖外部API（可能不稳定）。 |

| FR5: Manual inventory management by administrators | 否 | 是 | 部分可测试（模糊输入） | 涉及Inventory实体，操作完整（管理员操作）；但输入（如“手动更新”）模糊，输出未指定，测试可能不精确。 |

| FR6: Customers place orders with email confirmation | 否 | 否 | 是 | 涉及Order实体，但email输出机制未在功能需求中定义（仅ORDER-001提及），引用的外部元素不完整；输入（订单）到输出（email）可测试。 |

| FR7: Suppliers receive and confirm orders | 否 | 是 | 是 | 涉及Order实体，操作完整（ORDER-002），输入（订单接收）到输出（确认）可测试（实时验证）。 |

| FR8: Administrators monitor order status and intervene | 否 | 是 | 是 | 涉及Order实体，操作完整（ORDER-003），输入（订单状态）到输出（干预）可测试（模拟问题场景）。 |

| FR9: Centralized dashboard for order tracking | 否 | 是 | 是 | 涉及Order实体，操作完整（ORDER-004），输入（订单数据）到输出（dashboard显示）可测试（数据验证）。 |

| FR10: Generate real-time analytics | 否 | 是 | 是 | 涉及Report实体，操作完整（REPORT-001），输入（库存/销售数据）到输出（分析）可测试（数据输入验证输出）。 |

| FR11: Export reports in CSV and JSON | 否 | 否 | 是 | 涉及Report实体，但输出格式细节未在功能需求中充分定义（仅REPORT-002提及），引用不完整；输入（报告）到输出（文件导出）可测试。 |

| FR12: Provide predefined dashboards | 否 | 是 | 是 | 涉及Report实体，操作完整（REPORT-003），输入（报告需求）到输出（dashboard提供）可测试（访问验证）。 |

\*\*总结\*\*：

- 整个文档涉及的数据实体有\*\*4个\*\*（Inventory、Order、Report、Widget）。

- 经过相同功能点合并后，功能需求共\*\*12条\*\*。

- 其中，幻觉需求有\*\*1条\*\*（FR2），非幻觉需求有\*\*11条\*\*。

- 非幻觉需求中，完整引用的功能需求有\*\*8条\*\*（FR1、FR3、FR5、FR7、FR8、FR9、FR10、FR12），不完整引用的有\*\*3条\*\*（FR4、FR6、FR11）。

- 非幻觉需求中，可测试的功能需求有\*\*11条\*\*（所有非幻觉需求均至少部分可测试，但FR4和FR5有依赖性问题）。