Conversational AI in Games: An Empirical Study of LLM-Powered NPC Dialogue Systems

Anthon K. S. Petersen, Emil Rimer, Rasmus Ploug, Marco Scirea

Motivation

→ **NPC dialogue** is central to *immersion*, *agency*, and *narrative depth*

→ **Traditional branching systems** → predictable & repetitive

→ **LLMs** enable *dynamic, open-ended conversations*

→ **Gap:** little empirical research on their *actual impact* in games

→ **Need:** compare designs to *guide future dialogue systems*

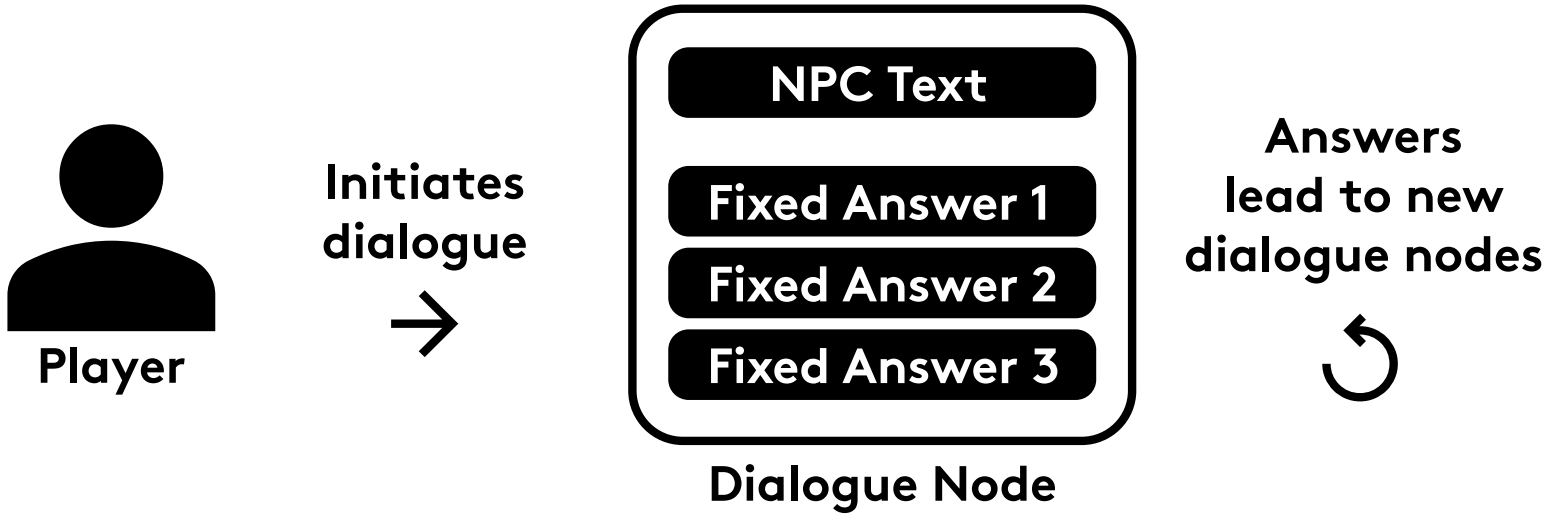
Research Questions

RQ1: How does the integration of different LLM-driven features in NPC dialogues influence player interaction in video games?

RQ2: Did any LLM-driven feature enhance player interaction compared to the Control Version?

