University of Brasília

Institute of Exact Sciences
Department of Computer Science

UXAPP: Evaluation of the User Experience of Digital Products through Emotion Recognition

Ricardo Cordeiro Galvão Sant'Ana van Erven

Dissertation submitted in partial fullfilment of the requirements for the Professional Master's Degree in Applied Computing

Advisor
Prof. Dr. Edna Dias Canedo

Brasília
2024

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

VV217u

VAN ERVEN, RICARDO CORDEIRO GALVÃO SANT'ANA
UXAPP: Evaluation of the User Experience of Digital Products through Emotion Recognition / RICARDO CORDEIRO GALVÃO SANT'ANA VAN ERVEN; orientador Edna Dias Canedo. -- Brasília, 2024.
157 p.

Dissertação(Mestrado Profissional em Computação Aplicada) -- Universidade de Brasília, 2024.

1. Emotion Recognition. 2. User Experience Evaluation. 3. Artificial Intelligence. I. Canedo, Edna Dias, orient. II. Título.

University of Brasília
Institute of Exact Sciences
Department of Computer Science

UXAPP: Evaluation of the User Experience of Digital Products through Emotion Recognition

Ricardo Cordeiro Galvão Sant'Ana van Erven

Dissertation submitted in partial fullfilment of the requirements for the Professional Master's Degree in Applied Computing

Prof. Dr. Edna Dias Canedo (Advisor)
University of Brasília (UnB)

Prof. Dr. Rejane Maria da Costa Figueiredo
University of Brasília (UnB)

Prof. Dr. Ana Paula Bernardi da Silva
Catholic University of Brasília (UCB)

Prof. Dr. Gladston Luiz da Silva
Coordinator of the Postgraduate Program in Applied Computing

Brasília, February 28, 2024

Dedication

To my wife, Andreza, and my children, Sabrina and Mateus.

May the light of science always guide our steps.

Acknowledgements

I thank everyone who directly or indirectly brought me here.

Firstly, to my wife and children, Andreza, Sabrina, and Mateus, who supported me unconditionally in this search for knowledge.

To my professors, who shared their precious time and knowledge, especially to my advisor, Professor Edna Dias Canedo, who guided me on my master's journey but always gave me the autonomy to decide the next step.

To the participants in the usability tests, who voluntarily made their time available to contribute to this work.

To my master's colleagues, who lived this experience with me, and to my work colleagues, who supported me for almost two years.

I want to give special thanks to my father, João, who always encouraged me in science, and to my mother, Terezinha, a doctor in cognitive psychology from the University of Brasília, for her encouragement and example.

Resumo

Contexto: Medir a experiência do usuário (UX) é essencial para criar valor na transformação digital. A medição permite identificar intenções de compra futuras, fidelidade e retenção do usuário. A forma tradicional de medir a experiência do usuário por meio da autoavaliação apresenta problemas. Portanto, precisamos de uma abordagem mais direta que nos permita medir a experiência do usuário automaticamente. **Objetivo:** Implementar e validar o modelo de avaliação de UX através de uma ferramenta que calcule automaticamente a avaliação da experiência do usuário de um produto digital e apresente pontos positivos, neutros e negativos na utilização deste produto. O modelo incluiu um processo de trabalho, um experimento exploratório e um aplicativo, que chamamos de UXAPP. **Métodos:** Identificamos e selecionamos o estado da arte relacionado ao reconhecimento de emoções com Inteligência Artificial no contexto da satisfação do usuário. Depois disso, propusemos e implementamos o modelo de avaliação de UX. Desenvolvemos o aplicativo UXAPP e conduzimos o experimento exploratório seguindo o processo de trabalho. Convidamos nove participantes para realizar quatro tarefas cada. Coletamos dados de entrada manual e capturamos emoções do vídeo e da fala do usuário. Em seguida, realizamos uma análise de ambos os dados e geramos um relatório de UX. Por fim, comparamos os resultados obtidos do UXAPP e os dados de entrada do usuário. **Resultados:** A avaliação de UX do UXAPP corresponde diretamente às entradas do usuário em 50% das tarefas e dá resposta próxima em outros 47,22% dos resultados. Cada elemento UX, como usabilidade, afeto e valor para o usuário, foi avaliado e analisado de forma independente. A análise de usabilidade do UXAPP tem correspondência direta com a resposta do usuário em 30,56% das tarefas e dá resposta próxima em outros 33,33% dos resultados. A análise de afeto do UXAPP tem correspondência direta com a resposta do usuário em 52,78% das tarefas e dá resposta próxima em outros 41,67% dos resultados. O UXAPP também identificou 616 pontos de sentimentos positivos ou negativos em vídeo e fala para todas as tarefas. A análise de valor do usuário com UXAPP tem correspondência direta com a resposta do usuário em 36,11% das tarefas e dá resposta aproximada em 50,00% dos resultados. Registramos 5h 35m 31s de duração das tarefas. Com base em nossa experiência, estimamos que uma pessoa precisaria de pelo menos quatro vezes a duração dessa tarefa para processar todas as informações apresentadas no relatório de UX, incluindo encontrar todos os pontos de sentimento e distinguir picos de sentimento positivos e negativos. **Conclusões:** As possibilidades de aplicação do reconhecimento de emoções são inúmeras em termos de contextos, técnicas, formas e componentes. Apesar disso, foi possível desenvolver o UXAPP e validar o modelo de avaliação de UX para medir a experiência do usuário de um produto digital através do reconhecimento de emoções e Inteligência Artificial. Descobrimos que a análise UXAPP e a análise de avaliação do usuário são importantes e complementares. A principal contribuição deste trabalho é a entrega de um modelo validado para medir automaticamente a experiência do usuário e identificar pontos positivos, neutros e negativos na utilização de um produto digital. O UXAPP pode ser utilizado diretamente pelo usuário final, possibilitando automatizar todos os esforços para obter uma pontuação de avaliação de UX e identificar os pontos positivos e negativos da interação do usuário. Com isso, ele pode reduzir drasticamente os custos envolvidos no desenvolvimento de uma competência de experiência do usuário em uma organização. Esta pesquisa também define um processo de experimentação de medição que qualquer organização ou pesquisador pode reproduzir.

Palavras-chave: Reconhecimento de Emoções, Avaliação de Experiência do Usuário, Inteligência Artificial

Resumo Expandido

UXAPP: Avaliação da Experiência do Usuário de Produtos Digitais por meio do Reconhecimento de Emoções

Introdução

A transformação digital pode ser definida como uma mudança na forma como uma empresa emprega tecnologias digitais para desenvolver um novo modelo de negócios digital que ajude a criar e apropriar mais valor para a empresa [1].

Os principais desafios da transformação digital são o compromisso organizacional, a criação de valor, a proposta de valor, a entrega de valor, a captura de valor, a infraestrutura de TI e a segurança de dados. A criação de valor está relacionada ao modelo de negócio e seus elementos, como processos, recursos, capacidades e parcerias [2].

Entendemos que a criação de valor e seus elementos correlatos podem ser visto como um dos principais elementos no desempenho organizacional. A partir disso, observamos que a principal métrica para medir o desempenho de uma forma não financeira é medir a satisfação do usuário [3].

A satisfação do usuário é um atributo dos elementos da experiência do usuário (UX). A experiência do usuário pode ser vista como uma combinação de usabilidade, afeto e valor do usuário [4] [5]. Para medir a experiência UX, precisamos avaliar todos os elementos.

