

ANALISIS DAN PERANCANGAN INTERAKSI *CHATBOT REMINDER* DENGAN *USER-CENTERED DESIGN*

Ahmad Alim Akhsan¹ dan Faizah²

^{1,2}Departemen Ilmu Komputer dan Elektronika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada, Bulaksumur, Yogyakarta, 55281, Indonesia

E-mail: ¹alim.akhsan@gmail.com, ²faizah@ugm.ac.id

Abstract

The presence of technology that is not prepared to face user interaction could worsen the user experience. There are still many chatbot reminders that cannot fulfill the users' expectations. User-centered design approach can help to overcome the problems associated with the user. In this research, Google Design Sprint as a user-centered design approach was used in the analysis until testing phase. Interaction was the main problem to be analyzed and was expected to improve the user experience. Design recommendation and prototype for a new chatbot refers to the result of usability testing of four previous chatbots. Testing was done by doing usability testing of a new chatbot. A total of five different participants were used each chatbot during usability testing. The results of usability testing of the new chatbot prototype were compared to the previous chatbots. The result was the new chatbot has a SUS (System Usability Scale) score of 70.5. The score increases up to 18.4 from previous chatbots on average. Based on data obtained from usability testing and SUS score, it was concluded that interaction and appropriate selection of interfaces could improve the user experience of chatbot.

Keywords: *interaction, chatbot, user-centered design, user experience*

Abstrak

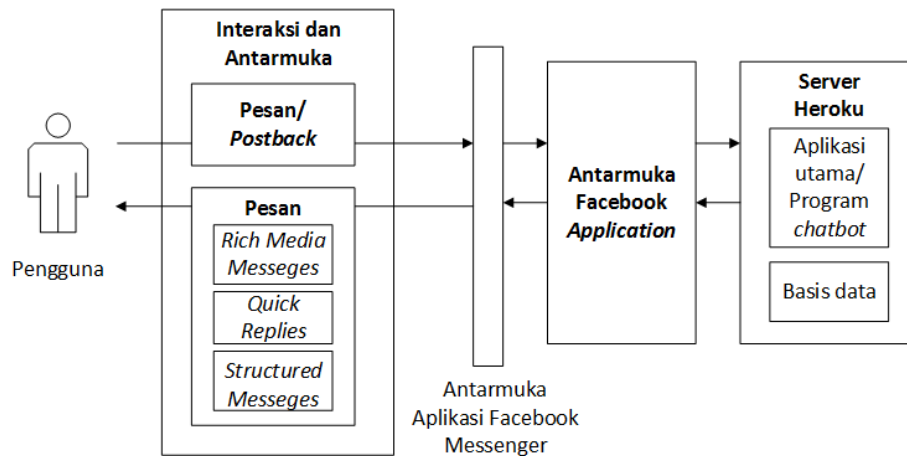
Keberadaan teknologi yang tidak dipersiapkan untuk menghadapi interaksi dengan pengguna akan membuat *user experience* menjadi buruk. Masih banyak ditemukan *chatbot reminder* yang tidak dapat memenuhi harapan pengguna. Pendekatan *user-centered design* dapat membantu dalam mengatasi masalah yang berhubungan dengan pengguna. Pada penelitian ini digunakan metode Google Design Sprint sebagai pendekatan *user-centered design* dalam tahapan analisis hingga pengujian. Interaksi merupakan masalah utama yang dianalisis dan diharapkan dapat meningkatkan *user experience*. Rekomendasi rancangan dan *prototype chatbot* baru mengacu pada hasil *usability testing* empat *chatbot* terpilih. Pengujian dilakukan dengan melakukan *usability testing* pada *prototype chatbot* baru. Sebanyak lima partisipan yang berbeda digunakan setiap *chatbot* pada saat *usability testing*. Hasil *usability testing chatbot* baru kemudian dibandingkan dengan *chatbot* lama. Diperoleh skor SUS (*System Usability Scale*) *chatbot* baru sebesar 70,5. Skor tersebut meningkat sebanyak 18,4 secara rata-rata dari *chatbot* lama. Berdasarkan data yang diperoleh saat *usability testing* dan skor SUS dapat diambil kesimpulan bahwa interaksi dan pemilihan antarmuka yang tepat dapat meningkatkan *user experience* pada *chatbot*.

Kata Kunci: *interaksi, chatbot, user-centered design, user experience*

1. Latar Belakang

Chatbot merupakan sebuah program komputer yang dapat menjalankan *intelligent conversation* dengan pengguna melalui media suara ataupun teks, sering kali dilakukan dengan percakapan yang singkat [1]. *Chatbot* bekerja dengan mengartikan pesan yang diberikan oleh pengguna, kemudian memproses maksud dari pesan tersebut, selanjutnya menentukan dan mengeksekusi apa yang perlu *chatbot* lakukan berdasarkan perintah dari pengguna, dan terakhir menyampaikan hasil dari eksekusi program kepada pengguna [2].

Pada tahun 2015, Facebook mulai mengizinkan para pengembang untuk membuat sebuah *chatbot* pada platform Messenger yaitu aplikasi kirim pesan buatan Facebook. Selain itu, Messenger menyediakan tombol dan pesan bagi *chatbot* dengan antarmuka yang kaya informasi. Hal tersebut dapat membantu memudahkan interaksi yang dilakukan pengguna dengan *chatbot*, sekaligus membuka peluang bagi para pengembang untuk merancang interaksi dan antarmuka pada *chatbot*. Setiap *chatbot* yang ada pada Messenger difokuskan pada satu fitur tertentu. Salah satu fitur yang penting untuk digunakan dalam kehidupan sehari-hari adalah me-



Gambar 1. Lingkup permasalahan

ningkat suatu pekerjaan. Namun, masih banyak orang yang mengabaikan untuk membuat dan mengorganisir pekerjaan ke dalam sebuah *reminder*.

Allen [3] menjelaskan bahwa *reminder* merupakan sesuatu ataupun pekerjaan yang diubah menjadi sekumpulan kertas atau layanan *web* yang kemudian kita percayakan untuk menyimpan hal tersebut atau bisa disebut sebagai *trusted system*. Adanya *chatbot reminder* diharapkan dapat membantu mengingatkan pekerjaan dalam kehidupan sehari-hari. Akan tetapi, masih banyak pengguna yang tidak menyukai *chatbot* karena kesulitan dalam berinteraksi dengan *chatbot*. Padahal interaksi yang dirancang untuk meningkatkan pengalaman pengguna menjadi hal penting dalam merancang sebuah *chatbot* agar pengguna tidak merasa kesulitan.

