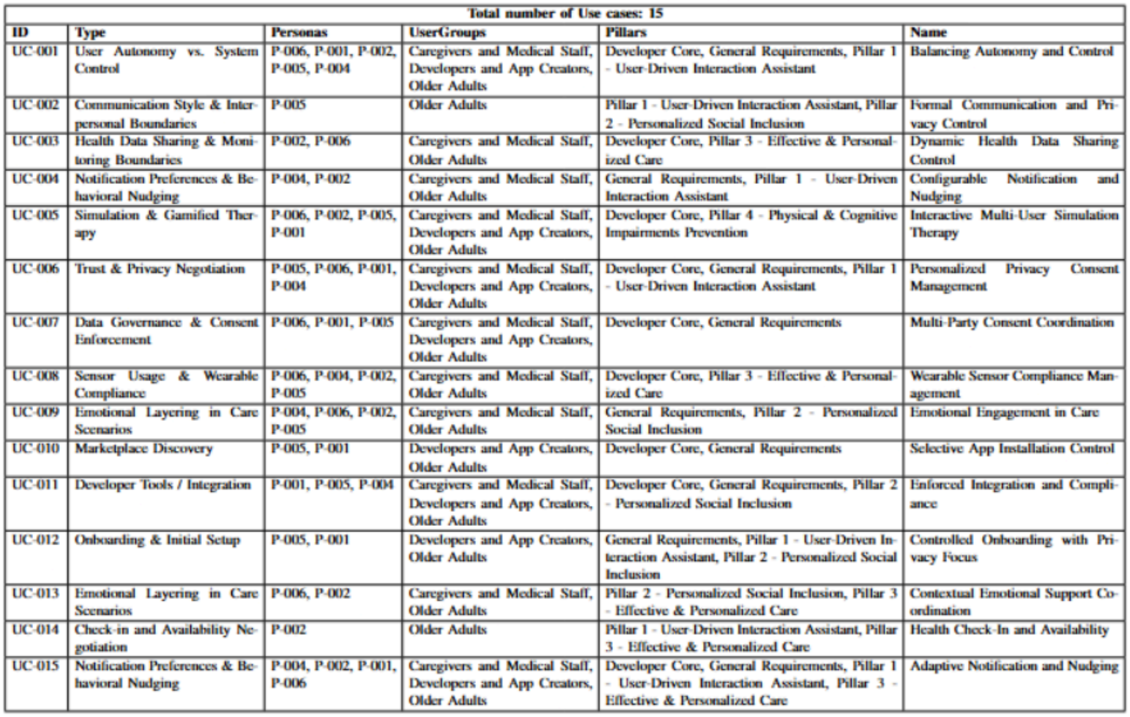
FINAL SUMMARY REPORT

Experiments conducted

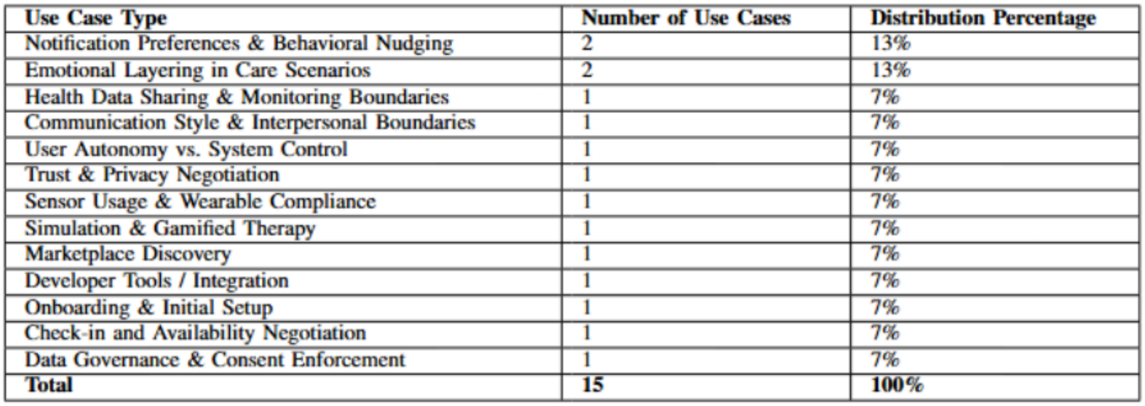
Generate use cases evaluation

We have executed with the five personas and retrieved responses using gpt-4.1-mini as the main case study.