Considerando a importância da utilização de avaliações contínuas de desempenho de um produto digital em seu desenvolvimento, surge a questão que esta pesquisa busca responder: **É possível medir automaticamente a experiência do usuário de produtos digitais, identificando aspectos positivos, neutros e pontos negativos da experiência desses usuários?**

Este trabalho implementa o modelo de avaliação de UX através de uma ferramenta para medir a experiência do usuário com produtos digitais de forma repetível e escalável. Isso permite que as organizações obtenham informações valiosas para melhorar os produtos e serviços oferecidos e, ao fazê-lo, reduzir desperdícios e entregar mais valor à sociedade.

Metodologia

Este trabalho é realizado em quatro etapas. A primeira consiste em identificar estudos sobre experiência do usuário que realizem avaliações de satisfação do usuário com foco no reconhecimento de emoções por meio de Inteligência Artificial, incluindo técnicas, benefícios e desafios. A segunda etapa é a seleção das técnicas mais adequadas, de acordo com o mapeamento anterior, com especial atenção aos desafios e benefícios observados.

A terceira etapa refere-se à implementação de uma ferramenta para validar a hipótese de pesquisa. O modelo proposto considera as descobertas feitas no mapeamento sistemático para utilização de técnicas recomendadas como boas práticas com resultados relevantes. O quarto passo é validar a ferramenta em cenário real e fazer os ajustes necessários. A metodologia utilizada neste trabalho é apresentada na Figura 1.1

Resultados e Discussão

Com base nas informações identificadas no mapeamento sistemático da literatura, foi proposto um modelo de avaliação de UX para realizar o reconhecimento de emoções utilizando Inteligência Artificial. Decidiu-se integrar as soluções *Google* existentes para o usuário final e na nuvem, que desempenhavam as responsabilidades dos componentes frontend, backend e IA, com a execução de um motor UXAPP para analisar os resultados da IA.

O aplicativo recebe a câmera e a tela do usuário, bem como os dados de entrada no UXAPP. O aplicativo entrega um vídeo com avaliações instantâneas de satisfação e afeto ao longo do vídeo, além de um relatório de UX com a pontuação da avaliação de UX, todos os pontos positivos e negativos com detecção de sentimento e informações adicionais para ajudar o usuário a entender a experiência.

Foi realizado um experimento em ambiente controlado. Esta abordagem é baseada em um teste de usabilidade com nove participantes e quatro tarefas por participante. O teste é orientado para falar sobre ações e sentimentos. Uma avaliação manual da experiência do usuário foi realizada ao final do experimento. Após o teste, o UXAPP analisou os dados da tarefa para elaborar o relatório de UX. Para uma duração média de 9m 19s, o UXAPP analisou a tarefa para um tempo médio de 24m 37s.

A avaliação do UXAPP nos deu um resultado consistente e equilibrado em relação ao resultado dos dados de entrada do usuário, mas não pode ser confundido com a avaliação realizada pelo usuário.

A avaliação UX do UXAPP corresponde diretamente à avaliação do usuário em 50% das tarefas e dá resposta próxima em outras 47,22% dos resultados. Ao comparar a análise dos dados de entrada do usuário e a análise do UXAPP, verificamos que o usuário demonstrou menos emoção do que sentia. Afirmamos que a avaliação UX a partir dos dados de entrada do usuário difere da avaliação UXAPP. Enquanto o primeiro se conecta com a complexidade dos pensamentos e sentimentos mais profundos de um indivíduo, o segundo mede a experiência perceptível do usuário. Numa analogia com um iceberg, a primeira é a parte submersa do iceberg e a segunda é a parte visível. Ambos são importantes e complementares.

Nosso trabalho identificou vários problemas na avaliação da experiência do usuário. Primeiramente, identificamos barreiras externas que impactaram a experiência do usuário, como má conexão com a internet, que impediu o carregamento completo do site, falta de conhecimento sobre o idioma do site testado e linguagem técnica para descrever as funcionalidades do site testado. Em segundo lugar, observamos que a experiência perceptível não é necessariamente aquela que o usuário registrou na tarefa. Alguns usuários demonstraram emoções com valência e registraram um estado emocional com valência oposta. Terceiro, mesmo depois de alguns usuários terem tido uma experiência ruim, observamos que eles não mudaram seu estado emocional inicial, o que significa que o estado emocional do usuário difere da experiência perceptível do usuário ao usar um produto digital.

A análise de usabilidade do UXAPP corresponde à avaliação do usuário em 30,56% das tarefas e dá resposta próxima em outras 33,33% dos resultados. Observamos que a abordagem SUS consome muito tempo registrando a avaliação do usuário. O processo de análise de usabilidade do UXAPP é muito complexo devido à necessidade de identificar ocorrências de sentenças do SUS com similaridade e mesma valência da fala do usuário. Uma abordagem melhor poderia ser o uso de grandes modelos de linguagem (LLMs) ou modelos multimodais geradores de IA para determinar a usabilidade da interpretação direta do contexto do usuário.

A análise de afeto do UXAPP corresponde à avaliação do usuário em 52,78% das tarefas e dá resposta próxima em outras 41,67% dos resultados. O UXAPP também identificou 616 pontos de sentimentos positivos ou negativos em vídeo e fala para todas as tarefas. Observamos que o afeto identificado correspondeu ao resultado da tarefa para tarefas com diferenças significativas entre afeto positivo e negativo. Isso significa que o afeto identificado impulsiona corretamente a experiência perceptível do usuário. Afirmamos que o estado emocional do usuário está relacionado à experiência do usuário de um produto digital imediatamente antes, durante e imediatamente após a utilização do produto. Ainda assim, o estado emocional é um conceito mais profundo e complexo do que a experiência perceptível do utilizador de um produto digital.

A análise de valor do usuário UXAPP corresponde à avaliação do usuário em 36,11% das tarefas e dá resposta aproximada em 50,00% dos resultados. Este resultado se comporta de forma semelhante à avaliação de afeto. Observamos uma incompatibilidade entre a experiência perceptível do usuário dos produtos digitais e a emoção interna do usuário representada pela satisfação do usuário. Afirmamos que existe uma relação entre a satisfação do usuário e a experiência perceptível do usuário em produtos digitais. No entanto, o conceito de satisfação do usuário é mais profundo e complexo do que a experiência perceptível do usuário em produtos digitais.

Registramos 5h 35m 31s de duração das tarefas. Com base em nossa experiência, estimamos que uma pessoa precisava de pelo menos quatro vezes a duração dessa tarefa para processar todas as informações apresentadas no relatório de UX, incluindo encontrar todos os pontos de sentimento e distinguir picos de sentimento positivos e negativos.

Conclusões

O processo de avaliação de UX e os experimentos funcionaram para extrair os dados necessários para o experimento. Os resultados do UXAPP demonstram que o modelo de avaliação UX pode impulsionar a experiência do usuário.

Este trabalho traz o benefício essencial de mapear o estado da arte relacionado a estudos que medem a satisfação do usuário por meio do reconhecimento de emoções por meio de expressões faciais. A disponibilização desse mapeamento permite que novos estudos sejam propostos para identificar carências de tecnologias, algoritmos e estudos sobre a necessidade de aplicação desses recursos em novas áreas ou áreas ainda não atendidas por eles.

Além disso, este trabalho entrega um modelo testado para avaliar a experiência do usuário a partir do reconhecimento emocional. O modelo UXAPP nos permite dimensionar a avaliação da experiência do usuário de produtos digitais a um novo nível e obter feedback contínuo para melhorar o desenvolvimento de novos produtos e serviços digitais.

Esperamos que os procedimentos de medição da experiência do usuário sejam facilitados com o uso de ferramentas de reconhecimento de emoções, e isso possibilite identificar os principais pontos de melhoria e atrito do produto nos serviços digitais para que esses produtos e serviços possam evoluir para entregar mais valor aos prestadores de serviços e sociedade.