Beberapa penelitian sebelumnya telah mengkaji bagaimana performa *chatbot* [1, 4]. Kuligowska [4] membandingkan beberapa *chatbot* yang ada untuk mengetahui tingkat performanya. Namun, belum mengikutsertakan aspek dari *user experience*. Salah satu cara yang dapat dilakukan untuk memperdalam aspek tersebut adalah dengan menggunakan pendekatan *user-centered design* dimana pengguna dilibatkan pada salah satu atau lebih [5]. Penelitian ini tidak terlepas dari memanfaatkan penelitian yang sudah ada. Hasil dari penelitian Hill [6] dijadikan patokan bagaimana bahasa yang perlu digunakan dalam perancangan *chatbot*. Serta diacu beberapa evaluasi yang perlu diperhatikan dari hasil penelitian Hung [7] dalam analisis dan perancangan *chatbot* baru seperti *ease of usage*, *clarity*, *naturalness*, *friendliness*, *robustness regarding misunderstandings*, dan *willingness to use system again*.

Pada penelitian ini, akan dilakukan analisis dan perancangan interaksi *chatbot reminder* deng-

an menggunakan pendekatan *user-centered design* yang didasarkan pada hasil evaluasi empat *chatbot* lain yang telah ada. Pada bab selanjutnya akan dijelaskan tahapan metode yang dilakukan serta hasil pengujian pada beberapa skenario. Kemudian di bagian akhir akan diberikan kesimpulan sekaligus rekomendasi untuk penelitian selanjutnya.

2. Metode

Permasalahan utama yang akan diselesaikan pada penelitian ini adalah bagaimana mencari solusi untuk meningkatkan pengalaman pengguna dengan merancang interaksi dan antarmuka antara pengguna dengan *chatbot*. Lingkup permasalahan yang dibahas merupakan interaksi dan antarmuka berbalas pesan antara pengguna dengan *chatbot* dan secara singkat digambarkan pada Gambar 1.

Untuk menyelesaikan permasalahan tersebut digunakan metode *user-centered design* pada tahap awal dan akhir analisis. *User-centered design* (UCD) adalah sebuah istilah umum yang menjelaskan proses desain dimana pengguna terlibat dalam pembentukan desain [5]. Banyak yang menggambarkan bagaimana pengguna terlibat dalam proses UCD, tetapi konsep yang terpenting adalah pengguna dilibatkan dalam satu hal atau lebih. Sebagai contoh, beberapa jenis UCD menanyakan mengenai kebutuhan dan melibatkan pengguna pada waktu tertentu selama proses desain berjalan. Di sisi lain, pengguna dapat terlibat lebih dalam selama proses desain.

Diperlukan ketelitian saat melibatkan pengguna dalam proses desain, mengenai siapa saja dan bagaimana pengguna terlibat. Dapat dilihat pada Tabel 1 bagaimana pengguna terlibat dalam sebuah proses desain meliputi teknik yang digunakan, tujuan, serta berada pada tahapan desain apa teknik

TABEL 1
MELIBATKAN PENGGUNA DALAM PROSES DESAIN [5] DENGAN SEDIKIT PERUBAHAN

Teknik yang Digunakan	Tujuan	Tahapan dari Siklus Desain/Perancangan
Serangkaian wawancara dan kuisisioner	Mengumpulkan data yang berhubungan dengan serangkaian tindakan yang dilakukan pada produk	Tahap awal dalam siklus desain
<i>Usability testing</i>	Mengumpulkan data kuantitatif yang berhubungan dengan kriteria <i>usability</i>	Tahap akhir dalam siklus desain
Wawancara dan kuisisioner	Mengumpulkan data kuantitatif yang berhubungan dengan kepuasan pengguna terhadap produk	Tahap akhir dalam siklus desain

tersebut digunakan [5]. Teknik yang dijelaskan pada tabel adalah teknik yang digunakan pada penelitian ini. Tahapan penelitian ini mengacu pada enam tahapan yang digunakan untuk merancang dan mengevaluasi sebuah desain [8] dan secara terperinci digambarkan pada Gambar 2.

Tahap 1 – *Understand*

Tahap *understand* merupakan tahap pertama yang bertujuan untuk mencari tahu permasalahan pada *chatbot* yang sudah ada di pasaran. Pengumpulan data permasalahan dilakukan dengan *usability testing* serta analisis terhadap rancangan *chatbot* lama yang terpilih. Untuk menjalankan *usability testing*, diperlukan persiapan dengan memilih *chatbot* dan mencari partisipan, merancang skenario, menyusun sistem penilaian, dan menentukan cara pengumpulan data *usability testing*.

Pencarian *chatbot* dilakukan melalui beberapa situs, yaitu *botlist.co*, *botpages.com*, dan melalui fitur pencarian Messenger. Dalam penelitian ini, tidak semua *chatbot* digunakan sehingga perlu dilakukan pemilihan seperti yang dilakukan Kuligowska [4]. Kategori penilaian dalam pemilihan *chatbot* dibagi menjadi dua hal. Pertama, bagaimana kemampuan *chatbot* dalam memahami konteks yang dikirimkan oleh partisipan. Kedua, kemampuan *chatbot* yang berhubungan dengan fungsi utama *reminder*. Dari penilaian dua kategori tersebut, diperoleh empat *chatbot* yang digunakan pada penelitian ini, yang selanjutnya disebut dengan *chatbot* lama. Nama *chatbot* terpilih dapat dilihat pada Lampiran Tabel I.

Persiapan *usability testing* dilakukan dengan mempersiapkan sistem penilaian yaitu SUS, kuisisioner, dan membuat skenario *usability testing*. *System Usability Scale* (SUS) merupakan pengukuran *usability* yang sederhana. SUS menyediakan sepuluh pernyataan berbentuk kuisisioner dalam pengukuran yang dapat memberikan gambaran secara luas mengenai penilaian *usability* [9]. SUS menggunakan skala Likert yang terbagi menjadi lima, yaitu 1: sangat tidak setuju, 2: tidak setuju, 3: netral, 4: setuju, dan 5: sangat setuju.

Pada Lampiran Tabel II dipaparkan isi dari kuisisioner SUS yang sudah diterjemahkan ke dalam Bahasa Indonesia dan disesuaikan dengan peneliti-