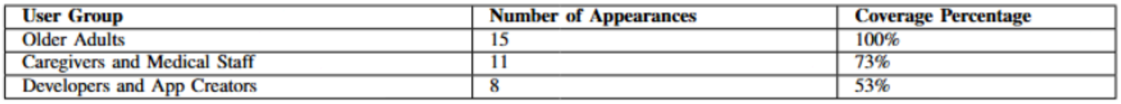
Total number of Use cases: 15					
ID	**Type**	**Personas**	**UserGroups**	**Pillars**	**Name**
UC-001	User Autonomy vs. System Control	P-006, P-001, P-002, P-005, P-004	Caregivers and Medical Staff, Developers and App Creators, Older Adults	Developer Core, General Requirements, Pillar 1 - User-Driven Interaction Assistant	Balancing Autonomy and Control
UC-002	Communication Style & Interpersonal Boundaries	P-005	Older Adults	Pillar 1 - User-Driven Interaction Assistant, Pillar 2 - Personalized Social Inclusion	Formal Communication and Privacy Control
UC-003	Health Data Sharing & Monitoring Boundaries	P-002, P-006	Caregivers and Medical Staff, Older Adults	Developer Core, Pillar 3 - Effective & Personalized Care	Dynamic Health Data Sharing Control
UC-004	Notification Preferences & Behavioral Nudging	P-004, P-002	Caregivers and Medical Staff, Older Adults	General Requirements, Pillar 1 - User-Driven Interaction Assistant	Configurable Notification and Nudging
UC-005	Simulation & Gamified Therapy	P-006, P-002, P-005, P-001	Caregivers and Medical Staff, Developers and App Creators, Older Adults	Developer Core, Pillar 4 - Physical & Cognitive Impairments Prevention	Interactive Multi-User Simulation Therapy
UC-006	Trust & Privacy Negotiation	P-005, P-006, P-001, P-004	Caregivers and Medical Staff, Developers and App Creators, Older Adults	Developer Core, General Requirements, Pillar 1 - User-Driven Interaction Assistant	Personalized Privacy Consent Management
UC-007	Data Governance & Consent Enforcement	P-006, P-001, P-005	Caregivers and Medical Staff, Developers and App Creators, Older Adults	Developer Core, General Requirements	Multi-Party Consent Coordination
UC-008	Sensor Usage & Wearable Compliance	P-006, P-004, P-002, P-005	Caregivers and Medical Staff, Older Adults	Developer Core, Pillar 3 - Effective & Personalized Care	Wearable Sensor Compliance Management
UC-009	Emotional Layering in Care Scenarios	P-004, P-006, P-002, P-005	Caregivers and Medical Staff, Older Adults	General Requirements, Pillar 2 - Personalized Social Inclusion	Emotional Engagement in Care
UC-010	Marketplace Discovery	P-005, P-001	Developers and App Creators, Older Adults	Developer Core, General Requirements	Selective App Installation Control
UC-011	Developer Tools / Integration	P-001, P-005, P-004	Caregivers and Medical Staff, Developers and App Creators, Older Adults	Developer Core, General Requirements, Pillar 2 - Personalized Social Inclusion	Enforced Integration and Compliance
UC-012	Onboarding & Initial Setup	P-005, P-001	Developers and App Creators, Older Adults	General Requirements, Pillar 1 - User-Driven Interaction Assistant, Pillar 2 - Personalized Social Inclusion	Controlled Onboarding with Privacy Focus
UC-013	Emotional Layering in Care Scenarios	P-006, P-002	Caregivers and Medical Staff, Older Adults	Pillar 2 - Personalized Social Inclusion, Pillar 3 - Effective & Personalized Care	Contextual Emotional Support Coordination
UC-014	Check-in and Availability Negotiation	P-002	Older Adults	Pillar 1 - User-Driven Interaction Assistant, Pillar 3 - Effective & Personalized Care	Health Check-In and Availability
UC-015	Notification Preferences & Behavioral Nudging	P-004, P-002, P-001, P-006	Caregivers and Medical Staff, Developers and App Creators, Older Adults	Developer Core, General Requirements, Pillar 1 - User-Driven Interaction Assistant, Pillar 3 - Effective & Personalized Care	Adaptive Notification and Nudging

Fig 1. Summary of use cases generated



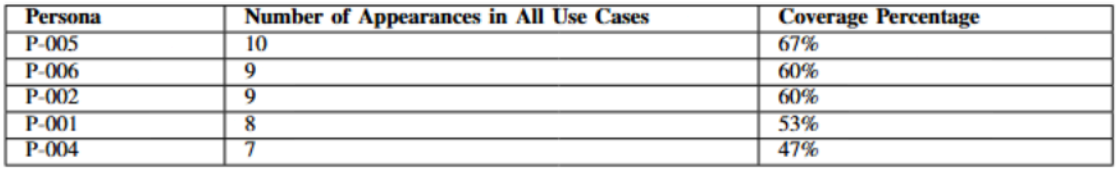
Use Case Type	Number of Use Cases	Distribution Percentage
Notification Preferences & Behavioral Nudging	2	13%
Emotional Layering in Care Scenarios	2	13%
Health Data Sharing & Monitoring Boundaries	1	7%
Communication Style & Interpersonal Boundaries	1	7%
User Autonomy vs. System Control	1	7%
Trust & Privacy Negotiation	1	7%
Sensor Usage & Wearable Compliance	1	7%
Simulation & Gamified Therapy	1	7%
Marketplace Discovery	1	7%
Developer Tools / Integration	1	7%
Onboarding & Initial Setup	1	7%
Check-in and Availability Negotiation	1	7%
Data Governance & Consent Enforcement	1	7%
Total	**15**	**100%**

Fig 2. Use case analysis by predefined types



User Group	Number of Appearances	Coverage Percentage
Older Adults	15	100%
Caregivers and Medical Staff	11	73%
Developers and App Creators	8	53%

Fig 3. Use case Coverage across user groups



Persona	Number of Appearances in All Use Cases	Coverage Percentage
P-005	10	67%
P-006	9	60%
P-002	9	60%
P-001	8	53%
P-004	7	47%

Fig 4. Use case involvement by personas

We analyzed the 15 use cases generated by the pipeline, detailed in Figs 1, 2, 3, 4. The use cases cover a wide variety of types mentioned above (aligned with the personas' distinct goals and challenges), reflecting ALFRED's multi-faceted system objectives.