Palavras-chave: Reconhecimento de Emoções, Avaliação de Experiência do Usuário, Inteligência Artificial

Abstract

Context: Measuring user experience (UX) is essential to create value in digital transformation. Measurement allows you to identify future purchase intentions, user loyalty, and retention. The traditional way of measuring user experience using self-assessment has problems. Therefore, we need a more straightforward approach to allow us to measure user experience automatically. **Objective:** Implement and validate the UX evaluation model through a tool that automatically calculates the user experience rating of a digital product and presents positive, neutral, and negative points in using this product. The model included a work process, an exploratory experiment, and an application, which we call UXAPP. **Methods:** We identified and selected the state-of-art related to emotion recognition with Artificial Intelligence in the context of user satisfaction. After that, we proposed and implemented the UX evaluation model. We developed the UXAPP application and conducted the exploratory experiment following the work process. We invite nine participants to perform four tasks each. We collect manual input data and capture emotions from the user's video and speech. Then, we analyze both data and generate a UX report. Finally, we compare the results obtained from UXAPP and the user input data. **Results:** The UXAPP UX evaluation directly matches the user evaluation in 50% of the tasks and gives a close answer in others 47.22% of the results. Each UX element, like usability, affect, and user value, was evaluated and analyzed independently. The UXAPP usability analysis matches the user evaluation in 30.56% of the tasks and gives a close answer in others 33.33% of the results. The UXAPP affect analysis matches the user evaluation in 52.78% of the tasks and gives a close answer in others 41.67% of the results. UXAPP also identified 616 points of positive or negative sentiments in video and speech for all tasks. The UXAPP user value analysis matches the user evaluation in 36.11% of the tasks and gives a close answer in 50.00% of the results. We registered 5h 35m 31s of tasks' duration. Based on our experience, we estimate that a person needed at least four times this task duration to process all the information presented in the UX report, including finding all sentiment points and distinguishing positive and negative sentiment peaks. **Conclusions:** The possibilities of applying emotion recognition are countless in terms of contexts, techniques, forms, and components. Despite this, it was possible to develop UXAPP and validate the UX evaluation model to measure the user experience of a digital product through emotion recognition and Artificial Intelligence. We found that UXAPP analysis and user evaluation analysis are important and complementary. The main contribution of this work is the delivery of a validated model to automatically measure user experience and identify positive, neutral, and negative points in the use of a digital product. UXAPP can be used directly by the final user, making it possible to automate all the efforts to obtain a UX evaluation score and identify user interaction's positive and negative points. It can drastically reduce the costs involved in developing a user experience competence in an organization. This research also defines a measurement experimentation process that any organization or researcher can reproduce.

Keywords: Emotion Recognition, User Experience Evaluation, Artificial Intelligence

Contents

1 Introduction **1**
1.1 Research Problem 2
1.2 Research Hypothesis 3
1.3 Justification 3
1.4 Objectives 5
1.4.1 General Objective 5
1.4.2 Specific Objectives 5
1.5 Expected Results 5
1.6 Research Methodology 6
1.7 Publications 6
1.8 Work Structure 8

2 Literature Review **9**
2.1 User Experience (UX) 9
2.1.1 User Experience Evaluation Tools 11
2.2 User Satisfaction Measurement 13
2.3 Emotion Recognition 14
2.4 Chapter Summary 17

3 Theoretical Framework **19**
3.1 User Experience Elements 19
3.1.1 System Usability Scale (SUS) 22
3.2 Related Work 23
3.3 Chapter Summary 29

4 Systematic Literature Mapping **30**
4.1 Planning 31
4.1.1 Research Protocol 31
4.1.2 Quality Assessment Criteria 36

4.1.3 Data Extraction Form . . . 37
4.2 Conducting . . . 37
4.2.1 Conducting Research in the Bases . . . 37
4.2.2 Selection of Studies . . . 38
4.2.3 Application of Quality Criteria . . . 38
4.2.4 Data Extraction . . . 40
4.3 Results . . . 40
4.3.1 Data Analysis . . . 40
4.3.2 Response to Research Questions . . . 41
4.3.3 Threats to Results . . . 49
4.4 Chapter Summary . . . 49

5 Proposed Model **51**
5.1 Proposal Overview . . . 51
5.1.1 Emotional Usability Evaluation . . . 52
5.1.2 Affect Evaluation . . . 58
5.1.3 Emotional Satisfaction Evaluation . . . 58
5.1.4 User Experience Evaluation . . . 62
5.1.5 Proposed Model Activities . . . 63
5.1.6 Exploratory Experiment . . . 66
5.2 Application . . . 66
5.2.1 Architecture . . . 66
5.2.2 Overview . . . 69
5.3 Chapter Summary . . . 74

6 Discussion **75**
6.1 Implementation Overview . . . 75
6.1.1 Model Activities . . . 75
6.1.2 Exploratory Experiment . . . 76
6.1.3 Application . . . 83
6.2 Results . . . 84
6.2.1 First Round of Analysis . . . 96
6.2.2 Second Round of Analysis . . . 106
6.3 Chapter Summary . . . 123

7 Conclusions **124**
7.1 Future Work . . . 127

References **129**

List of Figures

1.1 Methodology used in this dissertation. 6

2.1 A UXmood tool's view [6]. 13

3.1 User experience model [4]. 20
3.2 Framework of product experience [7]. 21
3.3 Mapping emotional states by Kolakowska et al. for Ekman according to Landowska [8]. 26

4.1 Systematic literature mapping used in this dissertation. 31
4.2 Systematic literature mapping workflow. 38
4.3 Paper rejection statistics by exclusion criteria. 40
4.4 Papers selected by type of study. 41

5.1 Proposed Model 52
5.2 Working process of the proposed model. 65
5.3 Exploratory experiment illustration. 67
5.4 Proposed model architecture. 69

6.1 View of the home screen of the tested site uxapp.com.br 77
6.2 The error message of obstacle O2. 80
6.3 Login. 83
6.4 About. 83
6.5 Task. 83
6.6 Task 1 recording before UXAPP analysis. 85
6.7 Task 1 recording after UXAPP analysis. 86
6.8 User image from task 1 recording in Vision API. 86
6.9 Confidence score of user image from task 1 recording in Vision API. . . . 87
6.10 Main. 88
6.11 Screen 1. 88
6.12 Screen 2. 88

6.13 Screen 3. 88
6.14 Screen 4. 88
6.15 Screen 5. 88
6.16 Screen 6. 89
6.17 Screen 7. 89
6.18 Screen 8. 89
6.19 Screen 9. 89
6.20 Screen 10. 89
6.21 Screen 11. 89
6.22 Screen 12. 90
6.23 Comparative task success rate. 98
6.24 Distribution of usability evaluation match for the first analysis. 98
6.25 Distribution of affect evaluation match for the first analysis. 100
6.26 Distribution of sentiment changes according to task results. 100
6.27 Tasks in the state "stable with task success" per initial emotional state. . 100
6.28 Tasks in the state "change without task success" per final emotional state. 102
6.29 Positive and negative sentiments identified in video and speech for the first analysis. 102
6.30 Distribution of user value evaluation match for the first analysis. 104
6.31 UXAPP shows joy. 104
6.32 Resized and cropped user image of the joy figure. 105
6.33 UXAPP shows neutrality. 105
6.34 Distribution of user value evaluation match for the first analysis. 106
6.35 Distribution of UX evaluation with user input data for the first analysis. 106
6.36 Distribution of usability evaluation match of the second analysis. 109
6.37 Distribution of affect evaluation match of second analysis. 111
6.38 Distribution of positive and negative affect identification in the second analysis. 111
6.39 Positive and negative sentiments identified in video and speech of second analysis. 113
6.40 Emotional final state vs Affect identified for tasks with great differences between positive and negative affect. 113
6.41 Task result vs Affect identified for tasks with great differences between positive and negative affect. 114
6.42 Distribution of final emotional state in second analysis for unsuccess tasks. 114
6.43 Distribution of affect identified in second analysis for unsuccess tasks. . . 115
6.44 Distribution of the number of users that repeat an initial emotional state. 115