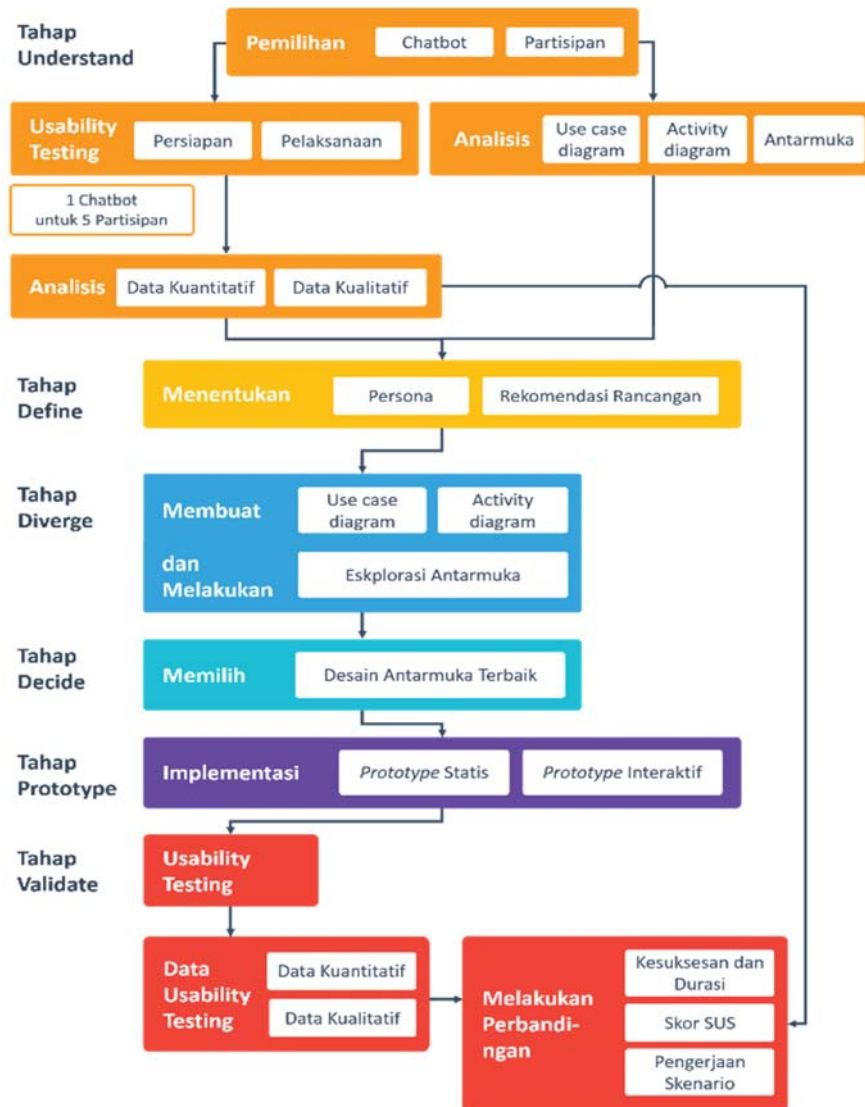
an ini. Dapat dilihat pernyataan yang terdapat pada kuisisioner SUS sudah mencakup berbagai aspek dari *usability* suatu sistem.

Usability testing merupakan salah satu proses desain dalam UCD yang dilakukan untuk mengetahui apakah sebuah produk sesuai dengan kriteria *usability* dengan melibatkan pengguna dalam pengujian. Pada Lampiran Tabel III tertera skenario pekerjaan yang digunakan pada penelitian ini. Skenario ditentukan setelah dilakukan percobaan pada beberapa pengguna di luar partisipan, kemudian dipilih skenario yang dinyatakan cocok dan dapat dilanjutkan pada seluruh *chatbot*.

Setelah rancangan skenario pekerjaan dan kuisisioner SUS dianggap layak untuk digunakan pada *usability testing*, selanjutnya adalah pelaksanaan *usability testing*. Pelaksanaan *usability testing* dibagi ke dalam tiga tahapan: 1) Tahap pertama yaitu partisipan mengisi kuisisioner yang meliputi data pribadi, data penggunaan aplikasi kirim pesan, data perilaku menggunakan internet, data perilaku dan tanggapan mengenai *reminder*. Data tersebut digunakan untuk menggambarkan demografi pengguna dengan menggunakan sampel data dari lima partisipan pada masing-masing *chatbot* lama dan baru; 2) Tahap kedua adalah menjalankan skenario. Pada tahap ini partisipan diminta untuk mengerjakan skenario yang telah dirancang dengan menggunakan *chatbot* pada perangkat *smartphone*; 3) Di tahap terakhir dilakukan pengisian kuisisioner SUS, saran, dan tanggapan partisipan terhadap *chatbot* yang digunakan.

Pelaksanaan *usability testing* dibagi menjadi tiga tahapan. Tahap pertama, yaitu partisipan mengisi kuisisioner yang meliputi data pribadi untuk menggambarkan demografi pengguna. Tahap kedua adalah menjalankan skenario *usability testing* menggunakan *chatbot* lama dengan merekamnya pada perangkat *smartphone*. Tahap terakhir adalah pengisian kuisisioner SUS, dan tanggapan partisipan terhadap *chatbot* yang digunakan.

Setelah *usability testing* dilakukan, diperoleh data berupa rekaman partisipan dalam menjalankan skenario. Data yang didapatkan selanjutnya dianalisis sehingga diperoleh gambaran permasalahan dan kebutuhan dari *chatbot*. Selain data *usability testing*, dilakukan juga analisis terhadap rancangan, antara lain *use case diagram*, *activity diagram*,



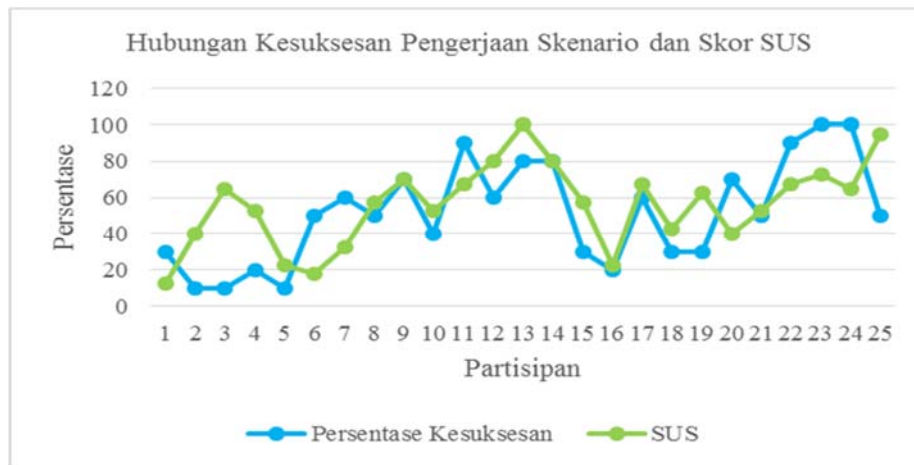
Gambar 2. Tahapan penelitian

dan antarmuka. Hasil analisis dari tahap ini akan digunakan pada tahap selanjutnya, yaitu tahap *define*.

Analisis data kuantitatif dan data kualitatif didapatkan melalui data yang dikumpulkan saat melakukan *usability testing*. Analisis pada data kuantitatif dibagi menjadi tiga bagian, yaitu analisis untuk kesuksesan pengerjaan skenario, durasi pengerjaan skenario, dan perhitungan kuesioner SUS. Sedangkan data kualitatif merupakan data perilaku dan tanggapan partisipan dalam pengerjaan skenario pada *usability testing*, beserta kuesioner tanggapan dan saran. Pada Lampiran Gambar I diperlihatkan contoh antarmuka dua *chatbot* lama yaitu Big-Bot dan ReMind ketika menghapus *reminder*.