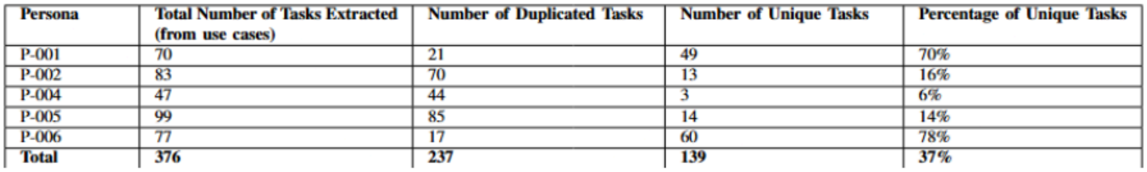
When it comes to the type of Use Case Type, only two type have the highest number and distribution percentage, with 2 use cases and 13%, respectively. The rest of all type are the same each other, with only 1 use case and 7%.

All user groups are well represented: Older Adults always engage in all use cases (100% coverage) with 15 times appearing, while Caregivers and Medical Staff takes the second place in participation with 11 time appearances (80% coverage). Developers comes last in approximately 67% coverage and 8 time appearances

Persona-level involvement quite varies, P-005 (Older Adult) among the highest with 10 time appearances in every use cases with 67% coverage, then P-006 (Informal Caregiver) with P-002 (Difficult Older Adult) are the second highest with around 60% coverage and 9 appearances in all use cases. P-001 (Developer) and P-004 (Nurse) are among two of the last place in number of appearances with percentage of coverage, with respective 8, 7 times and 53%, 47%

This distribution confirms the pipeline's effectiveness in capturing relevant system interactions per persona and user group.

Extracted tasks



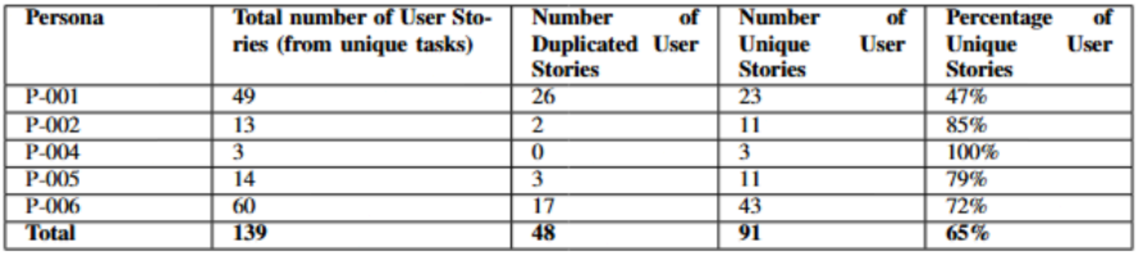
Persona	Total Number of Tasks Extracted (from use cases)	Number of Duplicated Tasks	Number of Unique Tasks	Percentage of Unique Tasks
P-001	70	21	49	70%
P-002	83	70	13	16%
P-004	47	44	3	6%
P-005	99	85	14	14%
P-006	77	17	60	78%
Total	**376**	**237**	**139**	**37%**

Fig 5. Task extraction (Phase 2d) and deduplication (Phase 2e) analysis (by each persona)

The Figure 5 summarize task extraction and deduplication outcomes across personas. From 376 initially extracted tasks (all personas), 237 were identified as duplicates or semantically overlaping, leaving 139 unique, valid tasks (37% of the original)

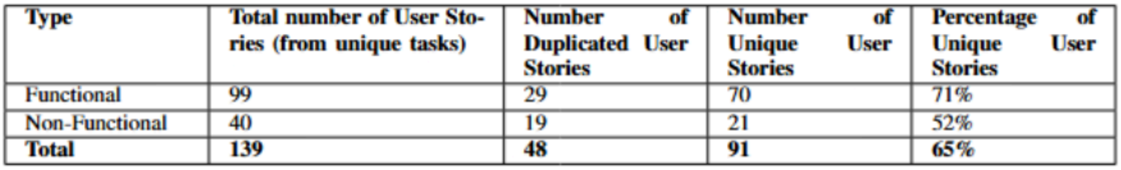
Also, the no-need-to-deduplicate ratio (unique/total), as also shown in the Fig 5, aried on every persona, with the highest around 78% (60/77), in comparison with P-004 with only 6% (3/47), even though they are from the same Caregiver group. Then, P-001 (Developer) has the second highest unique rate, with below than the top just only about 6% below (49/70). The other two from Older group, which are P-002 and P-005, stand at the second and third place respectively, with 16% and 14% (13/83 and 14/99). Overally, this suggests that Older Adults are most affected by the duplicated tasks problem when extracting them from use cases, followed by the Caregiver and Developer user groups