6.45 Distribution of user value evaluation match for the second analysis. . . . 117
6.46 Distribution of overall user value in second analysis for unsuccess tasks. . 117
6.47 Distribution of user value identified in second analysis for unsuccess tasks. 117
6.48 Distribution of user value evaluation match for the second analysis. . . . 118
6.49 Distribution of UX evaluation with user input data for the second analysis. 118
6.50 Distribution of UX evaluation match for the second analysis. 122
6.51 Distribution of user satisfaction score with UXAPP. 122

List of Tables

1.1 Definition of user satisfaction and its components [9]. 2

3.1 Summary of the use of emotion recognition techniques in usability testing according to Landowska [8]. 27
3.2 Summary of the main characteristics of related works and this study [10] [11] [12] [8]. 29

4.1 Presentation of the PICOC approach criteria applied to systematic literature mapping. 32
4.2 Systematic Literature Mapping Research Questions. 33
4.3 Presentation of relevant keywords and synonyms in systematic literature mapping. 33
4.4 Digital Bases Used in Systematic Mapping of Literature. 34
4.5 Presentation of the inclusion and exclusion criteria used in systematic literature mapping. 35
4.6 Presentation of questions for quality assessment of selected papers. 36
4.7 Presentation of levels of compliance with the criteria for quality assessment of selected papers. 37
4.8 Studies selected in systematic literature mapping to answer research questions. 39
4.9 Summary of systematic literature mapping papers. 46

5.1 Portuguese sentences used to obtain baseline for SUS similarity score. . . 54
5.2 English sentences used to obtain a baseline for SUS similarity score . . . 55
5.3 Initial similarity score and sentiment analysis' valence conversion to Likert scale. 56
5.4 Final similarity score and sentiment analysis' valence conversion to Likert scale. 57
5.5 Initial emotional usability scale. 57
5.6 Final emotional usability scale. 57

5.7 Final affect evaluation. 58
5.8 Valence of emotions used. 59
5.9 Initial association of intensities with intensity score. 60
5.10 Final association of intensities with intensity score. 60
5.11 Association of confidence degree with confidence score. 60
5.12 Initial manual review and final level conversion to satisfaction score. . . . 61
5.13 Final manual review and final level conversion to satisfaction score. . . . 61
5.14 First UX evaluation points. 62
5.15 Final UX evaluation points. 62
5.16 Initial UX evaluation scale. 62
5.17 Final UX evaluation scale. 63
5.18 Constraints for product usability testing. 66
5.19 List of technologies used in the proposed model. 68
5.20 Sentiment Analysis Scores 73

6.1 Attributes of the experiment carried out. 77
6.2 Summary of macrofunctionalities of the current MVP of the UXAPP site. 78
6.3 Style Guide definitions and its motivations. 79
6.4 Obstacles were intentionally added to the site. 80
6.5 Usability Test Tasks. 81
6.6 Experiment task IDs. 82
6.7 UXAPP's Functionalities 84
6.8 Summary of the task success and duration. 97
6.9 First analysis of usability evaluation data. 99
6.10 Affect evaluation data for the first analysis. 101
6.11 User value evaluation data for the first analysis. 103
6.12 User experience evaluation data for the first analysis. 107
6.13 User input data used to calculate UX evaluation for the first analysis. . . 108
6.14 Usability evaluation data of second analysis. 110
6.15 Affect evaluation data of second analysis. 112
6.16 User value evaluation data for the second analysis. 116
6.17 User experience evaluation data for the second analysis. 119
6.18 User input data used to calculate UX evaluation for the second analysis. 120
6.19 Comparative results of user experience data for the second analysis. . . . 121
6.20 Comments of user satisfaction score with UXAPP. 123

Chapter 1

Introduction

Digital transformation can be defined as a change in how a company employs digital technologies to develop a new digital business model that helps create and appropriate more value for the company [1].

Digital transformation has emerged as an important and, at the same time, complex issue for society. Digital transformation requires action in terms of strategy, culture, process, organizational structure, and information technology (IT). Access to digital technologies has changed consumer behavior and increased consumer expectations about the service provided to carry out transactions [13] digitally.

Digital technologies allow new value propositions related to service provision to emerge. New environments can be created that encourage the sale of products and services. On the other hand, it is possible to extract usage data to boost these same services [1].

The main challenges of digital transformation are organizational commitment, value creation, value proposition, value delivery, value capture, IT infrastructure, and data security. Value creation is related to the business model and its elements, such as processes, resources, capabilities, and partnerships [2].

Product development in the context of digital transformation is associated with the implementation of new IT solutions. Digitizing product development requires a data-driven approach and considers physical components, services, and software in addition to traditional [2] requirements.

Organizational performance measurement has moved from the supply side to the demand side, so new dimensions such as quality, time, and user satisfaction have been observed. Regardless of whether organizations provide products or services, the value chain must be viewed from the user's perspective and how they use these products and services in their lives [14]. The main metric for measuring performance in a non-financial way is measuring user satisfaction [3].

User satisfaction can be described as the relationship between the service provider's performance regarding the brand and the user's expectation of that service. While supplier performance is related to brand value, user expectations are related to their emotion [9]. It is possible to describe user satisfaction as shown in Table 1.1.

Concept	Definition
Customer Satisfaction	Brand Trust / Customer Emotion
Trust in Brand	f(Spread Trust x Trust in Character x Trust in Affection)
Customer Emotion	f(Emotion Evocation x Emotion Formation x Behavior Response)

Table 1.1: Definition of user satisfaction and its components [9].

To increase user satisfaction, we need to act on the trust the consumer acquires about the brand or reinforce positive emotions in the user's interaction. The evocation of emotion is associated with the user's sentimental understanding of the products provided. This understanding is formed after evaluating the product image and description, as well as the supplier's reputation and evaluation [9].

1.1 Research Problem

Five factors influence user satisfaction with digital services: security, privacy, trust, accessibility, awareness of services, and quality of services. There are some challenges in measuring user satisfaction, including [15]:

- user satisfaction is subjective and influenced by factors such as personal prejudices, previous experiences, and cultural differences;
- there may be external barriers, such as technical difficulties, that prevent users from using the services, even if the level of satisfaction is satisfactory;
- user needs and expectations can change, challenging measuring satisfaction over time.

User satisfaction is an attribute of the user experience (UX) elements. User experience can be seen as a composite of usability, affect, and user value [4] [5]. To measure UX experience, we need to evaluate all elements. Considering the importance of using continuous assessments of the performance of a digital product in its development, the question that this research seeks to answer arises: **Is it possible to automatically measure the user experience of digital products by identifying positive, neutral, and negative points from the experience of these users?**

This research aims to resolve the challenges presented:

- About the subjectivity of evaluation, this research contributes by bringing an approach to measuring user experience that does not depend on manual responses but rather on information obtained naturally from the user;
- About the technical difficulties that users may present when using services, measuring user experience from the use of digital products can provide clues to points where the product creates resistance to users and prevents them from proceeding or limits the use of some functionality;
- Regarding changes over time in users' needs and expectations, this research provides an approach simple and automated enough to be replicated countless times for numerous digital products and services, making it possible to track user experience without requiring dedicated teams or overloading users with research.

1.2 Research Hypothesis

Automatic emotion recognition, in the context of facial expression recognition, is related to the task of classifying emotion archetypes, which would be, in this context, models of facial expressions that represent that emotion [16].