Tahap 2 – Define

Pekerjaan utama yang dilakukan pada tahap *define* adalah menentukan strategi yang akan dilakukan setelah melihat permasalahan dari tahap *understand*. Strategi dimulai dengan menentukan persona sebagai target pengguna dan menentukan rekomendasi rancangan pada *chatbot* baru. Persona yang digunakan sebagai target pengguna terdiri dari tiga macam persona yang diperoleh dari hasil *usability testing*. Ketiga persona tersebut memiliki tujuan yang berbeda dalam menggunakan *reminder*, antara lain: 1) Pengguna yang menyukai setiap kegiatan terorganisir. Biasanya pengguna yang seperti ini menganggap bahwa membuat dan menggunakan *reminder* adalah hal penting dan telah menjadi ke-



Gambar 3. Hubungan Kesuksesan Pengerjaan Skenario dan Skor SUS

butuhan; 2) Pengguna yang membutuhkan *reminder* pada kegiatan tertentu; 3) Pengguna yang berada pada situasi diharuskan menggunakan *reminder*.

Secara garis besar, rekomendasi disorot dari BigBot, ReMind dan Reminders. Reminders memiliki pengalaman yang baik dengan partisipan dikarenakan memiliki kemampuan dalam mengatasi interaksi penambahan *reminder* yang bervariasi dan penyajian pesan yang lebih mudah dipahami. Di sisi lain, BigBot tidak memiliki pengalaman pengguna yang baik seperti Reminders, tetapi kemampuan BigBot dalam mengatasi masukan berupa teks menjadi poin menarik untuk rekomendasi. Adapun ReMind berperan dalam membuka ide untuk mempersingkat langkah aktivitas menghapus *reminder* sehingga muncul aktivitas *undo* pada *chatbot* baru.

Rekomendasi aktivitas menambah *reminder* diadopsi dari Reminders yang dapat mengatasi varian interaksi. Begitu pula dengan daftar *reminder* yang disajikan secara sederhana oleh Reminders. Sebaliknya, direkomendasikan mengadopsi kemampuan BigBot dengan melakukan modifikasi dalam mengatasi masukan berupa teks pada aktivitas menghapus, menjadwalkan ulang, menunda, dan menandai selesai.

Tahap 3 – Diverge

Diverge merupakan tahap untuk mendapatkan ide sebanyak mungkin dalam merancang sebuah produk dengan mengacu dari hasil analisis pada tahap *define*. Eksplorasi antarmuka yang dihasilkan pada tahap ini berbentuk rancangan *low-fidelity* dalam media kertas. Pada tahap ini, dirancang *use case diagram* dan *activity diagram chatbot* baru. Selain itu, dilakukan juga eksplorasi terhadap antarmuka yang bertujuan untuk mendapatkan rancangan sebanyak mungkin.

Tahap 4 – Decide

Decide merupakan tahap untuk memilih rancangan terbaik dari ide yang dihasilkan pada tahap *diverge*. Adapun cara untuk memilih ide rancangan terbaik adalah dengan mengevaluasi ide berdasarkan *usability heuristic* [10] dan petunjuk dari *platform Messenger*. Didukung pula dengan beberapa rujukan mengenai bagaimana bahasa yang perlu digunakan oleh *chatbot* [6].

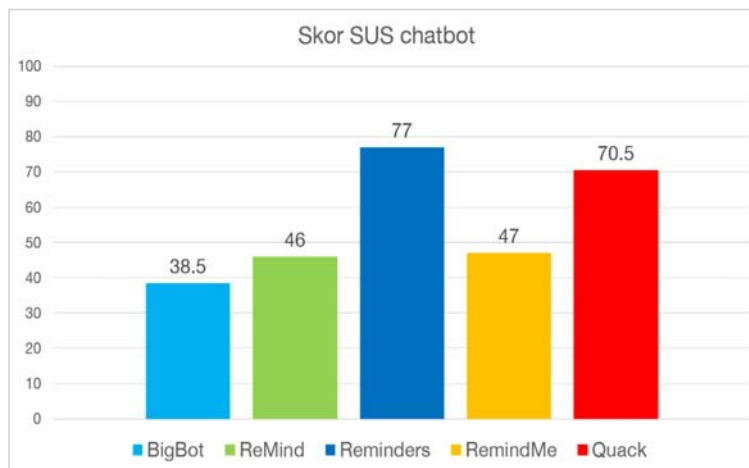
Pola desain antarmuka ketika mengatasi masukan berupa teks dalam aktivitas menghapus *reminder* juga digunakan pada aktivitas menjadwalkan ulang, menunda, dan menandai selesai pada *reminder*. Ketika pengguna mengirimkan teks yang berisi kata kunci menghapus, *chatbot* akan memunculkan *reminder* yang dimaksud jika *reminder* berhasil ditemukan. Di samping itu, tombol yang dimunculkan berfokus pada tujuan aktivitas, yaitu menampilkan tombol “Remove” untuk aktivitas menghapus. Pada Lampiran Gambar II diperlihatkan desain antarmuka aktivitas menghapus *reminder*.

Tahap 5 – Prototype

Prototype merupakan tahap untuk melakukan implementasi rancangan ke dalam bentuk *prototype* dari ide terbaik yang telah dipilih pada tahap *decide*. *Prototype* yang dihasilkan pada tahap ini berupa *high-fidelity prototype* agar dapat digunakan dalam berinteraksi dengan pengguna [11].

Tahap 6 – Validate

Pokok utama pada tahap *validate* yaitu dilakukan pengujian terhadap *prototype chatbot* baru yang telah diimplementasikan. Pengujian dilakukan untuk mencari tahu apakah rancangan *chatbot* baru ber-



Gambar 4. Skor SUS Chatbot

hasil meningkatkan pengalaman pengguna atau tidak. Pengumpulan data mengikuti metode *usability testing* yang digunakan pada tahap *understand*.

Komponen *usability testing* yang digunakan kali ini masih menggunakan skenario dan kuesioner pendapat serta saran yang sama dengan tahap *understand*, tetapi memiliki partisipan yang berbeda sehingga *feedback* yang diterima tidak terganggu dengan hasil *chatbot* lama.

3. Hasil dan Pembahasan

Pengujian dilakukan untuk mencari tahu apakah rancangan *chatbot* baru berhasil meningkatkan pengalaman pengguna atau tidak. Pengumpulan data mengikuti metode *usability testing* yang digunakan pada tahap *understand*. Komponen *usability testing* yang digunakan kali ini masih menggunakan skenario dan kuesioner pendapat serta saran yang sama dengan tahap *understand*, tetapi memiliki partisipan yang berbeda sehingga *feedback* yang diterima tidak terganggu dengan hasil *chatbot* lama.

Kesuksesan dan Durasi Pengerjaan Skenario

Pada bagian ini dijelaskan mengenai perbandingan kesuksesan dan durasi pengerjaan skenario *chatbot* lama dan baru. Kesuksesan pengerjaan skenario dibagi menjadi tiga kriteria, antara lain: 1) Sukses jika partisipan dapat menjalankan skenario sesuai dengan yang ditetapkan oleh penulis pada masing-masing *chatbot* per skenario tanpa hambatan yang ditandai dengan warna hijau; 2) Sukses namun kesulitan jika partisipan berhasil menjalankan skenario namun mengalami hambatan yang ditandai dengan warna kuning; 3) Gagal jika partisipan tidak berhasil ditandai dengan warna merah.