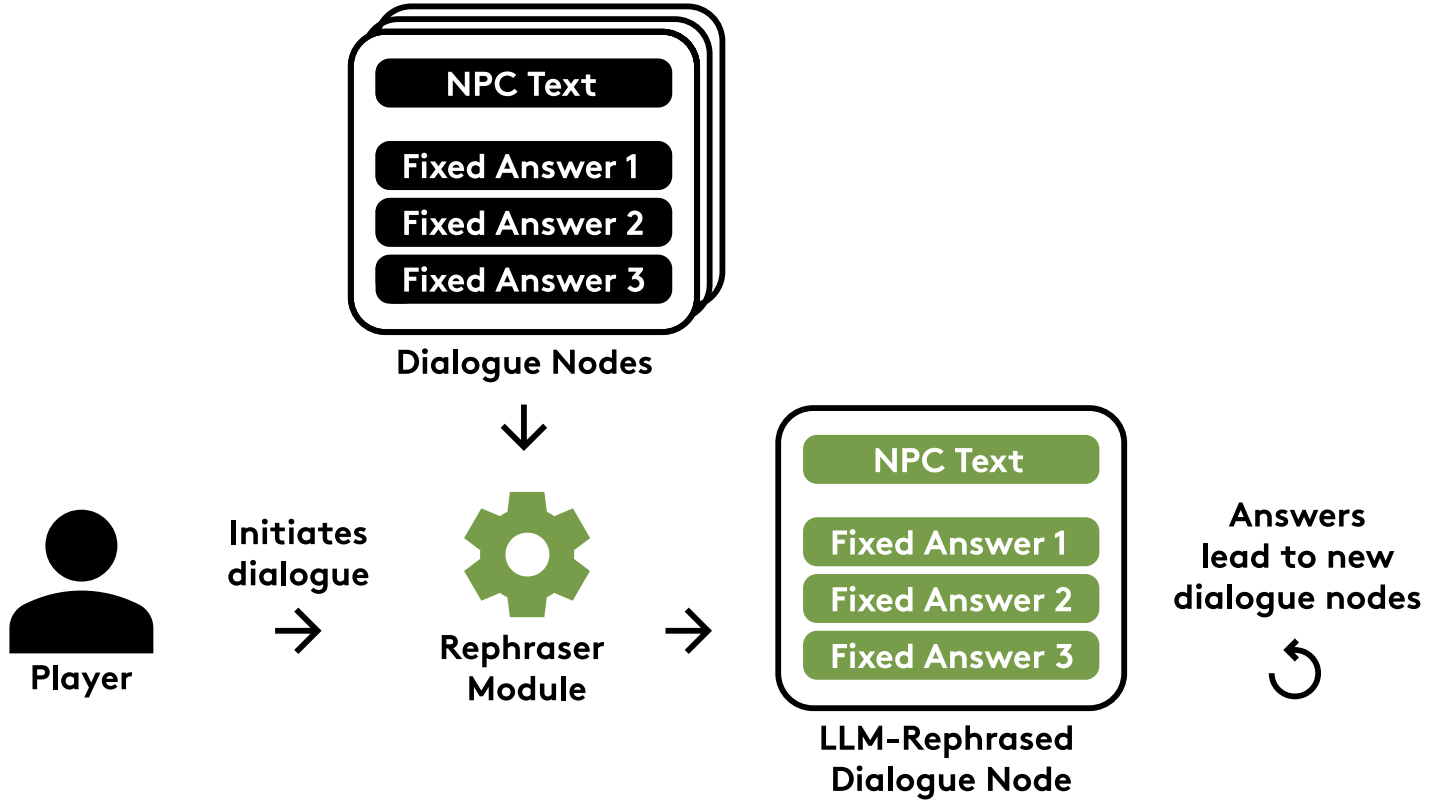
Control Version

→ *Static branching dialogue tree* (traditional RPG style)



