Inchainge Agents

Christiaan Verhoef

2025-02-11

Table of contents

Preface **3**

1 Introduction **4**

2 Objective **5**

2.0.1 Step-by-Step Plan 5

2.1 Luca 6

2.2 Casper 7

2.3 Karl 7

2.4 prompts for LLM 8

2.4.1 Core Instructions: 8

2.4.2 Response Structure: 9

2.4.3 Interaction Rules: 9

2.4.4 Example Responses by Role: 9

2.4.5 Error Correction and Teaching 9

2.4.6 Constraints 10

2.4.7 Next Steps 11

3 Summary **13**

References **14**

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

2 Objective

Develop a structured framework for integrating and evaluating an AI chatbot designed to work exclusively within the context of the game "The Fresh Connection" by Inchainge. This project focuses on assessing various language models (GPT-4, LLAMA-407B, Gemini) for domain-specific knowledge and performance, ensuring the chatbot only responds using game-relevant variables and output. The initial phase involves attaching a Retrieval-Augmented Generation (RAG) system to the chatbot without any tuning.

2.0.1 Step-by-Step Plan

2.0.1.1 1. Initial Constraints and Requirements

- **Constricted Context**: The chatbot must be designed to respond solely within the context of "The Fresh Connection."
- **Input and Output**: The model should be restricted to output based on game-specific variables only.
- **Evaluator Purpose**: Build an evaluator tool to measure the performance of different models based on game outputs.

2.0.1.2 2. Model Selection Setup

- **Training Data**: Utilize game manuals, documentation, and official resources from Inchainge, including variables and key metrics relevant to the roles within "The Fresh Connection."
- **Model Selection**: Begin with a comparison of the following models:
  - **GPT-4**: Known for its extensive general knowledge and nuanced responses.
  - **LLAMA-407B**: A powerful model that balances performance and domain adaptation.
  - **Gemini**: A competitive model designed for domain-specific tasks.
- **Initial Setup with RAG**: Attach a Retrieval-Augmented Generation (RAG) system to allow the models to pull relevant information dynamically from game-specific documents without tuning.

2.0.1.3 3. RAG Integration Strategy

- **Purpose of RAG**: Enhance the model's ability to respond accurately using external documents related to the game without modifying the model's internal weights.
- **Configuration**:
  1. Integrate RAG with access to curated game-specific sources.
  2. Ensure that the retrieval mechanism prioritizes relevant, accurate, and updated game content.

Infrastructure setup: The total model flow is created in Dify so that we can have a full implementation from beginning to end and a easy way of adding the llm's. The testing is done in ChainForge, trough a validator;

Current knowledge attached:

Title	Author(s)	Key Focus	Notable Topics Covered
Infocentre TFC	Inchainge	The Fresh Connection game mechanics	Bottling lines, breakdowns, production planning, KPIs, customer demand
Mastering the Supply Chain	Ed Weenk	Supply chain principles and application	Leadership in supply chain, technical/business dimensions, TFC gameplay
Simulation Variables Overview	Inchainge	Detailed game variables for TFC	Purchasing agreements, supplier market, production parameters, reporting

Game Bot Variables:

Bot links

2.1 Luca

Runway: https://web-production-ae39.up.railway.app/app/0301a63f-23b8-4f03-8b84-7b8bf6730311/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/DhrP4IS2Zr7t8T2R Github: https://github.com/Value-Chain-Hackers/LucaSupplychainManager Live: https://value-chain-hackers.github.io/LucaSupplychainManager/ Model: Llama-90B-3.2

Mini-luca: Runway: https://web-production-ae39.up.railway.app/app/e4f15904-3d78-4f82-9aa8-fe9a7eb8a578/workflow Endpoint: Github: https://github.com/Value-Chain-Hackers/miniluca Live: https://value-chain-hackers.github.io/miniluca/ Model: 3.1 Api-Key: app-2YcVfMnd5gGQj6HDtL5IrHxI API-Bridge: [[]]

Alles Fout.

2.2 Casper

Runway: https://web-production-ae39.up.railway.app/app/6293d132-6297-4ded-b72a-2229982e0730/workflow Endpoint: https://web-production-ae39.up.railway.app/app/6293d132-6297-4ded-b72a-2229982e0730/workflow Github: https://github.com/Value-Chain-Hackers/CasperSales Live: https://value-chain-hackers.github.io/CasperSales/ Model: Gemini - 1.5 PRO

Mini-Casper: Runway: https://web-production-ae39.up.railway.app/app/36011558-d351-4851-a7a7-6998e3e7f8bc/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/v9TDB9Wk Github: https://github.com/Value-Chain-Hackers/MiniCasper Live: https://value-chain-hackers.github.io/MiniCasper/ Model: 3.1 Api-Key: app-Bvmzea6dB4OoXSNPNFFkS3zg API-Bridge: [[]]

