Testplan

Rick van Schijndel  
Open Innovation minor

**Geschiedenis**

|  |  |  |
| --- | --- | --- |
| Datum | Auteur | Wijzigingen |
| 7 januari 2020 | Rick van Schijndel | Eerste opzet |
| 9 januari 2020 | Rick van Schijndel | Uitvoering tests |
| 13 januari 2020 | Rick van Schijndel | Toevoeging gesprekken |
| 14 januari 2020 | Rick van Schijndel | Conclusies en SWOT |
|  |  |  |

Inhoudsopgave

[Context 3](#_Toc29976713)

[Structuur 4](#_Toc29976714)

[Test setup 4](#_Toc29976715)

[Observaties 7](#_Toc29976716)

[Conclusies 8](#_Toc29976717)

Gebruikte onderzoeksmethoden:

|  |  |  |
| --- | --- | --- |
| Methode | Strategie | Bladzijde |
| SWOT analysis | Library | 8 |
| Interview | Field | 4 |
| Non-functional test | Lab | 4 / 7 |
| Usability testing | Lab | 4 |
| Prototyping | Workshop | 4 |

# Context

Voor de Open Innovation minor (semester 7, richting Software Engineering) werken studenten zelfstandig of in een groep aan een project. De op te leveren producten, betrokken tooling/technieken en planning is geheel te bepalen door de studenten – mits er hier een externe product owner of stakeholder bij betrokken wordt met overzicht op het gehele traject.

In mijn project realiseer ik een chatbot voor de verschillende reserveerbare kamers op Fontys locaties. Deze chatbot maakt gebruik van persoonlijkheid, karaktereigenschappen en emoties om te achterhalen hoe gebruiker hierop reageren en of deze technologie meerwaarde bevat voor toekomstige projecten.

In dit document beantwoord ik uitgebreid de volgende deelvraag:

“Wat voor bijdrage levert een complex AI-systeem/chatbot aan een interactief gebouw?”

Wel moet daarbij opgemerkt worden dat de focus in het antwoorden van deze vraag vooral ligt in het bepalen of er **wel** of **niet** een positieve bijdrage wordt geleverd en of de emotionele complexiteit in de weg komt van de kernfunctionaliteit.

# Structuur

Het doel is simpel. Achterhalen wat de **meerwaarde** is van mijn chatbot. Heeft het een positieve invloed dat deze emoties uit en de gebruiker attent maakt op zijn/haar emoties? Of juist een negatieve invloed? Maakt de bot het reserveren van kamers eenvoudiger? Of juist onnodig complex? Het klinkt raar om pas laat in het project deze vragen te stellen, maar in mijn geval valt het juist perfect in het onderzoeksdoel van de chatbot. Eerder in het project heb ik alleen maar aannames gemaakt en (op basis van onderzoek) ingeschat wat de resultaten zullen zijn. Nu is het tijd om naar actuele reacties te kijken. Ik heb voldoende neutrale testpersonen nodig uit verschillende kanten van de doelgroep.

## Test setup

Ik wil graag één voor één personen individueel de chatbot laten testen, zo blind mogelijk. Wat betreft teststructuur vertel ik alleen het standaardscenario; de testpersoon wil graag een kamer reserveren. De rest mogen ze zelf invullen. Ik blijf erbij om te observeren hoe het gesprek verloopt en te noteren in onderstaande tabel welke acties vernomen worden.

Na het uitvoeren van de initiële testfase ga ik het testplan specifieker opstellen afhankelijk van de functionaliteiten die niet ter sprake zijn gekomen.