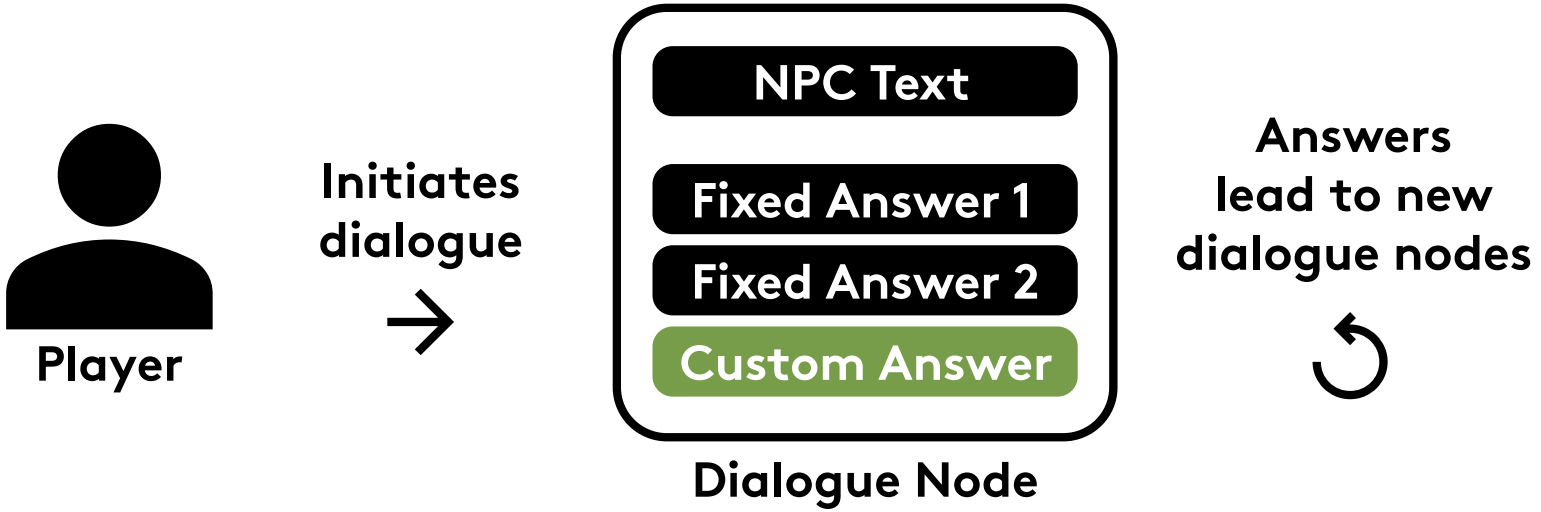
Version A

→ *LLM rephrases* static lines → linguistic variation, but still choice-based



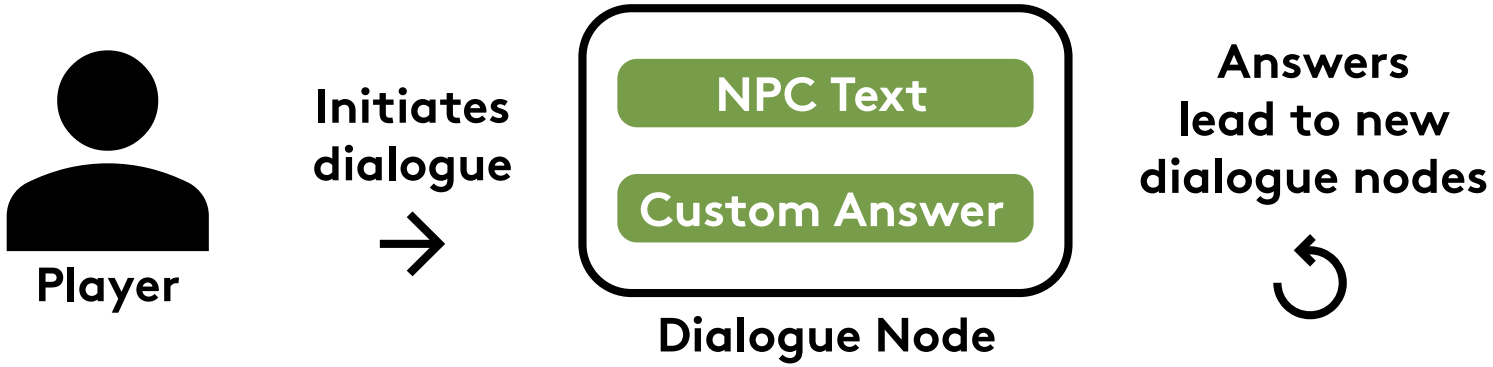
Version B

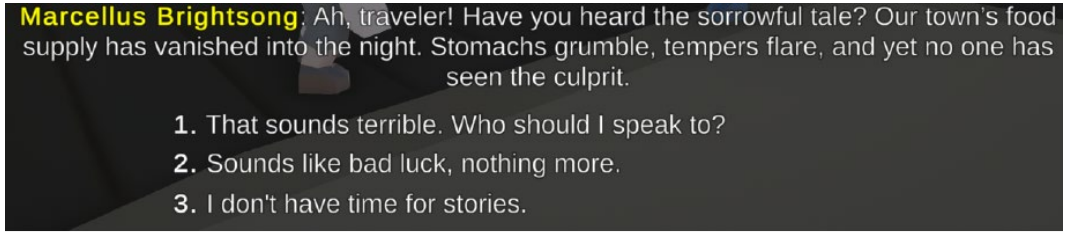
→ *Hybrid system* → fixed options **+** free-text input



Version C

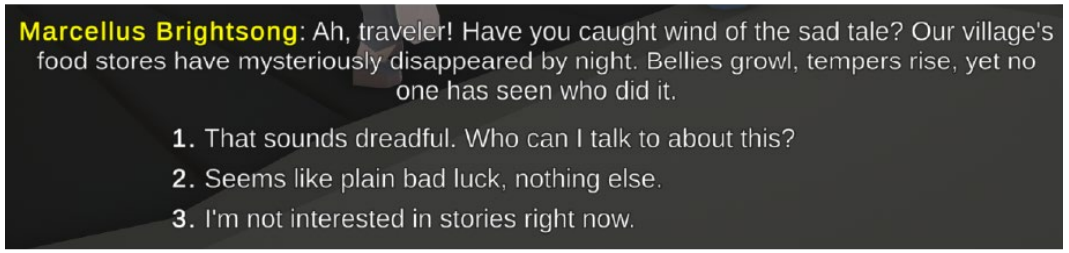
→ *Fully open-ended* → all player input as free text, LLM generates NPC replies





Control Version

Marcellus Brightsong: Ah, traveler! Have you heard the sorrowful tale? Our town's food supply has vanished into the night. Stomachs grumble, tempers flare, and yet no one has seen the culprit.
1. That sounds terrible. Who should I speak to?
2. Sounds like bad luck, nothing more.
3. I don't have time for stories.


Marcellus Brightsong: Ah, traveler! Have you caught wind of the sad tale? Our village's food stores have mysteriously disappeared by night. Bellies growl, tempers rise, yet no one has seen who did it.
1. That sounds dreadful. Who can I talk to about this?
2. Seems like plain bad luck, nothing else.
3. I'm not interested in stories right now.

