

# Lembar Kerja

Sistem Interaksi Genap 2023/2024

**Dosen Pengajar:** Prof. Harry Budi Santoso, Ph.D., Dr. Baginda Anggun Nan Cenka, Dr. Eng. Lia Sadita, Suci Fadhillah, M.A, Syifa Nurhayati, M.Kom

**Nama Mhs 1 (NPM):** Novita Mulia Sari (2206032785)

**Nama Mhs 2 (NPM):** Alden Luthfi (2206028932)

## Lembar Kerja 11 – Evaluasi Sistem

PETUNJUK: Pada paper yang berjudul “Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?”, dipaparkan sebuah penelitian *usability* pada aplikasi WeightMentor untuk pemeliharaan penurunan berat badan di Inggris. Jawablah beberapa pertanyaan berikut ini terkait penelitian tersebut:

1. Mengapa evaluasi pada penelitian tersebut dilakukan?

Evaluasi dilakukan untuk menilai apakah metode-metode konvensional dalam usability testing dapat efektif digunakan untuk mengevaluasi antarmuka pengguna berbasis percakapan seperti chatbot. Penelitian ini menganggap penting karena chatbot memiliki interaksi yang berbeda dibandingkan dengan antarmuka pengguna tradisional.

2. Aspek apa saja yang dievaluasi pada penelitian tersebut?

Aspek yang dievaluasi meliputi kegunaan chatbot WeightMentor, korelasi antara berbagai kuesioner kegunaan konvensional, jumlah pengguna optimal yang diperlukan untuk mengidentifikasi masalah kegunaan, dan jumlah repetisi tugas yang diperlukan untuk mencapai kinerja tugas optimal.

3. Metode apa yang digunakan pada evaluasi? Jelaskan mekanisme evaluasi yang dilakukan!

Metode yang digunakan adalah pengujian kegunaan dengan desain penelitian observasional. Evaluasi melibatkan penggunaan kuesioner kegunaan khusus yang dirancang untuk chatbot, yang dibandingkan

# Lembar Kerja

## Sistem Interaksi Genap 2023/2024

**Dosen Pengajar:** Prof. Harry Budi Santoso, Ph.D., Dr. Baginda Anggun Nan Cenka, Dr. Eng. Lia Sadita, Suci Fadhillah, M.A, Syifa Nurhayati, M.Kom

**Nama Mhs 1 (NPM):** Novita Mulia Sari (2206032785)

**Nama Mhs 2 (NPM):** Alden Luthfi (2206028932)

dengan kuesioner kegunaan yang sudah ada. Pengujian melibatkan tugas-tugas yang harus diselesaikan oleh peserta, diikuti dengan pengisian kuesioner pre-test dan post-test untuk mengukur kesulitan dan keefektifan dalam menyelesaikan tugas tersebut.

4. Siapa saja responden yang dilibatkan dalam evaluasi? Berapa orang? Apa dasar alasan pengambilan responden?

Sebanyak 30 partisipan (dewasa sehat) direkrut untuk evaluasi. Responden dipilih untuk mendapatkan persepsi yang beragam dari pengguna umum, yang tidak harus memiliki latar belakang teknis atau medis, mewakili pengguna chatbot dalam konteks umum.

5. Bagaimana hasil evaluasinya?

Hasil evaluasi menunjukkan bahwa skor kegunaan WeightMentor umumnya tinggi, dan korelasi antar kuesioner kegunaan kuat. Jumlah pengguna optimal untuk mengidentifikasi masalah kegunaan adalah 26, yang menantang penelitian sebelumnya yang menyarankan lebih sedikit pengguna.

6. Apakah menurut kamu hasil evaluasi tersebut *reliable*? Mengapa?

Hasil evaluasi dapat dianggap *reliable* karena melibatkan metode yang komprehensif dan jumlah partisipan yang cukup untuk menggambarkan variasi umum dalam penggunaan chatbot. Penelitian juga menggunakan pendekatan statistik untuk memvalidasi temuan.

7. Berdasarkan *scope* evaluasi yang dilakukan pada penelitian tersebut, apakah model rancangan evaluasi yang sama dapat digunakan juga untuk mengevaluasi ChatGPT atau Google Bard?

Model rancangan evaluasi yang sama dapat digunakan untuk mengevaluasi ChatGPT atau Google Bard dengan beberapa penyesuaian pada kuesioner untuk memasukkan aspek-aspek khusus dari masing-masing teknologi tersebut. Namun, prinsip dasar pengujian kegunaan dan analisis respons pengguna dapat diaplikasikan secara luas.

# Lembar Kerja

Sistem Interaksi Genap 2023/2024

**Dosen Pengajar:** Prof. Harry Budi Santoso, Ph.D., Dr. Baginda Anggun Nan Cenka, Dr. Eng. Lia Sadita, Suci Fadhillah, M.A, Syifa Nurhayati, M.Kom

**Nama Mhs 1 (NPM):** Novita Mulia Sari (2206032785)

**Nama Mhs 2 (NPM):** Alden Luthfi (2206028932)

8. Apa hasil yang didapatkan pada *section Discussion* pada penelitian tersebut?

Pada bagian Discussion, penelitian tersebut mengemukakan bahwa metode tradisional mungkin tidak sepenuhnya memadai untuk menguji kegunaan chatbot karena chatbot memiliki karakteristik interaksi yang berbeda. Hasil menunjukkan bahwa evaluasi kegunaan chatbot mungkin memerlukan lebih banyak responden daripada sistem konvensional untuk mengidentifikasi semua masalah kegunaan. Selain itu, ditemukan bahwa pengguna dapat mencapai kecakapan dalam mengoperasikan chatbot dengan cepat, yang menunjukkan tingkat kemudahan penggunaan yang tinggi.