User story generation



Persona	Total number of User Stories (from unique tasks)	Number of Duplicated User Stories	Number of Unique User Stories	Percentage of Unique User Stories
P-001	49	26	23	47%
P-002	13	2	11	85%
P-004	3	0	3	100%
P-005	14	3	11	79%
P-006	60	17	43	72%
Total	**139**	**48**	**91**	**65%**

Fig 6. User story uniqueness analysis by personas



Type	Total number of User Stories (from unique tasks)	Number of Duplicated User Stories	Number of Unique User Stories	Percentage of Unique User Stories
Functional	99	29	70	71%
Non-Functional	40	19	21	52%
Total	**139**	**48**	**91**	**65%**

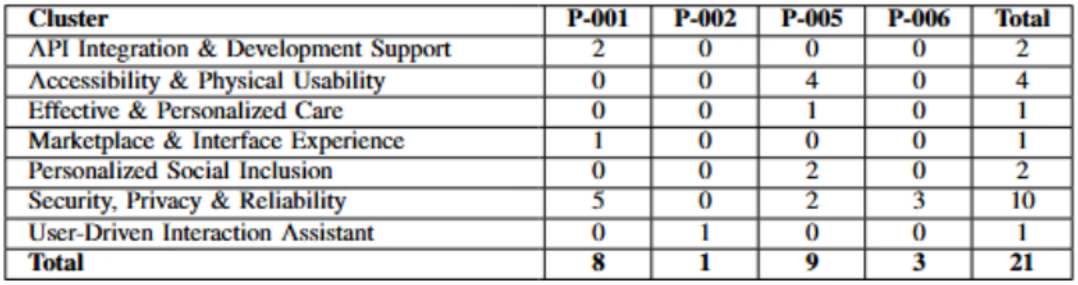
Fig 7. User story uniqueness by types

The figs 6 and 7 shows how the user stories are generated, and deduplicated. Note that the duplication check for User stories is executed after the clustering phases. Overally, there are about 1/3 user stories associated with all personas are detected as duplicated and removed

However, the problem of similar information in user stories still exists dramatically and the team has not completely solved it, despite trying to change the prompt many times. The team has thought that the only way is that, instead of inputting a number of user stories (by cluster and persona) into LLM and letting it show duplicated ones as it is now in the pipeline, each time only input 2 user stories into the prompt, so the result will be as accurate as possible. However, we decided not to choose this option because of the huge limitation in Execution time (Max Time complexity = $O(n^2)$)

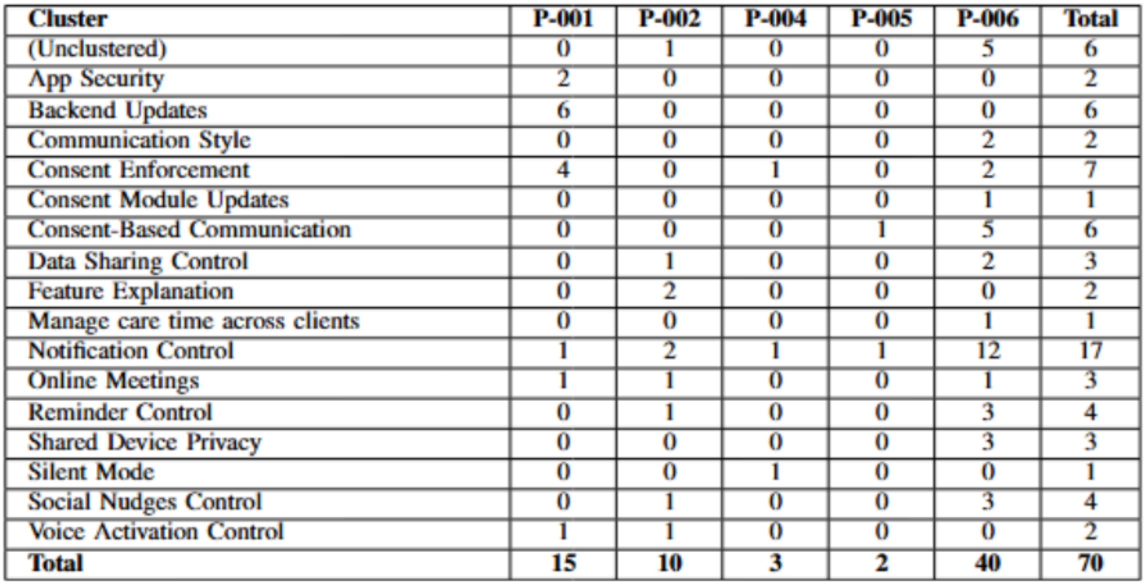
The number of FUSs is generally twice the number of NFUSs (as shown in fig 7). This ratio also means the ratio of number of clusters of FUSs vs NFUSs (≈ 15/8), as calculated using the formula Num_FUS_Clusters = (Num_FUSs / Num_NFUSs) * Num_NFUS_Clusters

The detailed of clusters (for both functional and non-functional user stories, from each persona) are included in Figs 8 and 9.