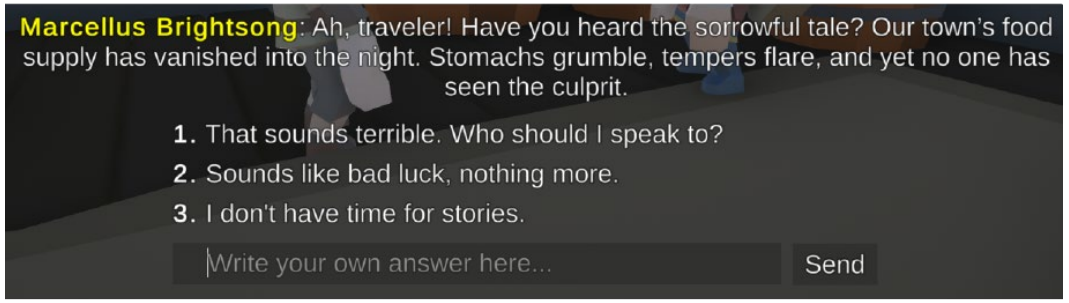
Version A

Marcellus Brightsong: Ah, traveler! Have you heard the sorrowful tale? Our town's food supply has vanished into the night. Stomachs grumble, tempers flare, and yet no one has seen the culprit.
1. That sounds terrible. Who should I speak to?
2. Sounds like bad luck, nothing more.
3. I don't have time for stories.
Write your own answer here...
Send

Marcellus Brightsong
Sigh It's a gloomy day, isn't it? Even my lute strings weep. Is there anything you need, adventurer?
You
What happened?
Marcellus Brightsong
Our town is in dire straits. The food supply was stolen, leaving us all in peril. You should speak to the Innkeeper, she might have some leads.
Send

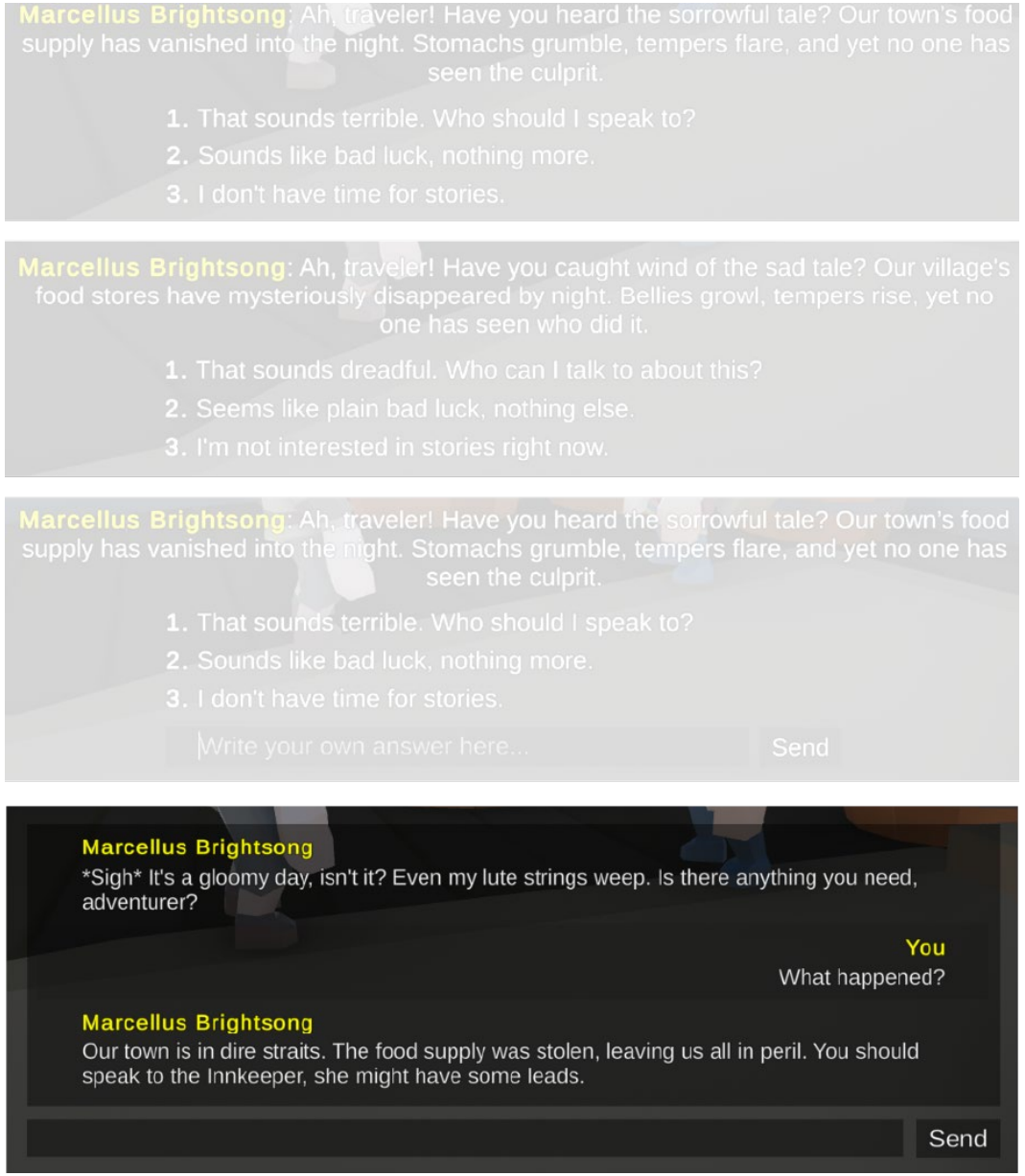
Marcellus Brightsong: Ah, traveler! Have you heard the sorrowful tale? Our town's food supply has vanished into the night. Stomachs grumble, tempers flare, and yet no one has seen the culprit.
1. That sounds terrible. Who should I speak to?
2. Sounds like bad luck, nothing more.
3. I don't have time for stories.