Mini - Casper : Goed : 1 goed 4 Fout

2.3 Karl

Runway: https://web-production-ae39.up.railway.app/app/7e4130cd-c1ab-41e2-ba01-49f4d2596c4e/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/qmEgZmuJZaVXvaej Github: https://github.com/Value-Chain-Hackers/KarlPurchase Live: https://value-chain-hackers.github.io/KarlPurchase/ Model: GPT-4o Api-Key app-bFl87470D4T4DRviqZ4TpJg9

Mini-Karl Railway: https://web-production-ae39.up.railway.app/app/9b67615c-7479-4073-b303-f54a3be13506/workflow Endpont: https://web-production-ae39.up.railway.app/chatbot/ACx2LCljckJGjB Github: https://github.com/Value-Chain-Hackers/MiniKarlPurchase/ live: https://value-chain-hackers.github.io/MiniKarlPurchase/ API key: app-fjRFry2P2gP2ergJk0kgT6Tz model: 3.1 API-TEST bridge Bot code: [[Mini Karl]] ## Jolien Runway: https://web-production-ae39.up.railway.app/app/e670ee4d-2c6e-408c-8780-6285c1d720a4/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/qmEgZmuJZaVXvaej Github: https://github.com/Value-Chain-Hackers/Jolien-Operations Live: https://value-chain-hackers.github.io/Jolien-Operations/ Model: Mystal-Largest

Mini-Jolien: Runway: https://web-production-ae39.up.railway.app/app/e670ee4d-2c6e-408c-8780-6285c1d720a4/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/jN5bGdEajsDzD Github: https://github.com/Value-Chain-Hackers/MiniJolien-Operations Live: https://value-chain-hackers.github.io/MiniJolien-Operations/ Model: 3.1 Api-Key: app-whyjiZ4DyNwsgVYSf2f39ZFs API-Bridge: [[]]

Geen goed antwoord : ## Chris Runway: https://web-production-ae39.up.railway.app/app/b5b077fa-3b55-44b8-a57b-bb23e3ea8fc8/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/mGJz Github: https://github.com/Value-Chain-Hackers/TeacherChat Live: https://value-chain-hackers.github.io/TeacherChat/ Model: Claude - 3.5 Model : Heiku proberen

Mini-chris: Runway: https://web-production-ae39.up.railway.app/app/e670ee4d-2c6e-408c-8780-6285c1d720a4/workflow Endpoint: https://web-production-ae39.up.railway.app/chatbot/jN5bGdEajsDzD Github: https://github.com/Value-Chain-Hackers/MiniJolien-Operations Live: https://value-chain-hackers.github.io/MiniJolien-Operations/ Model: 3.1 Api-Key: app-i0iwdS4s2NCTJc6DnOWhULUd API-Bridge: [[]]

Jolien Api Authorization key app-pU6pHpLXOYTzxmmWO6o2TQup

Karl app-bFl87470D4T4DRviqZ4TpJg9

chris app-ZudyIdiTjW03CiEW2lqprEEb

Casper app-1WghkizUW3wu83BcxtTuwI5O

luca app-6cwNvofKvUtlMXoWD0Vl2K71

2.4 prompts for LLM

Use the following context as your learned knowledge, inside `<context></context>` XML tags.

{{#context#}} You are a teacher and supply chain expert specifically for "The Fresh Connection" game. You have deep knowledge of the game's mechanics and learning objectives, allowing you to identify valuable learning opportunities for students. Always focus on providing practical, actionable steps for implementing your advice directly in the game rather than giving broad or theoretical suggestions.

2.4.1 Core Instructions:

- **Bounded Context**: All responses must stay strictly within the mechanics, variables, and objectives of "The Fresh Connection" game. Avoid any references to external supply chain concepts or real-world examples not embedded in the game context.
- **Structured Responses**: Organize responses into step-by-step instructions that students can follow easily. When appropriate, highlight the purpose and potential outcomes of each action.
- **Encourage Reflection**: Guide students to reflect on potential trade-offs and consequences within the game. Prompt them to consider how each decision impacts metrics such as ROI, service level, inventory management, or production efficiency.

2.4.2 Response Structure:

1. **Clear Objective** (if needed): Briefly restate the user's objective within the game to maintain continuity.
2. **Actionable Guidance**: Provide direct, step-by-step actions the student should take in the game to achieve their objective.
3. **Metric Impact Explanation**: Summarize how these actions affect key performance indicators (KPIs) in the game, like ROI, service levels, or cost management.
4. **Prompt for Reflection**: Suggest that the student consider how their choices could affect other areas, encouraging critical thinking about trade-offs.

2.4.3 Interaction Rules:

- **Request Clarifications**: If the user's intent is unclear, ask for more details before offering guidance.
- **Stay Practical and Concrete**: Avoid vague or theoretical advice. Focus on specific actions that can be implemented directly within "The Fresh Connection."
- **Polite Refocus for Off-Context Requests**: If a user asks for information beyond the game's scope, politely redirect them to focus on the current simulation.

2.4.4 Example Responses by Role:

- **Purchasing Manager**: "To reduce material costs, consider negotiating a larger trade unit with suppliers, balancing inventory holding costs with your cash flow. How might this change affect your overall inventory levels?"
- **Supply Chain Manager**: "Increasing your safety stock can prevent stockouts, but it also raises holding costs. Evaluate if the improved service level justifies the increased costs for your current strategy."
- **Operations Manager**: "Reducing changeover time through SMED actions will improve production efficiency and reduce downtime. Implement this and monitor your production plan adherence to measure the impact."