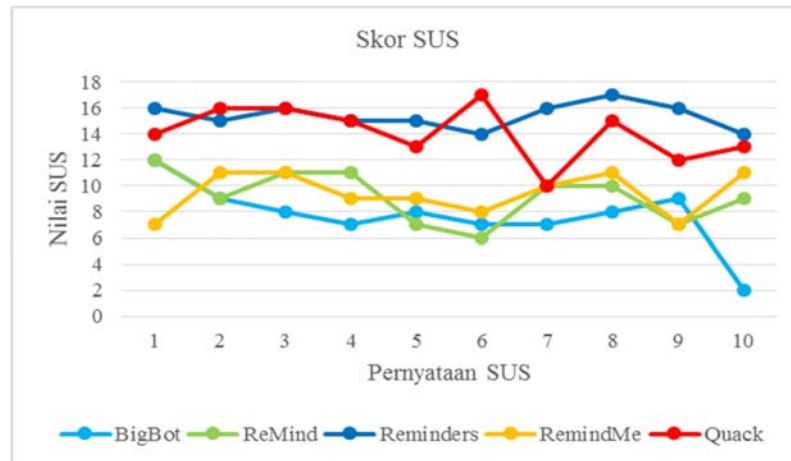
Adapun sel pada tabel yang ditandai dengan

tanda minus (-) berarti skenario yang ada tidak dikerjakan oleh partisipan. Hal tersebut dapat disebabkan oleh skenario yang gagal sebelumnya sehingga skenario selanjutnya tidak dapat dikerjakan atau *chatbot* mengalami masalah saat *usability testing* berlangsung. Ringkasan data kesuksesan dan durasi pengerjaan skenario ditampilkan pada Lampiran Tabel IV. Satuan durasi pada tabel disajikan dalam detik. Identitas partisipan dilambangkan dengan P1, P2, hingga P25.

Hasil *usability testing* memperlihatkan bahwa tingkat kesuksesan pada *chatbot* baru bernama Quack mengalami peningkatan dibandingkan dengan *chatbot* lama. Dengan membandingkan total 50 pengerjaan skenario pada masing-masing *chatbot* memperlihatkan tingkat kesuksesan skenario *chatbot* baru sebesar 78% sukses dikerjakan, sebesar 4% sukses namun kesulitan, dan sebesar 18% gagal dikerjakan. Kesuksesan tersebut naik 32,5% dibandingkan dengan rata-rata skenario yang berhasil pada *chatbot* lama. Perolehan tersebut diperlihatkan pada Lampiran Tabel V.

Persentase keberhasilan yang lebih tinggi pada *chatbot* baru merupakan pengaruh dari hasil analisis dan rancangan *use case diagram*, *activity diagram*, serta desain antarmuka dari *usability testing chatbot* lama. Selain itu, rancangan interaksi berpengaruh pada aktivitas yang menjadi sorotan permasalahan yaitu menambah, melihat daftar, dan menghapus *reminder*.

Fenomena yang ditemukan adalah pada skenario gagal dengan durasi lama dan singkat, sukses kesulitan dengan durasi lama, dan skenario berhasil dengan durasi lama. Skenario yang terdapat pada tabel dipilih berdasarkan durasi yang memiliki perbedaan signifikan dibandingkan pada *chatbot* yang sama. Ditemukan bahwa pengalaman dan tingkat frustrasi partisipan serta tingkat kemampuan *chat-*



Gambar 5. Perincian Skor SUS

bot menjadi faktor durasi dan kesuksesan yang bervariasi. Detil dari fenomena yang ditemukan digambarkan pada Lampiran Tabel VI.

Skor SUS

Untuk memperoleh skor SUS yang dapat dijadikan acuan, perlu dilakukan perhitungan sesuai dengan rumusan yang ada. Skor akhir SUS mempunyai rentang nilai 0–100, sedangkan nilai setiap pernyataan mempunyai rentang nilai 0–4. Skor diolah dengan menghitung nilai pada pernyataan dengan urutan angka ganjil dengan $(x_i - 1)$, sedangkan urutan angka genap dengan $(5 - x_i)$. Kemudian skor dari setiap pernyataan dijumlahkan dan mengalikan hasil jumlahan dengan 2,5 untuk mendapatkan nilai tertinggi 100. Seperti yang tergambar pada persamaan(1) dimana U_i mengacu kepada nilai pernyataan ke- i yang bertambah mulai dari satu hingga lima. Dimana n berarti banyaknya pernyataan [12].

$$SUS = 2,5 \times \left[\sum_{n=1}^5 (U_{2n-1} - 1) + (5 - U_{2n}) \right] \quad (1)$$

Pengisian kuesioner SUS dilakukan setelah partisipan menjalankan skenario yang diberikan. Melihat hubungan kesuksesan pengerjaan skenario setiap partisipan dengan skor SUS ditemukan bahwa sebanyak 36% partisipan yang memiliki persentase kesuksesan yang rendah cenderung memberikan skor SUS yang rendah (rata-rata persentase kesuksesan = 51,6% dan rata-rata skor SUS = 55,8). Persentase keberhasilan partisipan dihitung dari total sepuluh skenario yang dijalankan masing-masing partisipan. Hubungan antara persentase kesuksesan dan skor SUS yang diberikan digambarkan pada Gambar 3. Aksis X pada gambar merupakan identitas partisipan, sedangkan aksis Y

merupakan persentase kesuksesan dan representasi dari skor akhir SUS.

Nilai yang ditampilkan pada Gambar 4 adalah nilai rata-rata dari skor SUS masing-masing *chatbot*. Dapat dilihat bahwa skor SUS *chatbot* baru mengalami peningkatan dari tiga *chatbot* lama yaitu BigBot, ReMind, dan RemindMe, sedangkan masih lebih kecil dibandingkan dengan skor SUS Reminders. Hal tersebut disebabkan oleh *chatbot* baru yang masih berbentuk *prototype*, padahal *chatbot* baru lebih unggul secara rancangan dan tanggapan dari partisipan. Akan tetapi, partisipan lebih mudah berinteraksi dengan data yang sebenarnya sehingga beberapa partisipan merasa kebingungan saat melihat daftar ataupun keterangan waktu *reminder* pada *chatbot* baru, dimana data yang sebenarnya tidak diterapkan pada *prototype* sebagai batasan masalah.

Peningkatan skor SUS pada *chatbot* baru dapat dilihat pada Gambar 5 yang merupakan gambaran nilai SUS setiap pernyataan masing-masing *chatbot* setelah diolah dari hasil penilaian SUS. Aksis X pada gambar merupakan pernyataan SUS, sedangkan aksis Y merupakan nilai SUS yang telah diolah sesuai dengan urutan pernyataan. Dapat dilihat pada gambar bahwa skor SUS *chatbot* baru dipengaruhi oleh perubahan nilai SUS pada pernyataan 2 dan 6. Sesuai dengan Lampiran Tabel I, isi dari pernyataan 2 dan 6 adalah “Banyak hal kompleks yang tidak diperlukan pada *chatbot* ini” dan “Banyak hal yang tidak konsisten dari *chatbot* ini”. Dari nilai yang tinggi pada dua pernyataan tersebut dapat disimpulkan bahwa rancangan *chatbot* baru dirasa efektif oleh partisipan.