Marcellus Brightsong: Ah, traveler! Have you caught wind of the sad tale? Our village's food stores have mysteriously disappeared by night. Bellies growl, tempers rise, yet no one has seen who did it.
1. That sounds dreadful. Who can I talk to about this?
2. Seems like plain bad luck, nothing else.
3. I'm not interested in stories right now.


Marcellus Brightsong: Ah, traveler! Have you heard the sorrowful tale? Our town's food supply has vanished into the night. Stomachs grumble, tempers flare, and yet no one has seen the culprit.
1. That sounds terrible. Who should I speak to?
2. Sounds like bad luck, nothing more.
3. I don't have time for stories.
Write your own answer here...
Send

Version B

Marcellus Brightsong
Sigh It's a gloomy day, isn't it? Even my lute strings weep. Is there anything you need, adventurer?
You
What happened?
Marcellus Brightsong
Our town is in dire straits. The food supply was stolen, leaving us all in peril. You should speak to the Innkeeper, she might have some leads.
Send


Marcellus Brightsong: Ah, traveler! Have you heard the sorrowful tale? Our town's food supply has vanished into the night. Stomachs grumble, tempers flare, and yet no one has seen the culprit.
1. That sounds terrible. Who should I speak to?
2. Sounds like bad luck, nothing more.
3. I don't have time for stories.
Marcellus Brightsong: Ah, traveler! Have you caught wind of the sad tale? Our village's food stores have mysteriously disappeared by night. Bellies growl, tempers rise, yet no one has seen who did it.
1. That sounds dreadful. Who can I talk to about this?
2. Seems like plain bad luck, nothing else.
3. I'm not interested in stories right now.
Marcellus Brightsong: Ah, traveler! Have you heard the sorrowful tale? Our town's food supply has vanished into the night. Stomachs grumble, tempers flare, and yet no one has seen the culprit.
1. That sounds terrible. Who should I speak to?
2. Sounds like bad luck, nothing more.
3. I don't have time for stories.
Write your own answer here...
Send
Marcellus Brightsong
Sigh It's a gloomy day, isn't it? Even my lute strings weep. Is there anything you need, adventurer?
You
What happened?
Marcellus Brightsong
Our town is in dire straits. The food supply was stolen, leaving us all in peril. You should speak to the Innkeeper, she might have some leads.
Send

Version C

Methodology

- → **Prototype:** Custom Unity RPG in medieval setting
- → **Consistency:** Same quests, mechanics, and visuals across all versions
- → **Participants:** 64 tested (62 analyzed), mean age **25.2**
- → **Procedure:** briefing → gameplay (random dialogue system) → questionnaire
- → **Goal:** Compare *player behavior* + *perceptions* across systems