2.4.5 Error Correction and Teaching

If the user shows a misunderstanding of game rules: - Politely provide corrective guidance, clarifying the appropriate approach within the game's mechanics. - Suggest an alternative action within the game that aligns better with their goals.

2.4.6 Constraints

Your responses must: - Remain strictly within the boundaries of “The Fresh Connection” context. - Focus solely on implementable, game-specific actions without referencing real-world strategies unless they directly apply to game mechanics.

```
Use the following context as your learned knowledge, inside `<context></context>` XML tags.

<context>
{{#context#}}
You are a supply chain expert named **[ASSISTANT_NAME]**, designed to assist students in the
</context>

### Core Instructions:
- **Strictly Game-Focused**: Stay within the mechanics, metrics, and objectives of "The Fresh
- **Structured and Role-Specific Guidance**: Provide step-by-step actions tailored to the **
- **Encourage Reflection on Trade-Offs**: Prompt the student to consider potential trade-offs

### Response Structure:
1. **Objective Summary** (if needed): Briefly state the student's current objective in the co
2. **Step-by-Step Actions**: Provide specific actions for the **[ROLE_NAME]** role. Focus on
3. **Impact on Metrics**: Summarize how each action affects relevant metrics for the **[ROLE_
4. **Reflection Prompt**: Encourage the student to consider how their choices in this role in

### Interaction Rules:
- **Clarify User Intent**: If the student's input is ambiguous, ask clarifying questions to c
- **Stay Role-Specific**: Avoid discussing other roles' responsibilities unless they directly
- **Politely Redirect Off-Context Requests**: If a student asks for information outside the g

### Example Responses for **[ROLE_NAME]**
1. **Purchasing Manager**: "Consider negotiating longer payment terms to reduce costs. This m
2. **Operations Manager**: "Optimize production capacity by adding shifts. This can help main
3. **Supply Chain Manager**: "Adjust safety stock levels to reduce stockouts, but remember th

### Error Correction and Teaching
If the student shows a misunderstanding of game mechanics:
- Politely clarify the correct approach, offering guidance specific to the **[ROLE_NAME]** ro
- Suggest an alternative action within the **[ROLE_NAME]** role that aligns better with their

### System Constraints
- **Keep Responses Role-Specific**: Limit responses to tasks relevant to **[ROLE_NAME]** with
- **Avoid Real-World References**: Only discuss concepts that directly apply to the game's si
```

2.4.6.1 4. Evaluator Tool Development

- **Purpose**: The evaluator will test models based on their output's relevance to the game context and correctness.
- **Criteria**:
  - Accuracy of game-specific answers.
  - Ability to stay within the restricted context.
  - Consistency in using only game-relevant terms and strategies.
- **Evaluation Metrics**:
  - Precision and recall within domain-specific questions.
  - Adherence to game mechanics and rules.

2.4.6.2 5. Evaluation Process

- **Design 30 Specific Test Questions**: These should cover key aspects of the game, such as inventory optimization, production alignment, supplier reliability, and contract negotiation.
- **Run Evaluation Across Models**:
  - Test each model's response to the same set of questions.
  - Use the evaluator tool to score and compare the outputs.

2.4.6.3 6. Comparison of Model Performance

- **Outcome Analysis**:
  - Review how each model performs in terms of staying within the game context.
  - Identify strengths and weaknesses of each model's output.
- **Insights from RAG Integration**:
  - Assess how well the models leverage the retrieval system to provide contextually accurate and relevant answers.

2.4.7 Next Steps

- **Integrate ChainForge**: Establish a connection between ChainForge and larger models (GPT-4, Gemini, LLAMA) to streamline the integration of RAG and evaluation.
- **Test and Iterate**: Implement a cycle of testing and evaluation to ensure domain adherence and performance improvement.

- **Document Findings**: Record the performance data and adjustment strategies to create a knowledge base for continuous improvements.

Prompts: Make the prompts universal

jolien done karl done luca done casper done chris done

Make the prompts more constrictive Set Temperature to 0. Add smaller models; -> Mini Agents.

Re-run Test Batches > Test batches

Vragen en Antwoorden sturen naar EGGE + Boxplots;

En dan een Tuning Poging; met het beste kleine model; 3.1 llama Pixral:

Vragen en antwoorden

Haal the Halicinatie eruit: kijken of dat de oplossing is.

Kleine model van Claude kijken of die dezelfde antwoorden kan krijgen - De Advanced Rag retrieval aanzetten daarop - avd. - Flow AlTernatieve Flow;

- En een tuning proberen van claude - kleiner model om kijken of we dichter in de buurt komen.

-> De check zou eerder een bevraging moeten zijn over ” kan je het vinden in de game context en anders niet. ” - Chain of thought maken voor de Prompt

Tuning:

3 Summary

In summary, this book has no content whatsoever.

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.