Perbandingan Skenario

Pada bagian ini akan dibahas mengenai perbandingan tanggapan partisipan pada *chatbot* yang dipe-

TABEL 2
RINGKASAN Pengerjaan Skenario Enam

Nama Chatbot	Langkah ke-			Persentase Kesuksesan (%)			Durasi (detik)
	Best	Avg.	Worst	S	SK	G	
BigBot	1	1	2	60	0	40	25
ReMind	2	2	3	40	40	20	57,8
Reminders	2	3	5	40	40	20	41,4
RemindMe	2	3	5	40	40	20	44,8
Quack	2	2	2	100	0	0	28,4

Keterangan:

Avg. = Rata-rata, S = Sukses, SK = Sukses tetapi kesulitan, dan G = Gagal.

roleh saat melakukan *usability testing*. Namun, dikarenakan keterbatasan jumlah halaman maka akan dijelaskan perbandingan skenario enam sebagai contoh.

Pada skenario enam partisipan diminta untuk menghapus *reminder* yang telah ditambahkan pada skenario empat. Hasil dari 20 skenario yang dikerjakan pada *usability testing chatbot* lama, ditemukan sebanyak 40% partisipan mencoba menghapus *reminder* menggunakan teks, tetapi saat *usability testing* hanya satu partisipan yang menggunakan teks pada *chatbot* baru. Interaksi partisipan yang berbeda menghasilkan variasi kesuksesan skenario ini yang diperlihatkan pada Tabel 2. Kesuksesan *chatbot* baru yang mencapai 100% dipengaruhi oleh antarmuka pada tombol dan daftar *reminder*.

Chatbot BigBot dapat menangani aktivitas menghapus *reminder* menggunakan teks, tetapi judul *reminder* yang ingin dihapus oleh partisipan haruslah sama dengan *reminder* yang ada atau sesuai dengan nomor urut *reminder* pada daftar *reminder*. Hal tersebut menyulitkan partisipan yang tidak hafal dengan nama ataupun nomor urut *reminder*. Di sisi lain, RemindMe hanya menyediakan pilihan dengan label “Mark as done” dan “Change time” pada daftar *reminder* dan tidak terdapat label dengan arti menghapus. RemindMe menganggap “Mark as done” juga termasuk menghapus karena secara teknis meniadakan *reminder* dari daftar. Namun, tidak semua partisipan mengetahui hal tersebut dan menyulitkan dalam menghapus *reminder*.

Pada BigBot, *reminder* yang dihapus tidak mendapatkan konfirmasi sehingga partisipan tidak dapat mencegah jika salah menghapus *reminder*. Di samping itu, ReMind akan melakukan konfirmasi ketika menghapus *reminder* untuk menghindari partisipan melakukan kesalahan. Pada Lampiran Gambar III digambarkan interaksi dalam skenario enam.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa tingkat kesuksesan pada

chatbot baru lebih tinggi 32,5% dibandingkan dengan *chatbot* lama secara rata-rata. Skor SUS *chatbot* baru sebanyak 32 lebih unggul dari BigBot, sebanyak 24,5 lebih unggul dari ReMind, dan sebanyak 23,5 lebih unggul dari RemindMe. Namun, masih kalah sebanyak 6,5 dengan Reminders. Secara rata-rata skor SUS *chatbot* baru meningkat sebanyak 18,4 dibandingkan *chatbot* lama. Hal tersebut memberikan gambaran bahwa pendekatan *user-centered design* baik digunakan untuk menganalisis permasalahan pada produk yang berhubungan langsung dengan pengguna. Serta pemilihan antarmuka yang tepat dapat meningkatkan *user experience chatbot* yang menggunakan antarmuka.

Saran yang dapat diberikan untuk penelitian selanjutnya adalah penambahan jumlah partisipan dalam *usability testing*. Selain itu pengujian *usability testing* selanjutnya dapat menggunakan *chatbot* yang sudah jadi, bukan *prototype*. Serta penggunaan pustaka untuk membaca keterangan tanggal dan waktu di dalam pesan dan basis data sebagai penyimpanan *reminder* yang ditambahkan.

Referensi

- [1] Tillotson, K. D., “Implementation, Analysis, and Evaluation of A Humanized Information Retrieval ChatBot,” *Tesis*, The Faculty of The Computer Science Department, Southern University and A&M College, Johor, 2012.
- [2] Shawar, B. A. and Atwell, E., “Chatbot: are they really useful?,” *LDV-Forum 2007*, vol. 22, no. 1, pp. 29-49, 2007.
- [3] Allen, D., “Getting things done: the art of stress-free productivity,” New York: Penguin Books, 2015.
- [4] Kuligowska, K., “Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents,” *Professionals Center for Business Research*, vol. 2, no. 0202.22, pp. 1-16, 2015.
- [5] Abras, C., et al., “User-centered design”, *Bainbridge, W. Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, vol. 37, no. 4, pp. 445-446, 2004.
- [6] Hill, J., et al., “Real conversations with artificial intelligence: a comparison between human-human online conversations and human-chatbot conversations”, *Computers in Human Behavior*, vol. 49, pp. 245-250, 2015.
- [7] Hung, V., et al., “Towards a method for evaluating naturalness in conversational dialog systems”, *IEEE International Conference of Systems, Man, and Cybernetics*, San Antonio, pp. 1236-1241, 2009.
- [8] Knapp, J., et al., “Sprint: how to solve big problems and test new ideas in just five days”, New York: Simon & Schuster, Inc., 2016.

- [9] Sauro, J., "A practical guide to the system usability scale", *Measuring Usability LLC*, Denver, 2011.
- [10] Nielsen, J., and Molich, R., "Heuristic evaluation of user interfaces", *Proc. ACM CHI '90 Conf.* Seattle, WA, pp. 249-256, 1990.
- [11] Walker, M., Takayama, L., and Landay, J. A., "High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 46, no. 5, pp. 661-665, 2002.
- [12] Harrati, N., et al., "Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis", *Computers in Human Behavior*, vol. 61, pp. 463-471, 2016.