Cluster	P-001	P-002	P-005	P-006	Total
API Integration & Development Support	2	0	0	0	2
Accessibility & Physical Usability	0	0	4	0	4
Effective & Personalized Care	0	0	1	0	1
Marketplace & Interface Experience	1	0	0	0	1
Personalized Social Inclusion	0	0	2	0	2
Security, Privacy & Reliability	5	0	2	3	10
User-Driven Interaction Assistant	0	1	0	0	1
Total	**8**	**1**	**9**	**3**	**21**

Fig 8. Non-functional user stories clustering by personas



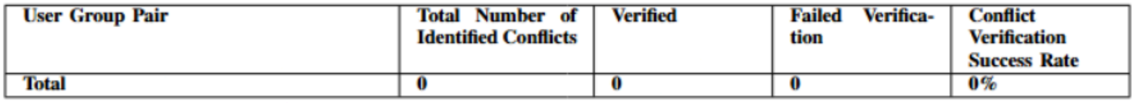
Cluster	P-001	P-002	P-004	P-005	P-006	Total
(Unclustered)	0	1	0	0	5	6
App Security	2	0	0	0	0	2
Backend Updates	6	0	0	0	0	6
Communication Style	0	0	0	0	2	2
Consent Enforcement	4	0	1	0	2	7
Consent Module Updates	0	0	0	0	1	1
Consent-Based Communication	0	0	0	1	5	6
Data Sharing Control	0	1	0	0	2	3
Feature Explanation	0	2	0	0	0	2
Manage care time across clients	0	0	0	0	1	1
Notification Control	1	2	1	1	12	17
Online Meetings	1	1	0	0	1	3
Reminder Control	0	1	0	0	3	4
Shared Device Privacy	0	0	0	0	3	3
Silent Mode	0	0	1	0	0	1
Social Nudges Control	0	1	0	0	3	4
Voice Activation Control	1	1	0	0	0	2
Total	**15**	**10**	**3**	**2**	**40**	**70**

Fig 9. Functiona user story clustering by personas

Conflict handling

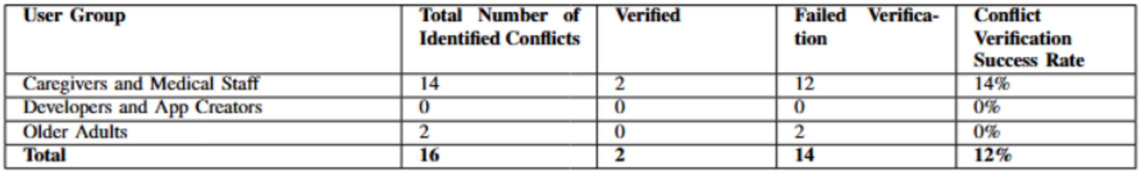
As we have mentioned, all conflict handlings have to go through 3 phases:

- Phase a: Conflict identifying
- Phase b: (Identified) conflict verifying
- Phase c: (Verified) conflict resolution



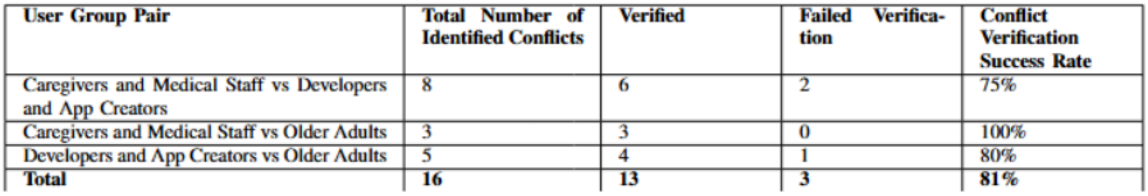
User Group Pair	Total Number of Identified Conflicts	Verified	Failed Verification	Conflict Verification Success Rate
Total	**0**	**0**	**0**	**0%**

Fig 10. Count of non-functional user story conflicts within each group



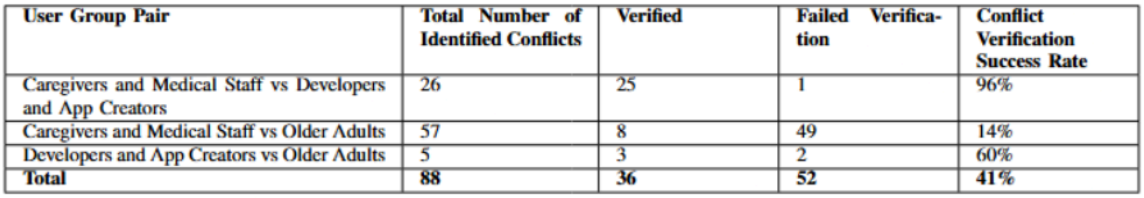
User Group	Total Number of Identified Conflicts	Verified	Failed Verification	Conflict Verification Success Rate
Caregivers and Medical Staff	14	2	12	14%
Developers and App Creators	0	0	0	0%
Older Adults	2	0	2	0%
Total	**16**	**2**	**14**	**12%**