Data Collection

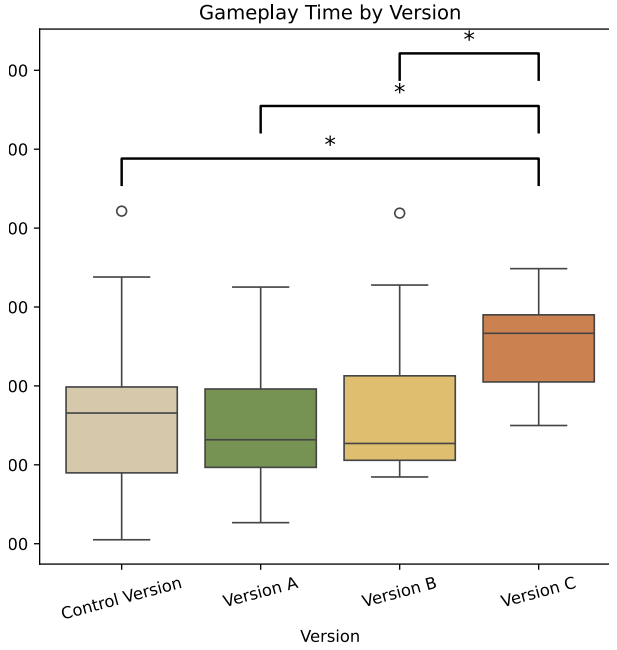
→ **Game Logging:**

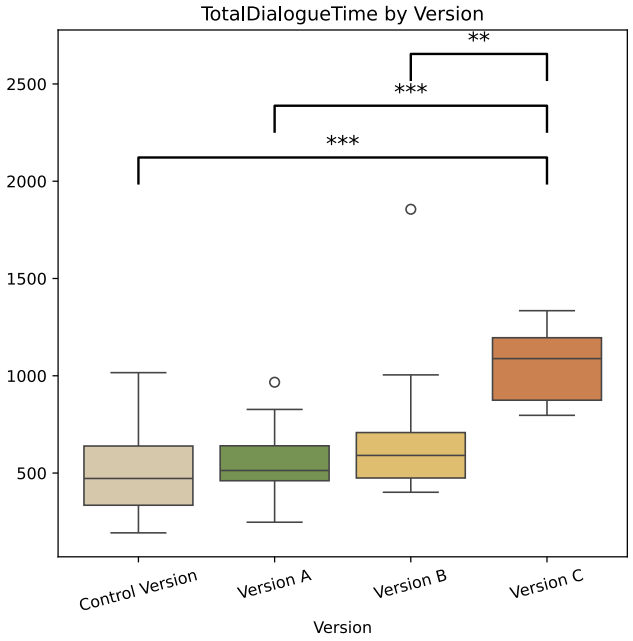
- → Dialogue time & number of exchanges
- → Movement, area transitions, quest outcomes
- → LLM wait/load times, input types (Version B)

→ **Survey Responses:**

- → 24 Likert-scale items + 4 open-ended questions
- → Covered *engagement, immersion, clarity, enjoyment*
- → Collected *technical data & background* for context

Results: Gameplay & Dialogue Time





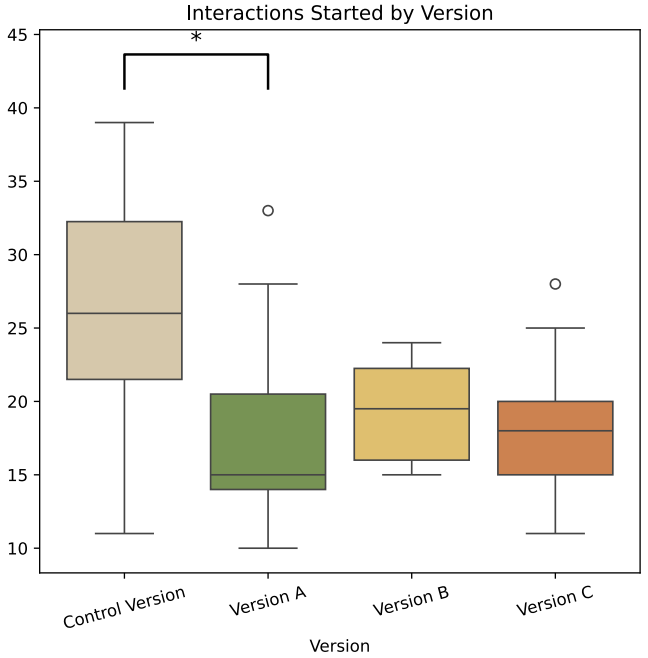
→ **Version C** → significantly longer *total gameplay time* (**p = 0.0064**)

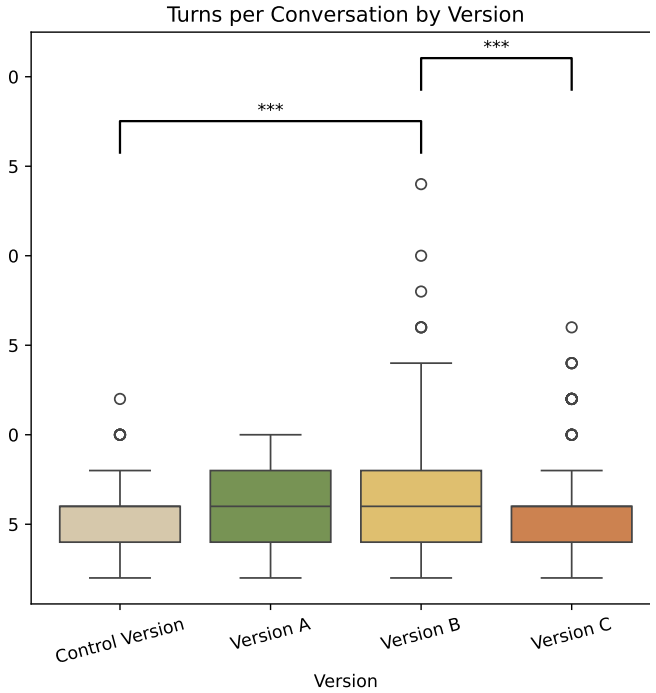
→ Extra time came from **dialogues**, not slower gameplay