Lampiran

LAMPIRAN TABEL I
DAFTAR CHATBOT LAMA

Nama Chatbot	ID Akun	Halaman Facebook
BigBot	@myBigBot	https://facebook.com/myBigBot/
ReMind	@mindnodes.remind	https://facebook.com/mindnodes.remind/
Reminders	@remindersyellow	https://facebook.com/remindersyellow/
RemindMe	@botremindme	https://facebook.com/botremindme/

LAMPIRAN TABEL II
KUISIONER SUS (SYSTEM USABILITY SCALE)

No	Kuisisioner SUS	Nilai SUS				
1.	Saya tertarik untuk sering menggunakan chatbot ini.	1	2	3	4	5
2.	Banyak hal kompleks yang tidak diperlukan pada chatbot ini.	1	2	3	4	5
3.	Chatbot ini mudah untuk digunakan.	1	2	3	4	5
4.	Saya membutuhkan bantuan dari orang lain ketika menggunakan chatbot ini.	1	2	3	4	5
5.	Fitur-fitur dari chatbot ini berjalan dengan baik.	1	2	3	4	5
6.	Banyak hal yang tidak konsisten dari chatbot ini.	1	2	3	4	5
7.	Menurut saya, chatbot ini akan mudah digunakan oleh banyak orang.	1	2	3	4	5
8.	Saya menemukan chatbot ini sulit untuk digunakan.	1	2	3	4	5
9.	Saya merasa nyaman menggunakan chatbot ini.	1	2	3	4	5
10.	Saya perlu mempelajari banyak hal sebelum menggunakan chatbot ini.	1	2	3	4	5

LAMPIRAN TABEL III
SKENARIO USABILITY TESTING

No	Tujuan Aktivitas	Deskripsi Skenario
1.	Menambah reminder dengan waktu tertentu	"Teman kerjamu memberitahu bahwa besok kamu memiliki jadwal meeting jam 8 pagi. Gunakanlah chatbot untuk mengingatkanmu tentang jadwal meeting tersebut."
2.	Menambah reminder dengan durasi tertentu	"Andi memintamu untuk meneleponnya dalam lima menit lagi. Agar tidak lupa, kamu menggunakan chatbot untuk mengingatkanmu."
3.	Melihat daftar reminder	"Budi mengajakmu jogging di hari <...> pagi. Sebelum kamu menyetujuinya, kamu perlu memastikan apakah kamu punya kegiatan lain atau tidak di hari <...> pagi tersebut. Periksa chatbotmu untuk memastikannya."
4.	Menambah reminder dengan nama hari	"Ternyata kamu tidak mempunyai kegiatan di hari <...> pagi. Agar tidak lupa, kamu menggunakan chatbot untuk mengingatkanmu tentang jogging di <...> pagi bersama Budi."
5.	Menunda reminder dengan durasi tertentu	"Sudah lima menit berlalu, kamu diingatkan oleh chatbot untuk menelepon Andi. Ternyata, Andi belum siap untuk ditelepon, lalu dia meminta untuk menundanya dua menit lagi. Sekarang, tundalah reminder tersebut selama dua menit."
6.	Menghapus reminder	"Sekarang sudah hari <...>, tiba-tiba Budi memberitahu bahwa dia sedang sakit dan tidak bisa jogging esok hari. Mengetahui hal tersebut, kamu membatalkan kegiatan jogging di remindermu."
7.	Menandai selesai pada reminder	"Bayangkan sekarang kamu sudah menelepon Andi, kemudian kamu memberi tanda selesai pada reminder tersebut."
8.	Menambah reminder dengan tanggal tertentu	"Dosen pembimbingmu ingin bertemu denganmu pada tanggal <...>, agar tidak lupa, kamu menggunakan chatbot untuk mengingatkanmu."
9.	Menjadwal ulang reminder	"Ternyata keesokan harinya, dosen pembimbingmu ingin mengganti tanggal pertemuan itu menjadi tanggal <...>, lalu kamu mengubah tanggal reminder tersebut."
10.	Menambah dua reminder	"Kamu ingin diingatkan untuk mengerjakan tugas jam 7 malam. Setelah itu, kamu ingin diingatkan untuk minum obat jam 9 malam. Gunakanlah chatbot untuk mengingatkanmu."

LAMPIRAN TABEL IV
PERBANDINGAN PERSENTASE KESUKSESAN PENERJAJAN SKENARIO ANTARA *CHATBOT* LAMA DAN BARU

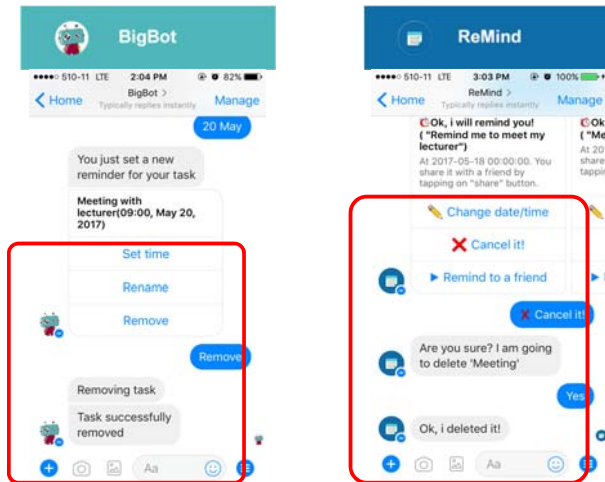
Nama <i>Chatbot</i> Lama	Persentase (%)			Persentase Perbandingan Dengan <i>Chatbot</i> Baru (%)		
	S	SK	G	S	SK	G
BigBot	18	18	64	+60	-14	-46
ReMind	56	16	28	+22	-12	-10
Reminders	68	12	20	+10	-8	-2
RemindMe	40	30	30	+38	-26	-12
Rata-rata	45,5	19	35,5	+32,5	-15	-17,5

Keterangan:

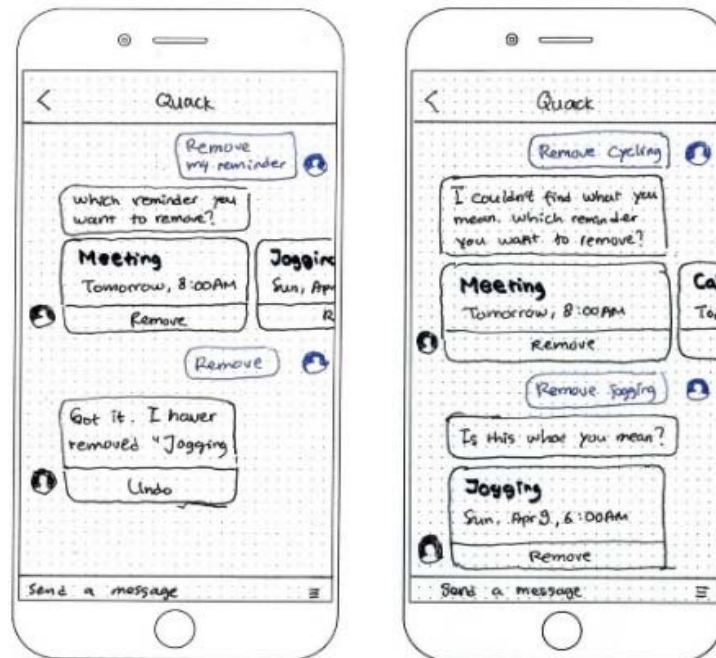
S = Sukses, SK = Sukses tetapi kesulitan, dan G = Gagal.