Fig 11. Count of functional user story conflicts within each group



User Group Pair	Total Number of Identified Conflicts	Verified	Failed Verification	Conflict Verification Success Rate
Caregivers and Medical Staff vs Developers and App Creators	8	6	2	75%
Caregivers and Medical Staff vs Older Adults	3	3	0	100%
Developers and App Creators vs Older Adults	5	4	1	80%
Total	**16**	**13**	**3**	**81%**

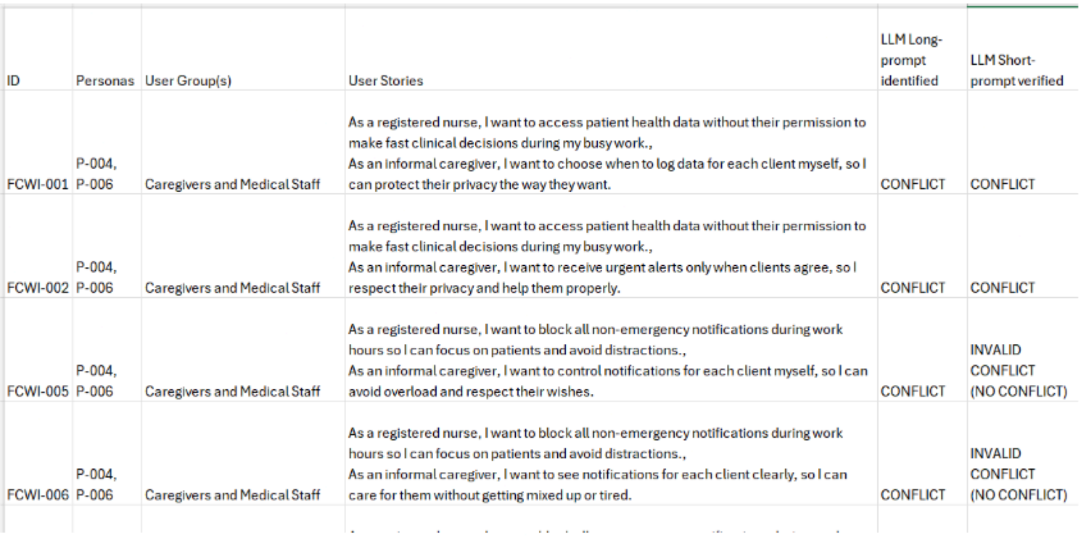
Fig 12. Count of non-functional user story conflicts across two groups



User Group Pair	Total Number of Identified Conflicts	Verified	Failed Verification	Conflict Verification Success Rate
Caregivers and Medical Staff vs Developers and App Creators	26	25	1	96%
Caregivers and Medical Staff vs Older Adults	57	8	49	14%
Developers and App Creators vs Older Adults	5	3	2	60%
Total	**88**	**36**	**52**	**41%**

Fig 13. Count of functional user story conflicts across two groups

The figures 10, 11, 12 and 13 hows that the Confliting Verification Succcess rate (which mean both Phase a and Phase b state that the pair is in Conflict) is under 50% in total (0 + 2 + 13 + 36) / (0 + 16 + 16 + 88) = 51 /120 = 42%)