Emotions are complex and multifaceted, depending on context, culture, and individual differences. Despite this, six basic emotions appear in literature in practically all lists: happiness, sadness, fear, anger, surprise, and disgust [16].

These basic emotions are considered universal and culture-independent [17]. The authors state that it is possible to measure user experience by recognizing emotions identified through the user's facial expressions and sentiment analysis of the user's speech. Therefore, this work has the following research hypotheses:

1. It is possible to use automatic emotion recognition to measure user experience in a repeatable and scalable way;
2. Using this technique, the positive, neutral, and negative points of the user experience using a digital product or service can be identified.

1.3 Justification

Financial metrics do not favor measuring the value of intangible assets, such as brand, loyalty, and user satisfaction. Most quality theories understand user satisfaction as a determining success factor in all industries and sectors of society [18].

It is also known that high user satisfaction leads to a stronger company image, increased user loyalty, and reduced complaints. To this end, understanding how user satisfaction evolves is crucial [18].

Discussions regarding user satisfaction have evolved. Some results identified were [18]: 1) Brand image remains vital in determining user satisfaction over time, from which it is observed that intangible assets, such as trust, have an impact on users' perception; 2) Product quality has less impact on user satisfaction, as new products, features or advanced interfaces have a volatile effect on user satisfaction; 3) Service quality has a more significant impact on user satisfaction so that aspects such as user relationships, availability and engagement have become preponderant; 4) Being close to users' core values deeply boosts user satisfaction.

Some organizations use the measurement of user satisfaction symbolically; that is, user satisfaction information is used only to give an appearance of legitimacy to the process to justify decisions already taken in the organization. These organizations have difficulty understanding measurements, and their processes do not lead to improvement actions [3].

Traditional satisfaction measurement has some problems [19]:

- the satisfaction assessment usually covers the entire user experience about using the product;
- the evaluation is carried out with date accuracy that does not capture the evolution of this user's satisfaction nor changes and justification for changes in the evaluation of this satisfaction;
- different users have different judgment standards, which may generate deviations in the overall evaluation result.

The use of more invasive methods to increase the accuracy of satisfaction assessment, such as blood pressure measurement, electroencephalogram, and heart rate measurement, among others, causes tension and anxiety in users, which contradicts the premise that tests need to be carried out in a consistent manner as close as possible to the actual use of the products [19].

Affective computing can avoid differences in users' evaluation standards, avoid anguish and anxiety in the environment, and allow user satisfaction to be quantified more assertively [19]. Affective computing can be defined as "computing that relates to, arises from, and deliberately influences emotion" [20].

An approach, therefore, that allows the continuous measurement of user experience in using a digital product makes it possible to obtain an indicator that can lead to a more assertive way to develop products and services.

1.4 Objectives

1.4.1 General Objective

This work aims to implement and validate a UX evaluation model through a tool that automatically calculates a user's experience rating when using a digital product and presents the positive, neutral, and negative points.

1.4.2 Specific Objectives

To achieve the general objective, the following specific objectives were defined:

1. identify studies that specify user experience elements and carry out user satisfaction assessment and how they do it;
2. select techniques for evaluating user experience through emotion recognition;
3. implement a tool to visualize the level of user experience throughout the use of the product and which presents a report at the end of use;
4. validate the tool by comparing the results with the manual application of user experience measurement, presenting discussions of the results, and implementing possible adjustments identified.

1.5 Expected Results

Investing in improving user satisfaction brings long-term economic returns. To achieve this, the organization must align its processes, resources, performance measurements, and organizational structures to treat users as assets [21].

Improved satisfaction provides greater user loyalty, reduced price elasticities, lower costs to attract users, reduced failure costs, and improved reputation for the company [21].

Measuring user experience and user satisfaction is an activity, as seen, essential for developing products and services so that it allows the creation of these products and services aligned with users' values.

This work implements the UX evaluation model through a tool to measure user experience with digital products in a repeatable and scalable way. This allows organizations to obtain valuable information to improve the products and services offered and, in doing so, reduce waste and deliver more value to society.

1.6 Research Methodology

This work is carried out in four stages. The first consists of identifying studies about user experience that carry out user satisfaction assessments focusing on emotion recognition using Artificial Intelligence, including techniques, benefits, and challenges. The second stage is the selection of the most appropriate techniques, according to the previous mapping, with particular attention to the challenges and benefits observed.

The third step refers to implementing a tool to validate the research hypothesis. The proposed model considers the discoveries made in systematic mapping to use recommended techniques as good practices with relevant results. The fourth step is validating the tool in a real scenario and making necessary adjustments. The methodology used in this work is presented in Figure 1.1

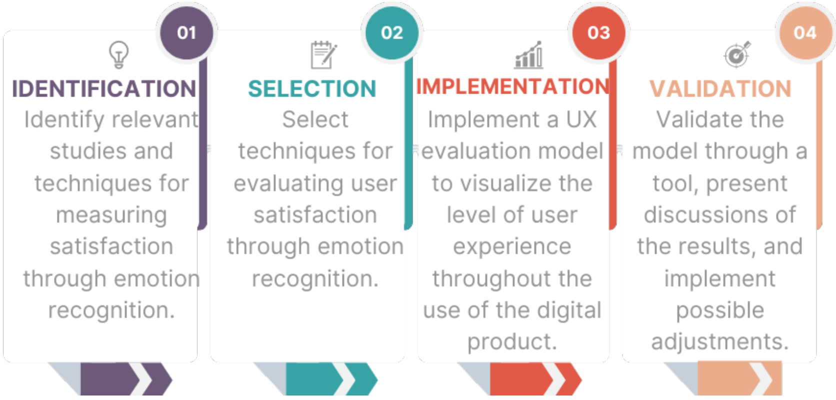

Figure 1.1: Methodology used in this dissertation.

1.7 Publications

We present below the articles published in conference proceedings during this master's degree.

- Ricardo Cordeiro Galvão Sant'Ana van Erven, Edna Dias Canedo: **Measurement of User's Satisfaction of Digital Products through Emotion Recognition [22].** SBQS 2023: 62-71. *Abstract: Context: Measuring user satisfaction is an essential tool to create value in digital transformation. Measurement allows you to identify future purchase intentions, user loyalty, and retention. The traditional way of measuring satisfaction using self-assessment has problems, such as subjectivity. Therefore, a more objective approach is needed that allows automatic measurement of satisfaction. Objective: This research aims to make a systematic literature mapping (SLM) to identify the state-of-the-art use of Artificial Intelligence tools and techniques related to emotion recognition visually and then use the results to propose a model that can lead to an automated user satisfaction measurement. We present the model with a work process, an exploratory experiment, and an application project to measure the user experience from visual emotion recognition, automatically calculate the satisfaction assessment of a user when using a digital product, and present the positive, neutral, and negative points of using this product. Methods: A systematic literature mapping was used to identify the primary studies related to techniques, benefits, and challenges that associate the measurement of user satisfaction through the recognition of emotions with the use of Artificial Intelligence (AI). In the proposed model, an experiment* *was planned in two phases to validate research hypotheses related to the objective of the work. The experiment, which will apply the projected application, will be carried out by videoconferencing with the approach of exploratory usability testing on a web product. Results: 10 primary studies were identified in different areas of knowledge: restaurants, television systems, retail stores, artistic shows, and usability testing in a call center system. Two systematic reviews of the literature were also identified. The technique most commonly used in primary studies is the convolutional neural network (CNN). The use of cloud services for emotion recognition was also verified. Benefits related to user feedback, such as user profile mapping, were reported, and challenges for emotion recognition were found, such as user privacy and capture environment inadequacies. Besides, a model to automatically measure user satisfaction in a scalable and replicable way was proposed. Conclusions: The possibilities of applying emotion recognition are countless in terms of contexts, techniques, forms, and components. Despite this, it was possible to identify good practices that could guide the creation of a tool to measure the satisfaction of using a digital product through emotion recognition and Artificial Intelligence. The main contribution of this work is the proposal of a model that guides the creation of an automatic satisfaction measurement tool to identify positive, neutral, and negative points in the use of a digital product. This research also defines a measurement experimentation process that any organization or researcher can reproduce.*