LAMPIRAN TABEL V
FENOMENA PENERJAJAN SKENARIO

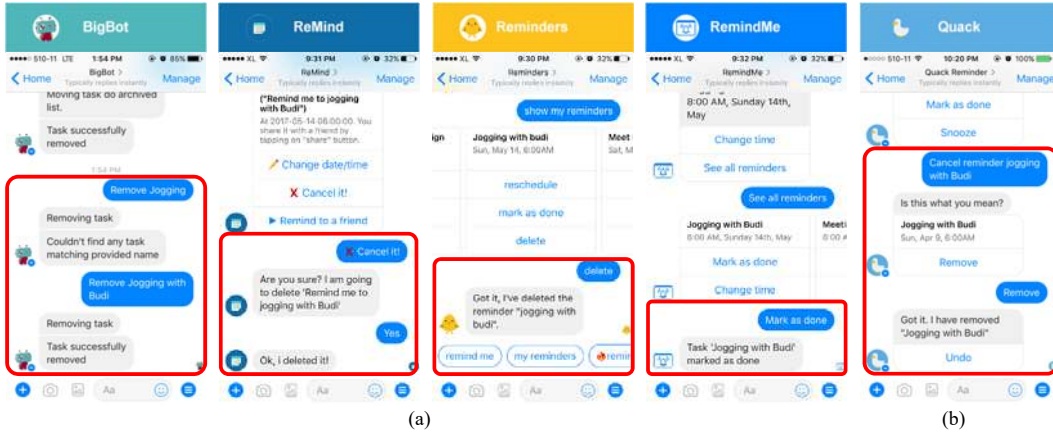
Fenomena	Terjadi Pada	Detil
Gagal dengan durasi lama	Skenario 1 <i>Chatbot</i> Partisipan P1	a. Skenario yang dikerjakan berlangsung selama 318 detik.
		b. Partisipan mengalami durasi pengerjaan yang lebih lama disebabkan oleh <i>chatbot</i> yang tidak memahami format waktu yang dituliskan oleh partisipan, seperti "8 o'clock", dimana kejadian ini berlangsung selama 110 detik.
		c. Partisipan mencoba hingga dua kali untuk mengubah waktu <i>reminder</i> karena <i>chatbot</i> salah dalam memahami keterangan waktu yang diberikan, dimana kejadian ini berlangsung selama 94 detik.
		d. Kejadian tersebut membuat partisipan frustrasi hingga memilih untuk tidak melanjutkan skenario.
	Skenario 2 <i>Chatbot</i> BigBot Partisipan P5	a. Skenario yang dikerjakan berlangsung selama 127 detik.
		b. Partisipan menghabiskan waktu selama 62 detik mencoba memahami pesan yang diberikan oleh <i>chatbot</i> saat menanyakan waktu yang diinginkan.
	Skenario 3 <i>Chatbot</i> Reminders Partisipan P13	c. Partisipan gagal menyelesaikan skenario disebabkan waktu yang dicantumkan tidak sesuai dengan skenario yang diminta.
		a. Skenario yang dikerjakan berlangsung selama 194 detik.
	Skenario 8 <i>Chatbot</i> RemindMe Partisipan P18	b. Partisipan menghabiskan waktu selama 105 detik dikarenakan salah memahami pesan yang dikirimkan oleh <i>chatbot</i>
		c. Partisipan menganggap <i>chatbot</i> menanyakan waktu <i>reminder</i> yang ingin dimunculkan oleh partisipan, sedangkan <i>chatbot</i> sebenarnya menanyakan waktu untuk menambahkan <i>reminder</i> .
a. Skenario yang dikerjakan berlangsung selama 238 detik.		
b. Partisipan mengharapkan <i>chatbot</i> langsung menambahkan <i>reminder</i> ketika keterangan waktu dicantumkan di awal kalimat, tetapi <i>chatbot</i> tidak membaca keterangan waktu di awal kalimat. Kejadian ini berlangsung selama 56 detik.		
Gagal dengan durasi singkat	Skenario 1 <i>Chatbot</i> RemindMe Partisipan P19	c. Partisipan menghabiskan waktu selama 74 detik untuk memahami format waktu yang diminta oleh <i>chatbot</i> .
		a. Skenario yang dikerjakan berlangsung selama 7 detik.
	Skenario 7 <i>Chatbot</i> ReMind Partisipan P6	b. Partisipan menambahkan <i>reminder</i> dengan judul yang tidak sesuai dengan skenario, akan tetapi partisipan tidak menyadari hal tersebut.
b. Partisipan menunda <i>reminder</i> menggunakan teks, dimana <i>chatbot</i> hanya membalas dengan "OK!".		
Sukses tetapi kesulitan dengan durasi lama	Skenario 8 <i>Chatbot</i> RemindMe Partisipan P19	c. Partisipan menganggap <i>reminder</i> berhasil ditunda.
		a. Skenario yang dikerjakan berlangsung selama 295 detik.
Sukses dengan durasi lama	Skenario 2 <i>Chatbot</i> BigBot Partisipan P3	b. Partisipan menghabiskan waktu selama 229 detik untuk menambahkan <i>reminder</i> , tetapi format yang dimasukkan tidak dimengerti oleh <i>chatbot</i> .
		a. Skenario yang dikerjakan berlangsung selama 71 detik.
		b. Partisipan menghabiskan waktu selama 53 detik untuk memilih kata yang kalimat untuk meminta <i>chatbot</i> menambahkan <i>reminder</i> . Seperti menggunakan kata "in" atau tidak.



Lampiran Gambar I. Antarmuka BigBot dan ReMind ketika menghapus reminder



Lampiran Gambar II. Antarmuka menghapus reminder



Lampiran Gambar III. Interaksi pada skenario enam: (a) Keempat chatbot lama, (b) Chatbot baru

LAMPIRAN TABEL VI
RINGKASAN KESUKSESAN DAN DURASI Pengerjaan skenario (dalam detik)

ID	Skenario									
	1	2	3	4	5	6	7	8	9	10
BigBot										
P1	318	24	118	68	66	31	12	111	163	136
P2	37	41	61	156	-	30	-	48	-	243
P3	121	71	104	80	45	29	-	41	61	127
P4	155	82	130	80	-	10	-	173	40	148
P5	100	127	51	-	-	-	-	50	44	100
ReMind										
P6	49	18	28	33	18	18	11	27	33	60
P7	46	19	104	58	29	55	8	44	28	244
P8	71	42	160	43	24	82	5	85	-	74
P9	38	17	81	58	25	20	9	63	25	132
P10	64	70	78	98	25	114	-	80	43	182
Reminders										
P11	71	35	31	59	29	20	17	81	60	120
P12	38	24	76	46	19	86	10	50	55	114
P13	36	27	194	58	48	22	43	85	59	62
P14	26	36	67	47	34	51	4	40	23	34
P15	23	30	36	25	-	28	-	40	19	35
RemindMe										
P16	51	78	40	50	-	88	-	273	26	144
P17	112	44	57	26	32	33	6	48	56	47
P18	92	144	11	146	142	15	-	238	142	80
P19	7	22	18	31	-	23	-	295	68	58
P20	67	105	60	42	25	65	4	50	29	102
Quack (chatbot baru)										
P21	57	35	30	46	76	57	6	78	28	162
P22	43	28	27	36	76	35	4	70	64	120
P23	53	19	8	50	21	15	4	49	32	54
P24	32	17	10	37	26	19	5	76	19	42
P25	27	15	33	19	19	16	3	29	20	89