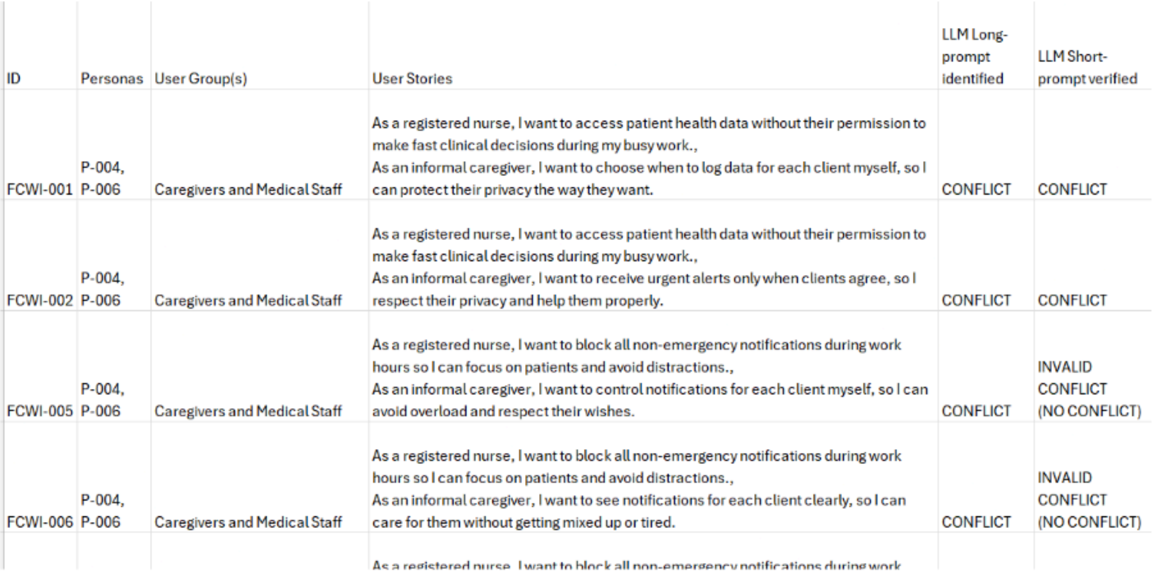


ID	Personas	User Group(s)	User Stories	LLM Long-prompt identified	LLM Short-prompt verified
FCWI-001	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to access patient health data without their permission to make fast clinical decisions during my busy work., As an informal caregiver, I want to choose when to log data for each client myself, so I can protect their privacy the way they want.	CONFLICT	CONFLICT
FCWI-002	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to access patient health data without their permission to make fast clinical decisions during my busy work., As an informal caregiver, I want to receive urgent alerts only when clients agree, so I respect their privacy and help them properly.	CONFLICT	CONFLICT
FCWI-005	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to block all non-emergency notifications during work hours so I can focus on patients and avoid distractions., As an informal caregiver, I want to control notifications for each client myself, so I can avoid overload and respect their wishes.	CONFLICT	INVALID CONFLICT (NO CONFLICT)
FCWI-006	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to block all non-emergency notifications during work hours so I can focus on patients and avoid distractions., As an informal caregiver, I want to see notifications for each client clearly, so I can care for them without getting mixed up or tired.	CONFLICT	INVALID CONFLICT (NO CONFLICT)

Fig 14 Some pairs of conflicts identified in phases a, then verified in phaces b

To illustrate the effectiveness of the short-prompt technique, we have manually checked all identified and verified conflicting pairs (partially shown in Fig 14). The result table of User Story Conflict Verification Analysis by Human includes 120 conflicting pairs of user stories identifying by Phase a, which are partially denied by the Verification Phase b. The team has manually check how much the Verification results (Conflict/Invalid conflict) would match the team's idea (Conflict/Not conflict). As a result, about 85/120 (about 71%) pairs of conflicting LLM verifications match the ideas of the team. The remaining 29% (35 pairs) are:

- mostly (33 conflicts) when the identified conflict is verified as a Valid Conflict as well, while the team does not think it is actually a clear conflict/inconsistency
- Only 2 conflicts (FCAT-041 and FCWI-016), which have been identified as conflicts in Phase a (long-prompt), but denied in Phase b (short- prompt verification), while the team thinks they are still the conflicts



ID	Personas	User Group(s)	User Stories	LLM Long-prompt identified	LLM Short-prompt verified
FCWI-001	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to access patient health data without their permission to make fast clinical decisions during my busy work., As an informal caregiver, I want to choose when to log data for each client myself, so I can protect their privacy the way they want.	CONFLICT	CONFLICT
FCWI-002	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to access patient health data without their permission to make fast clinical decisions during my busy work., As an informal caregiver, I want to receive urgent alerts only when clients agree, so I respect their privacy and help them properly.	CONFLICT	CONFLICT
FCWI-005	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to block all non-emergency notifications during work hours so I can focus on patients and avoid distractions., As an informal caregiver, I want to control notifications for each client myself, so I can avoid overload and respect their wishes.	CONFLICT	INVALID CONFLICT (NO CONFLICT)
FCWI-006	P-004, P-006	Caregivers and Medical Staff	As a registered nurse, I want to block all non-emergency notifications during work hours so I can focus on patients and avoid distractions., As an informal caregiver, I want to see notifications for each client clearly, so I can care for them without getting mixed up or tired.	CONFLICT	INVALID CONFLICT (NO CONFLICT)
			As a registered nurse, I want to block all non-emergency notifications during work		

Fig 15. Some pairs of conflicts verified by Phases b, which are then resolved in phases c

When it comes to conflict resolution (results shown paritially in Fig 15), the pipeline, after Verification with a short prompt, detects a total of 51 conflicts. When doing manual analysis, we passed the 33 pairs that the team did not think were conflicts earlier, which means the checked pairs are 51 – 33 = 18, then the team's agreement rate is 100% - (3/18) = 83% (the team has disagreed with 3 additional pairs in terms of resolution ideas)

Work completed checklist

Notes: This semester (Project B) has been separated into four sprints from 1-4 (three weeks each)

	Requirement	Status	Sprint(s)	Notes
Manual research and analysis	Research on the techniques applied in the project: - User story clustering - Conflict identification and resolution	Partially Completed	Project A's Sprints 1-3: Cannot be applied in the project Project B's Sprints 2-3: Find new techniques	The old findings are not suitable as they are compliocated to implement Applied 3 techniques found in 4 papers
	Research on the ALFRED document - Personas example	Partially Completed,	Project A's Sprints 2-3: Vague on the	Client did not happy with the understanding