- Ricardo Cordeiro Galvão Sant'Ana van Erven, Demétrius de Almeida Jubé, Helen Reis Santos, Sérgio Antônio Andrade de Freitas, Edna Dias Canedo: **Gamification Project to Receive Continuous Feedback in the Context of the Evolution of Public Service for Lawyers [23].** FIE 2023: 1-8. *Abstract: Contribution: The main contribution presented by this work was proposing a method for building gamification projects that can be replicated in teaching applied computing, developing new software solutions, or creating gamification projects for other areas. The exposed methodology not only proposes a solution to engage people in favor of learning objectives effectively but also presents two distinct gamification approaches using the same framework that can be used separately or together depending on the complexity of the application scenario. The implementation of this methodology can contribute to the advancement of knowledge in the field of gamification and motivation, enabling future studies to analyze the effectiveness of the proposed techniques and improve engagement strategies. Background: There are various contexts in which spontaneous user action is necessary or desirable, such as exercising the right to vote in a democracy or attending an algebra class. In both cases, motivation can be crucial in how an individual engages in these activities. Self-determination theory suggests that both extrinsic and intrinsic motivation can promote such engagement. Gamification, through motivational elements known as game techniques, allows for creating an immersive environment for specific activities. The focus is on adapting the approach to achieve a goal that is aligned with the motivational and behavioral profile of the participants. Intended outcomes: This approach was employed in a real-life case to propose a gamification solution that motivates participating lawyers in trial sessions of a Brazilian higher court to provide spontaneous, contextualized, neutral, and frequent feedback. Application design: This work proposes the development of a gamification project by merging two assessments based on the Octalysis gamification framework: one to identify motivation techniques based on expert opinions in the field of study and another to identify motivation techniques by mapping the motivational factors of the participant persona using self-determination theory. The final gamification project is the proposal of Octalysis techniques for each core driver resulting from the merging of the assessments. Findings: The proposed gamification project resulted in 16 techniques distributed among 7 core drivers of the Octalysis framework. At the end of the project, the techniques to be used in the gamification design are presented. The proposed approach, based on merging assessments and applying the Octalysis framework, can be applied in different contexts and activities, always to increase engagement and motivation among the individuals involved.*

- Ricardo Cordeiro Galvão Sant'Ana van Erven, Pollyanna C. O. Dias, Demétrius de Almeida Jubé, George Marsicano Corrêa, Edna Dias Canedo: **Assessment of Knowledge in Requirements Engineering at Startup Gov.br [Avaliação de Conhecimento em Engenharia de Requisitos no Startup Gov.br] [24].** WER 2023. *Abstract: In 2021, the Brazilian Government started the Startup Gov.br Program, which aims to support and promote the acceleration of strategic projects that make up the digital transformation of the Brazilian government. In 2022, after one year of the Program, a study was carried out which, among other things, identified that requirements engineering (RE) knowledge and practices were not well established among professional members of the startups. Given the importance and potential impacts that RE can have on the success of projects, it was then decided to deepen the study on RE in StartUps. In this sense, this article aims to discover whether the professionals hired for the Startup Gov.br Program feel motivated and have the necessary knowledge in requirements engineering. To this end, the IMI-Intrinsic Motivation Inventory questionnaire was used, as well as a relationship between skills and knowledge of the requirements discipline, present in the*

Guide for the Development of Products and Solutions (GPS). This guide was created for digital teams within the Brazilian government. Furthermore, we sought to map the environment in which these professionals are located and their motivation in learning Requirements Engineering tools and techniques. As a result, a map of the degree of knowledge of program participants in requirements engineering techniques and tools is presented. The elements that make up the motivation of the participants are also identified, which was presented separated by role. For the researchers, the result was satisfactory from a statistical point of view, as well as helping to generate insights for improving RE within the scope of Brazilian government startups.

- D. De Almeida Jubé, C. C. Wermelinger, R. C. G. S. van Erven and F. N. B. De Souza E Edison Ishikawa: **Optimization of vehicle routing in the distribution of electronic voting machines for elections [25].** 18th Iberian Conference on Information Systems and Technologies (CISTI), Aveiro, Portugal, 2023, pp. 1-6, doi: 10.23919/CISTI58278.2023.10211450. *Abstract: Optimizing the variables involved in the transport and distribution of a good is a complex task present in many types of business. This study presents a solution to define optimized routes for the delivery of electronic voting machines to polling places in the Federal District (Brazil) for an election. The challenge was characterized as a Vehicle Routing Problem (VRP) and the solution based on the open source software OptaPlanner.*

1.8 Work Structure

This master's thesis is organized into the following chapters, in addition to this one, consisting of:

- **Chapter 2:** presents the literature review relating to the knowledge used in the work;
- **Chapter 3:** presents the theoretical framework as well as interesting related works on the research topic;
- **Chapter 4:** describes the methodology used to identify studies and implement the tool;
- **Chapter 5:** presents the UX evaluation model obtained after systematic mapping and identification of related works.
- **Chapter 6:** presented the experiment used to test the model, the main findings, and the discussions about it;
- **Chapter 7:** presents the main conclusions of this work and lists future work.

Chapter 2

Literature Review

2.1 User Experience (UX)

Conscious experiences are related to the “context of experience”, a set of basic elements of experience, which involve person, event, environment, and point in time. It can be said that “an individual experiences an event (i.e., action) through a specific medium at a particular point in time”. It is clear that if any of the four elements change, the experience will also be different [26].

Besides the concept of experience can be structured, the experiences themselves are subjective, directed, and multi-dimensional. Saying that it is “subjective” implies that only the individual can report whether the content of the experience was positive or negative and its intensity.

Being “directed” means that the experience is related, as seen, to specific people, events, means, and times. The subject evaluates this input, generating a response, which may be an emotion. Experiences, therefore, have a beginning, middle, and end.

Being “multi-dimensional” means that experiences encompass different mental responses. The experience covers six dimensions: affective, cognitive, physical, relational, sensorial, and symbolic. People develop mental responses to each dimension [26]. Each of the six dimensions can be described as[26]:

- affective: related to the emotions experienced by the individual, interpreted as pleasurable, such as happiness and love, or unpleasant, such as anger and sadness;
- cognitive: related to intellectual stimuli and learning during interaction;
- physics: related to the individual’s perception of the movement and position of the body during the interaction;

- relational: refers to the individual's relationship with brands or other individuals during the interaction;
- sensory: refers to the use of the five senses during interaction - vision, hearing, smell, touch, taste;
- symbolic: refers to the individual's self-affirmation and self-expression during the interaction.

This way, user experience evolves multiple internal responses about an event and is not limited to task-oriented approaches. Beauty, fun, pleasure, and personal growth are relevant to human needs but have little instrumental value. We can see user experience as a system with three components that determine a user's evaluation of a system: perceptions of instrumental qualities, which are related to the usability and usefulness of the system; perceptions of non-instrumental qualities, which are related to the appeal and attractiveness of the system, and emotional reactions, influenced by these two quality aspects [27].

Some systems, such as business management systems, are predominantly instrumental, and others are predominantly non-instrumental or hedonic, such as a cinema *site*. In the former, perceived usefulness presents itself as the predominant value. In the second case, perceived pleasure is preponderant in explaining user acceptance [27].