→ Version C > Control, A, B in *time spent talking*

→ **Casual players:** longer dialogues in Version C, but *same completion time*

SDU





Results: Interaction Quality

→ **Control (CV):** more dialogue initiations than Version A (**p = 0.0105**)

→ **Version A:** fewer initiations, lower engagement

→ **Version B:** longer *turns per interaction* (**p < .0001**)

→ **Version C:** balanced *volume* + *depth* → most engaging

Survey Question	Control	*Version A*	*Version B*	*Version C*
The game was engaging and kept my interest throughout.	3.94	3.67	4.19	4.20
I found the NPCs interesting to talk to.	3.25	3.00	3.75	4.00
I felt motivated to gather information from NPCs.	3.94	3.80	3.94	3.93
The game's dialogue system made the experience enjoyable.	3.75	3.47	4.12	4.27
I felt in control of my interactions with NPCs.	3.75	3.53	3.12	3.87
The dialogue system allowed for meaningful conversations.	3.31	3.00	3.44	3.60
I felt the NPCs reacted naturally to my dialogue choices.	4.00	3.93	3.44	3.87
NPCs provided key information in an immersive way.	4.00	4.27	3.94	4.00
I felt encouraged to experiment with dialogue.	3.19	3.53	3.94	4.07
There were moments the dialogue felt unnatural or confusing.	1.94	2.00	3.19	2.87
After speaking with NPCs, I knew what to do next.	3.88	4.53	4.25	4.00
Finding and using info from NPCs was engaging.	4.19	3.93	3.94	4.07
NPCs conveyed information consistently.	4.19	4.20	3.75	4.20
I would be interested in playing a longer version of this game.	4.00	3.67	4.50	4.47
I would be interested in games with similar dialogue systems.	3.88	3.60	4.31	4.13
The dialogue system increased my interest in AI-powered NPCs.	3.19	3.07	3.75	3.93

TABLE 5.4: Likert-scale Survey Averages by Game Version

Results: Player Impressions

→ **Version C praised:** *"More natural"*, *"felt like role-playing"*

→ Players valued **freedom** and **expressiveness**

→ Stronger sense of **immersion** compared to fixed options

→ **Critiques:** occasional *off-topic answers* or NPCs *pushing dialogue forward*

→ Overall: Version C felt **intuitive** and **engaging**, especially for casual players

Discussion

Metric	Control Version	*Version A*	*Version B*	*Version C*
Engagement	+	-	++	+++
Flexibility	+	+	++	+++
Stability	+	+	- -	-
Technical Performance	+	-	- -	-
Player Agency	+	+	++	+++
Perceived Naturalness	+	+	±	++
Dialogue Control (Dev-side)	+	±	±	- -

TABLE 6.1: Comparative summary of each dialogue version across key metrics

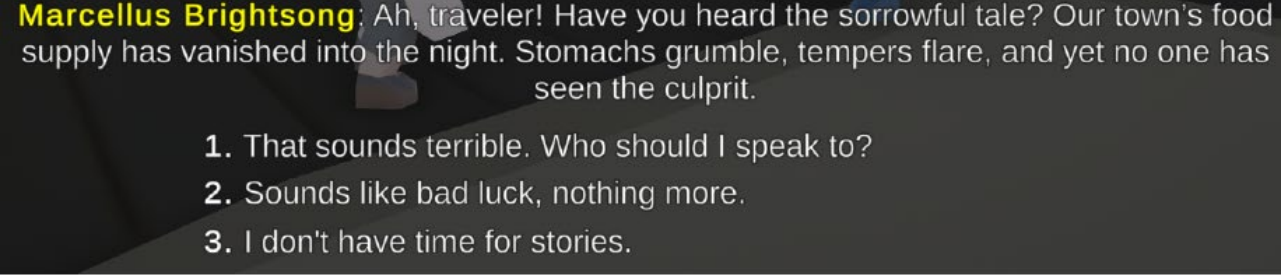
→ **Control (CV):** stable, predictable → but rigid & less immersive