	- Use case example - User story example - Diagram of how to create user stories	but not satisfied	ideas of the ALFRED doc Project B’s Sprints 1: Failed Project B’s Sprints 2: Completed	about the ALFRED document in Sprint 1
	Manually creating personas	Completed	Project A’s Sprints 1-3: Created but not used Project B’s Sprint 1-2: Re-created but the client did not approve Project B’s Sprint 3: Client has agreed	
	Manually create the ALFRED system contexts and found techniques summaries	Completed	Project B’s Sprint 3	
	Manually designing the use cases type	Completed	Project B’s Sprint 1-2: No clear design Project B’s Sprint 3: Design but cannot highlight conflicts Project B’s Sprint 4: Finalise the design of use case types	
	Manually created the user stories	Completed	Project B’s Sprints 2-3	
	Manually detected conflicts using the found techniques	Completed	Project B’s Sprints 2: Failed Project B’s Sprint 3: Completed	
Implementation - backend	Load the personas and system contexts	Completed	Project B’s Sprint 1: Personas loaded logic Project B’s Sprint 2-3: System (ALFRED) contexts loaded logic	

	Generate use cases: - Generate use case scenarios - Extracted unique tasks from each scenario	Completed	Project B’s Sprint 1-2: Completed but not satisfied Project B’s Sprints 3-4: Completed	
	Generate user stories: - Generate user stories from corresponding tasks - Classify and cluster user stories - Decomposite the NFUSs	Completed, **but not satisfied (almost failed)**	Project B’s Sprints 1-2: Failed Project B’s Sprints 3-4: Completed but not satisfied	The client think some user stories are vague, while some are unrealistic
	Handling conflicts: - Identify conflicts - Verify the identified conflicts - Resolve the verified conflicts	Partially Completed, **but not satisfied (almost failed)**	Project B’s Sprints 1-2: Failed Project B’s Sprints 3-4: Completed but not satisfied	This requirement is also not satisfied due to some user stories are already vague or unrealistic
	Save the results (use cases, tasks, user stories, and conflicts)	Completed	Project B’s Sprints 2: Save use cases and user stories Project B’s Sprint 3: Save tasks and conflicts	Save as JSON files
	Output the result analysis	Completed, **but not satisfied**	Project B’s Sprint 4	Save as CSV files
	Test with LLM	Partially completed	All sprints in project B	We have only test with OpenAI’s ones, due to the limited resources and creditss
Implementation – Frontend	Research and test the Streamlit library	Completed	Project A’s Sprint 3	
	Implement the panel to choose system, and LLM	Completed	Proejct B’s Sprints 3-4: Completed	
	Implement the panel to input personas	Completed	Project B’s Sprint 3: Failed Project B’s Sprint 4: Completed	

	Implement the panel to execute main.py and processing inputs	Partially completed	Project B’s Sprint 4	main.py run successfully, but the log is outputed after execution finished, not realtime (minor issue)
	Implement the Panel to download the results	Partially completed	Project B’s Sprint 4	Only results files (JSON) can be zipped and downloaded The CSV files are not included in the Frontend, and must access via Backend
Deployment	Deploy the application	Partially completed	Project B’s Sprint 4	Complete, but some minor issues as above
	User Manual	Completed	Project B’s Sprint 3	
	System Deployment Manual	Partially completed	Project B’s Sprint 3	Partially completed due to some minor issues at frontend implementation and deployment
Professional Report	Summarise the result analysis CSV files	Completed	Project B’s Sprint 4	Report is written using Overleaf, required by the Client
	Summarise the Pipeline (Phases 1 to 7)	Completed	Project B’s Sprint 4	
	Write the Report: - Introduction and Background - Methodology - Evaluation	Completed	Project B’s Sprint 4	
Project Summarizing and Handover	Get sign-off from the Client	In progress	Project B’s Sprint 4	Client still does not happy with the project

Key lessons learned

Enhance Communication skills (between the team and the client)

The team needs to improve the communication skills between the team itself and the client. Currently, we are still quite vague and unsure about the project requirements that the client has introduced during semesters

Enhance domain knowledge, especially about ALFRED-related

At the beginning of the project, the team lacked domain understanding of healthcare systems and assistive technologies like ALFRED. As a result:

- The persona creation was initially vague or unrealistic, as ALFRED's stakeholder (Caregiver/Medical Staff, Older Adult, or Developer) goals, constraints, and expectations were not accurately designed

Improving Requirements Engineering skills

Another key takeaway was recognizing our limited initial knowledge of requirement engineering methods. Concepts such as:

- User stories (definitions, clustering, generation strategies, ...)
- Use case modeling,
- Conflict identification and resolution strategies,

were unfamiliar at the project's start. We struggled to manually define realistic user stories or to anticipate conflicts between different stakeholder needs.

Found techniques (e.g., Poort and de With requirement grouping (clustering), Sadana & Liu requirement decomposition, Chentouf requirement classification) in the research papers, which should be done in the first Semester, came later in the semester and it was already too late to enhance the project deliverable

Leveraging LLMs power more Effectively

Although we used GPT-4.1, a very powerful and state-of-the-art model of OpenAI, to automate large portions of the pipeline, we underutilized its potential early in the project. For example: Initial prompts lacked precision, resulting in generic or incoherent outputs. The team was not confident in writing long prommpts guiding the model to generate domain-specific content (user story/requirement, conflicts, ...)