|  |  |  |  |
| --- | --- | --- | --- |
| Naam | Leeftijd | Gesprek in kort | Opmerkingen |
| Niels | 26 | * Niels vraagt om de kamer te reserveren * Chatbot vraagt wanneer * Niels zegt van 12 tot 1 * Chatbot zegt dat de reservering om 1 uur start (FOUT) * Niels vraagt waar hij heeft gereserveerd * Chatbot denkt dat hij een nieuwe kamer wil reserveren * Rick zegt gauw als excuus dat het oké is om de chatbot kapot te maken * Niels wil een reservering annuleren * Chatbot probeert weer te reserveren * Niels maakt een laatste reservering met duratie en starttijd * Chatbot reserveert, data klopt | De chatbot laat geen response zien wanneer er meerdere keren een vraag gesteld wordt dat niet begrepen wordt. In de back-end blijkt dat deze error veroorzaakt wordt door standaard DialogFlow gedrag en ik zal moeten kijken of ik dit kan omzeilen.  Een aantal beledigingen/negatieve opmerkingen werden niet juist opgevat door de chatbot (bv. “Kill yourself”). Deze resulteerden in uitleg over de chatbot (omdat de opmerking herkend wordt als “Tell me about yourself”.  Ook na een kleine push van mij af om erachter te komen welke functionaliteit de bot nog meer bezit is deze niet aan bod gekomen. |
| Karel | 25 | * Karel probeert eerst in het Nederlands te reserveren * Chatbot denkt dat “kamer” een naam is * Karel vraagt om te reserveren * Chatbot vraagt wanneer * Karel probeert volgende week maandag * Chatbot vraagt nu een specifieke tijd * Karel geeft een starttijd aan * Chatbot bevestigt de kamer met de correcte starttijd * Karel vraagt om de naam van de bot * Chatbot snapt niet wat er bedoeld wordt * Karel vraagt om meer info over de bot * Chatbot zegt dat het niet belangrijk is (klopt) | Positieve indruk. Karel vraagt om een kamer voor volgende week maandag. Dit wordt opgevangen door de chatbot maar deze wil nog wel een specifieke starttijd weten. Deze wordt vervolgens goed opgevangen.  Karel merkt op dat het niet uitmaakt hoe de chatbot onserieus reageert (of eigenlijk niet de informatie geeft waar om gevraagd werd) wanneer je bijvoorbeeld vraagt om meer informatie over de chatbot; het reserveren van kamers is kernfunctionaliteit maar de rest is toch bijzaak.  Ook wordt hier aangegeven (in tegenstelling tot de vorige test) dat het wel grappig kan zijn als een bot je begint te versieren – juist omdat het een bot is hecht je er geen serieuze waarde aan. |
| Niels | 20 | * Niels vraag om een kamer te reserveren * Chatbot vraagt wanneer * Niels zegt 1 uur * Chatbot bevestigt de tijd (met onnodige seconden) * Niels vraagt hoe lang de afspraak duurt. * Chatbot probeert opnieuw een kamer te reserveren. * Niels gaat vloeken. * De chatbot raakt geïrriteerd. * Niels merkt op dat het ongemakkelijk is als de chatbot begint verleidelijk te praten. * Niels gaat door met vloeken. * De chatbot wordt extreem boos. | Het valt snel op in het gesprek dat de chatbot gelimiteerd is op één kamer reserveren en deze nu niet in staat is om terug te kijken naar de laatste reservering of een lijst op kan halen van reserveringen. Dus betreft kernfunctionaliteit is deze nog zeer gelimiteerd.  Qua emoties is Niels het eens met de andere Niels dat de (extreme) uiting van karaktereigenschappen en emoties niet heel nuttig is, met als gedachtegang dat gebruikers toch weten dat ze praten met AI en dus gewoon straight-to-the-point de functionaliteit willen gebruiken. Als de bot misschien juist de persoon zou “opvrolijken” in plaats van spiegelgedrag tonen dat deze uitingen wel nut zouden hebben. |
| Marco | 56 | * Marco probeert in het Nederlands een kamer te reserveren. * Chatbot denkt dat de persoon ‘kamer’ heet. * Marco reserveert een kamer zonder starttijd of duratie. * Chatbot vraagt hoe laat de afspraak begint. * Marco geeft een startijd. * Chatbot geeft aan dat de afspraak is ingepland met de correcte starttijd. * Marco gaat nu al vloekend een kamer proberen te reserveren. * De chatbot raakt geïrriteerd maar reserveert nog steeds prima een kamer. | Marco geeft aan dat het fijn zou zijn als de chatbot aan geeft dat deze alleen Engels spreekt.  Verder geeft hij aan dat voor zijn gevoel gegenereerde antwoorden meer met een glimlach ontvangen wordt dan dat het iemand oncomfortabel maakt. Het zegt meer over de persoon die de bot gemaakt heeft.  Ook is Marco het ermee eens dat de uiting van karakter en emotie geen probleem is zolang het niet in de weg komt van vereiste functionaliteit (reserveren van deze kamer). |
| Eric | 58 | * Eric vraagt om een kamer maar maakt typfouten. * De chatbot herkent het commando niet. * Eric vraag om onmiddellijk een kamer te reserveren. * De chatbot lijkt het niet goed op gevangen te hebben dus vraagt voor de zekerheid nog om een starttijd. * Eric geeft een starttijd aan. * Nu is de reservering wel gelukt. | Het testen ging een beetje moeizaam omdat text-to-speech plotseling niet meer werkte. Dit heeft waarschijnlijk iets te maken met Chrome’s veiligheid tegen tabbladen die geluid afspelen zonder gebruikersinteractie.  Eric maakte een aantal spelfouten en het valt op dat de bot daardoor helemaal niet meer in staat is om een commando te herkennen. Dit zou aangezet kunnen worden. Maar Eric geeft wel aan dat dan het risico is dat een spellingchecker er juist weer een verkeerde intent van maakt. |
| Noah | 23 | * Noah vraagt om een kamer. * De chatbot vraagt wanneer. * Noah geeft een datum aan, geen tijd. * De chatbot vraagt om een tijd. * Noah geeft een specifieke tijd aan. * Chatbot bevestigt de tijd van de afspraak. * Noah vraagt om de naam van de chatbot. * Chatbot snapt het commando niet. * Noah probeert het nu in het Engels. * Chatbot geeft hier nu wel correct antwoord op. | Noah geeft aan dat de chatbot beter werkt dan verwacht en dat de meeste “onzin” (niet-functionele eisen) toch best goed opgevangen wordt.  Wat opviel is dat Noah ongevraagd begon te vloeken in de interactie met de chatbot, dit werd duidelijk wanneer de chatbot plotseling begon te snauwen. Dit vond Noah wel grappig (en het was voor ons beide onverwacht). Hij geeft aan dat het niet in de weg komt van het daadwerkelijk reserveren van kamers. |