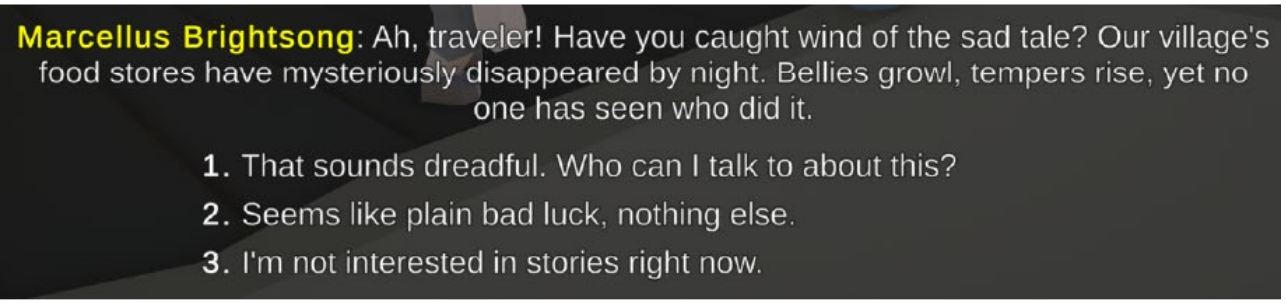
→ **Version A:** minor variation, added load time → limited impact

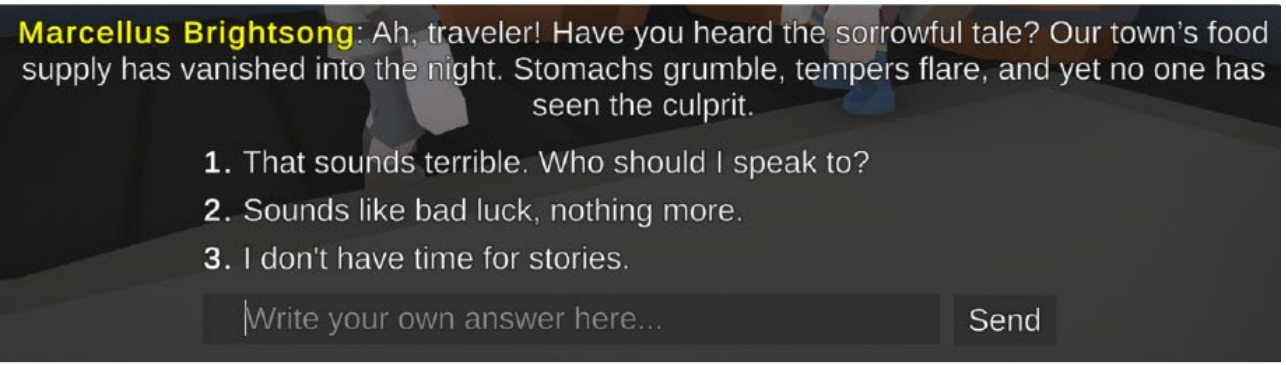
→ **Version B:** hybrid freedom + structure → engaging, but inconsistent

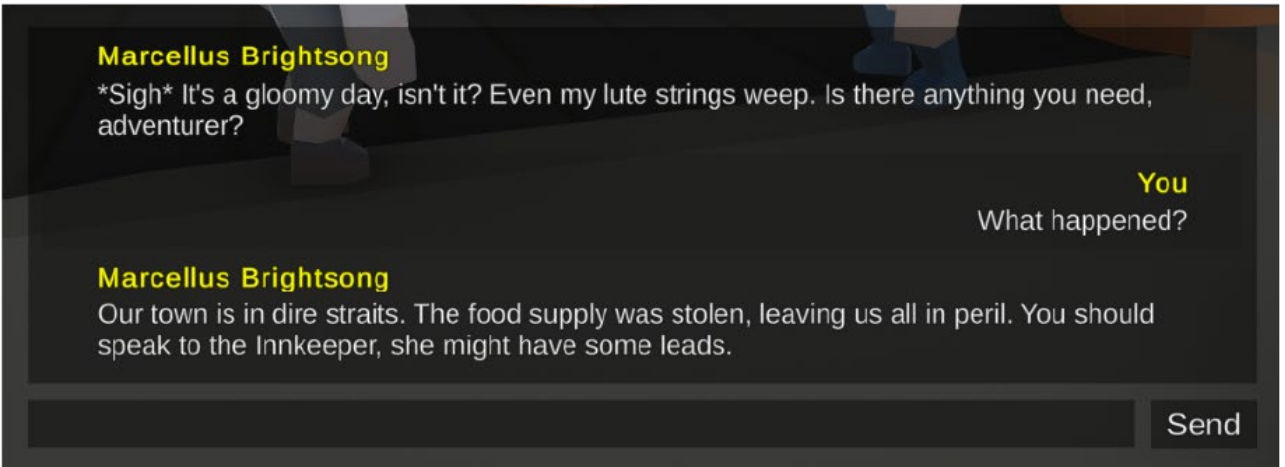
→ **Version C:** most engaging & natural → strong immersion, favored casual players

→ **Trade-off:** higher player agency vs. stability & developer control









Future Work & Conclusion

→ **Future Work**

- → **Broader testing** → larger, more diverse player groups and different genres
- → **Improve stability** → refine hybrid systems with better prompt design and tuning

→ **Conclusion**

- → **Open-ended dialogue (Version C)** → highest engagement and immersion
- → **Casual players** → showed the strongest benefits from natural conversation

Thank you!