Enchantment with utilitarian products is not commonly reported, but it is possible. Enchantment with technology refers to being carried away by the power behind technology. It can also be seen as a disorientingly pleasurable feeling of fullness and aliveness. The enchantment can last longer when discoveries can be made regarding aspects or qualities [28].

In the context of enchanting experiences, feeling, sensation, and emotional consciousness must be given the same relevance as thought, consciousness, and will. Lasting enchanting experiences need depth that allows the longer a product keeps a person engaged, the more new things are discovered about it [28].

Usability tests can be performed to measure a system. Measurement using traditional usability tests focuses on evaluating first-time experiences of systems and the ability to learn about them. These usability tests find immediate system usage issues that may have little long-term impact [27].

Exploratory studies conducted by Karapanos et al. [29] suggested that the types of errors and causes of user frustration change significantly over time. At the beginning of the experiment, the evaluation judgment is made based on pragmatic aspects of usefulness and usability. After four weeks, the judgment is made based on the identity that the products express about their owner. Another interesting point is that the news has a lot of relevance on the system's beauty evaluation, but this decreases over time as the product is used [29].

Another important point about user experience evaluation is the way how it's created. It's formed from memories of past experiences and not from the sum of details of individual experiences, mainly because people cannot remember all the details of these experiences, and the person's memory can be biased. The most relevant moments of the experience for evaluation are the general evaluation peaks and the final intensity of the experience [27].

In the same way, the reconstruction of memory through retrospective recall, even with possible biases, has great relevance in determining the final experience and guidance about the person's future behavior, including their willingness to report the experience to others. Another interesting aspect is that the reasons for improvements in product experience evaluations are associated with the social position the product provided to the user in their contacts [27].

The relationship among memory, user experience, and emotions has already been studied. Setchi and Asikhia [30] suggest "the user experiences are stored in the long-term memory along with emotion. During the retrieval process, the emotion acts as a reminder of the previously stored information. In user interactions with products, emotional responses are generated based on the 'affective tag' attached to similar situations experienced in the past" [30].

2.1.1 User Experience Evaluation Tools

We obtained studies related to user experience evaluation tools. Several tools analyzed are Likert scale [31] questionnaires to collect users' opinions about the product or their interaction with the product in the form of a self-assessment. Other tools aim to evaluate the overall usability of the product in an automatic or manual way. Some tools pretend to support user experience teams using UX metrics collected from the user.

Vaataja et al. [32] shown a tool for user experience evaluation based on a questionnaire containing three essential groups. The first one evaluates the pragmatic quality of the product, like usability and task/goal achievement. The second one evaluates the hedonic aspects of the product, like stimulation or identification characteristics. The third one evaluates judgments like satisfaction.

Lachner et al. [33] proposed a tool to measure, visualize, and communicate a product's UX within organizations. Their tool is based on analyzing 84 UX evaluation methods and 24 interviews with experts from academia and practice. They identified nine dimensions into three areas of the product's UX. Each dimension has several questions created using Likert's scale [31].

Desolda et al. [34] presented a tool to exploit Machine Learning (ML) for automatic UX evaluation by analyzing the log data of the users' interaction with websites. The tool uses a javascript snippet integrated into the website to extract mouse movements and keyboard presses. The ML model predicts the user's average emotions, shown in a heatmap dashboard. The heatmap color distinguishes the user's emotions using the website.

Ntoa et al. [35] presented a tool to support evaluators who performed user testing experiments. This tool provides data visualizations for the experiment, the system, and the user. It synchronizes data with video records and manual evaluator inputs. The tool can track structured or unstructured tests and show a timeline of the interaction session data (e.g., user tasks) and insights from the experiment based on all the users involved.

Oliveira et al. [36] developed the UXNator tool for recommending UX methods based on a questionnaire with two questions: "In what phase is your project?" and "Who is the evaluator of your project?". After receiving these two answers, the tool suggests a set of UX methods the user may need in your project phase.

Franco et al. [6] created UXmood, a tool to provide information to assist researchers and practitioners in evaluating user experience and usability. The tool works like a media player and allows a review of the user interaction associated with usability data and sentiment analysis techniques. UXmood focuses on visualizing experiment data. It has a project management menu, an interaction area to review the experiment, legends about presented data, and a user interaction visualization area with sentiment analysis and quantitative data. Its sentiment analysis manager module recognizes user sentiment through a multimodal classification of audio, video, and transcribed speech data. It classifies the emotion and determines the emotion's valence. A view of this tool is shown in Figure 2.1.

UXmood uses think-aloud protocol [6]. This method is one of the most popular methods of UX evaluation [37]. It is widely used in product design evaluation experiments and suggests people use language to express their thoughts about the experience with a product. The think-aloud protocol can allow sentiment analysis from the user's speech because the utterances contain task-related and affect-related words and phrases. Seitch and Asikhia [30] showed an example: “The use of the down button to increase the time seems weird”. The phrase “down button to increase” refers to the task of using a design feature, and the word “weird” refers to the experience of using this feature [30]

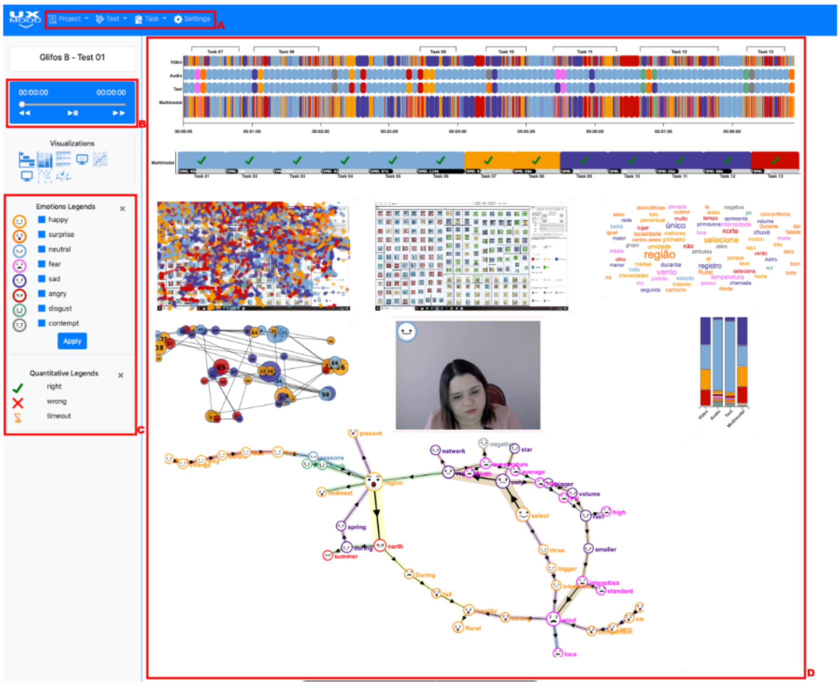

Figure 2.1: A UXmood tool's view [6].

2.2 User Satisfaction Measurement

User satisfaction can be defined as an overall assessment of the performance of various attributes that make up a product or service. User satisfaction increases user loyalty and retention. Given this, it is essential to know what the contributions of a product or service are to user satisfaction and which points reduce this satisfaction [38].

The level of satisfaction with a service results from the comparison between the reference that the user has and the service actually provided. If the reference is higher, this results in dissatisfaction; if it is lower, it results in satisfaction [38].

User satisfaction also can be seen as evaluating a product or service to meet that user's needs and expectations. Although satisfaction is necessary for user loyalty and retention, it is not a sufficient factor, as user commitment and involvement with the service also play an essential role [39].

The objective of evaluating user satisfaction is to provide interested parties with reliable information on aspects of their investment and provide managers with information that allows good decision-making about market behavior [40].