## Observaties

In het kort zijn de volgende punten opgevallen tijdens de uitvoering van bovenstaande tests:

* 100% van de gebruikers waren in staat om succesvol een kamer te reserveren.
* 66.7% van de gebruikers hadden een positieve reactie op de chatbot’s uiting van emotie en geven aan dat dit niet in de weg zit van de daadwerkelijke kernfunctionaliteit.
  + 16.7% van de gebruikers heeft een negatieve reactie en is van mening dat de uiting van emotie en karakter juist hindert, maar geeft aan dat dit misschien niet het geval zou zijn als (in plaats van spiegelgedrag) de chatbot juist probeert om de gebruiker op te vrolijken.
  + De resterende 16.7% heeft een neutrale mening. De uiting van emotie en karakter conflicteert niet met de kernfunctionaliteit van de chatbot (reserveren van kamers) maar voegt verder ook geen enkele waarde toe.
* 33.3% van de gebruikers liepen tegen het probleem aan dat ze Nederlands probeerde te spreken tegen de Engelse chatbot (50% als je de laatste testpersoon meetelt die dit voor de grap probeerde). Deze limitatie van de chatbot zou ergens aangegeven moeten worden.
* 83.3% van de tests zijn foutloos verlopen (in één user tests was de bot plotseling niet in staat om audio af te spelen).

# Conclusies

Met de informatie verzameld tijdens verschillende user tests en interviews kan ik nu de volgende deelvraag beantwoorden:

“Wat voor bijdrage levert een complex AI-systeem/chatbot aan een interactief gebouw?”

Gebaseerd op de test-uitvoering en genoteerde observaties/statistieken in het vorige hoofdstuk kan ik concluderen dat het voornamelijk een **positieve** bijdrage levert – zowel de implementatie van een chatbot in plaats van een standaard touch interface als het tonen van karaktereigenschappen en emoties. Het merendeel van de tests en interviews leiden tot een positieve uitspraak over de bot.

Echter geldt hier voor een aantal personen juist het tegendeel; ze voelen zich oncomfortabel met de “menselijke” chatbot of zijn van mening dat het in de weg staat van kernfunctionaliteit. Zelf haal ik hieruit op dat het belangrijk is om deze chatbot te implementeren in een dergelijk interactief gebouw – personen zijn het gewend dat er experimentele hardware en software gebruikt wordt. In een andere omgeving (zoals een standaard middelbare school) raad ik aan om gebruikers duidelijk te informeren van de chatbot voordat gebruikers deze proberen te gebruiken voor het reserveren van kamers.

Verder brengt het AI-systeem een complicatie dat de standaard touch interface niet bevatte; niet iedere taal wordt herkend. Het is dus belangrijk om de chatbot een Engelse naam te geven en een Engels introbericht om gebruikers duidelijk te maken dat deze bijvoorbeeld geen Nederlands spreekt.

Tot slot zou ik deze informatie over de chatbot kunnen visualiseren in een simpel SWOT-diagram:

|  |  |  |  |
| --- | --- | --- | --- |
| S  Strengths | W Weaknesses | O Opportunities | T Threats |
| Unieke functionaliteit (leidt tot opportunities). | Voor nu nog geen duidelijke indicatie van welke talen de chatbot begrijpt, kan leiden tot verwarring. | Sluit perfect aan bij interactieve gebouwen met dergelijke AI-systemen als vervangers voor standaardfunctionaliteit | Kan verwarrend (of zelfs frustrerend) zijn in gebouwen waarin de chatbot de enige applicatie is met dergelijk karakter en emoties. |
| Visualiseert subtiel (m.b.v. achtergrondkleur en stem) huidige emoties lang voordat deze zouden kunnen “hinderen”. | Heeft moeite met het correct interpreteren van zinnen met spelfouten. | Gebruikers kunnen de uiting van emoties opvangen als grappig, levert een positieve bijdrage aan de algemene werksfeer. | Gebruikers kunnen de uiting van emoties opvangen als irriterend en zullen zich frustreren als er geen standaard touch interface beschikbaar is. |
| Volledig in staat om kamers te reserveren met zelfs *meer* opties dan de vorige touch interface. |  |  | Niet makkelijk te gebruiken voor sommigen, hangt enorm af van correct opvangen van Engelse uitspraken. |