Several factors influence rating user satisfaction. During evaluation, temporary conditions and circumstances impact the user's grade. Also, there isn't a standard way to assess satisfaction. It can occur in writing, by interview, telephone, etc. Each assessment way can impact the user differently [39]. These factors are important to justify situations like when a test participant provides a favorable assessment with the SUS questionnaire [41] yet fails to complete the tasks being tested [42].

Measuring user satisfaction can be done using an assessment scale. More options tend to improve the quality of the analysis but may result in a reduction in the number of responses. The choice for the number of response options should vary depending on the nature of the problem, the participants' involvement in that problem, the demographic aspects of the participants, and the nature of the collection methods [43].

Given the need to compare the results of organizations, some institutions have proposed standardized methodologies for evaluating user satisfaction, such as *the American Customer Satisfaction Index* and *European Customer Satisfaction Index* [40].

Measuring user satisfaction can be used to indicate the usage and effectiveness of an information system. User satisfaction can be measured as the weighted sum of positive and negative user reactions to a set of information system factors. A system is said to be "good" when the user is very satisfied with the factors considered most important in the system [44]. Bailey & Pearson defined 39 factors that influence user satisfaction, like accuracy, timeliness, precision, reliability, and completeness [45].

User satisfaction can be measured using a questionnaire. To this end, three criteria must be considered. The first is the clarity with which the concept of satisfaction is defined when developing the questionnaire. The absence of this clarity can impact item generation, analysis, and validity investigation. The second is that the questionnaire must be developed by the psychometric method selected for the assessment to demonstrate its reliability and validity. The third is that the usefulness of the questionnaire must be evaluated, which is done by determining the degree to which the contexts of use and development are close [46].

2.3 Emotion Recognition

Emotions have an important role in thinking and rational behavior in everyday life. The same regions in the brain that process emotions also perform pattern recognition before information is passed on for rational processing, especially in terms of visual and auditory signals [47].

The ability of machines to understand and deal with emotions is related to their ability to learn preferences and adapt to what is important. Picard et al. [47] defined emotional intelligence as "the ability to recognize, express, and have emotions, along with the ability to regulate those emotions, harness them for constructive purposes, and deal skillfully with the emotions of others" [47].

Some theories point out that the human-machine relationship is based on the same aspects as human-human relationships, so it can also be considered a natural and social relationship. In this context, machines are expected to acquire some human emotional skills, especially recognizing affective *feedback*. This can be used for the machine to learn when to interrupt the user without making the user angry [47].

Considering that emotion alters several physiological conditions, such as facial expression, tone of voice, skin temperature, and others, methods that act on more than one condition to recognize emotions and obtain contextual information from the user have a better chance of being precise. In some cases, not even a human can identify a particular emotion other human experiences. Emotion recognition is accurate enough if it can understand the emotion in the same way a human would [47].

A human's recognition of emotions in speech with content that does not hint at the emotion is about 60% accurate when asked to classify the emotion of the speech in terms of six affective labels. Visual recognition of emotions by humans has greater accuracy and varies between 70% and 98% when choosing between six categories of facial expressions [47].

On the other hand, traditional human-computer interaction ignores the user's affective states. It focuses on sending and receiving explicit information, such as that sent by devices such as a mouse and keyboard. Much of the information in the interaction is not used, and the relationship with the system is seen as socially inept. User-centered systems must detect and respond to changes in the user's affective and non-affective behavior [48]. Most automatic affect analysis techniques consider the following items [48]:

- approaches trained in deliberately exaggerated affective expressions;
- approaches that recognize the set of basic expressions: happiness, sadness, anger, fear, surprise, and disgust);
- unimodal approaches limited to image or speech signal recognition.

Automatic emotion recognition can use basic emotions to describe affective situations in everyday life. Still, it is also quite limited to the possible emotions present in natural communication. Furthermore, the person's affective state and environment form the perceived emotion. Emotion recognition can be used with posed images or spontaneous images. Deliberate emotion recognition is more tractable than spontaneous emotion recognition. The difficulties inherent to the complexity of the spontaneous interpretation of emotions can impact the method's precision. The joint use of audio and video leads to better performance in recognizing affective behavior due to complementary information in both channels. Facial expressions, despite this, represent the most important affective cue [48].

Facial expression recognition should not be confused with emotion recognition. The former deals with classifying facial movement and its decomposition into classes based on visual information. The second deals with interpreting an emotional situation in a context and can involve several factors that culminate in the representation of emotions, such as voice, pose, gesture, direction of gaze, and facial expressions [49].

Facial recognition can occur through interpreting the emotion in the message received and evaluating linguistic signals. This method is considered more objective and leaves the subjective assessment of the message to a higher-level layer, such as a human evaluation, while focusing on mapping expressions and gestures to emotions. The most common method using this approach is the "Facial Action Coding System" (FACS). This system maintains a few dozen action units (AU), representing thousands of possible facial expressions [48].

Facial expressions occur through contractions of the facial muscle that last between 250ms and 5s. The measurement of facial expressions considers their location, intensity, and dynamics. The intensity is related to the expression's geometric deformation level about the original face. This is a point of attention in facial expressions obtained from poses, as they are usually exaggerated, unlike spontaneous facial expressions [49].

The dynamics of the expression must also be considered for its determination. This aspect considers the duration of the expression and the sequence of changes. Three temporal parameters are typically used to classify the expression: beginning or attack, apex or support, and displacement or relaxation [49]. Facial expression recognition is carried out in three stages [49]:

1. acquisition: this step detects the face in complex scenes and cluttered backgrounds using an automatic detector and is concerned with the pose used, that is, the distance and angle of the face in the scene, and the lighting, that is, differences in illumination incident on the face and levels of light reflection from eyes, teeth, and wet skin;

2. facial feature extraction: this step extracts information related to facial expression and can be:

    - carried out with holistic methods, which process the entire face, or with local methods, which process only the areas related to facial expression;
    - image-based, which extracts all information from the image without relying on extra resources, or model-based, uses 2D or 3D facial models to map facial features.

- deformation-based, which uses geometric deformations comparative to the neutral face, or movement-based, which focuses directly on the facial changes resulting from the performance of the expression;
- appearance-based, which focuses on the effects of facial activity, or muscle-based, which attempts to infer muscle activity from visual information;

3. facial expression classification: is the last stage of recognition and is carried out using facial action coding schemes based on signals or combination with judgments or structures based on dictionaries.

Facial expression measurement also has some concerns: a) distinguishing easy-to-recognize action scores from difficult ones; b) use of spontaneous actions instead of poses; c) differences in the characteristics of individuals in the population, such as babies and older people; d) definition of a minimum limit of intensity of facial expression; e) the level of experience of those measuring facial expression; f) the reliability of the type, intensity, and dynamics of facial action [49].

Voice recognition is also distinguished in terms of message content brought by verbal language and paralinguistic elements, such as tone or voice timing. It is possible to associate words with emotions, but this approach has proven ineffective. Paralinguistic elements provide more information that can be used to map affect in terms of objective linguistic signs [48].

Context is necessary for correctly identifying emotion in terms of voice and visual terms. Knowing where the facial expression or voice occurs, for example, on the street, at home, or at work, can be decisive for correctly identifying the emotion. Likewise, using deliberately created models of linguistic signals does not favor identifying signals that occur spontaneously, which are normally more subtle and can be captured with noise or inaccuracies resulting from the environment and the way of capture. [48].

2.4 Chapter Summary

This chapter presented the literature review necessary for understanding the issues that follow from here. We highlight that experiences are not limited to tasks and that aspects such as beauty and pleasure play an important role. However, some systems, such as a business management system, are strongly instrumental. Others focus more on non-instrumental features, such as a cinema site. Also, past and current experiences form the evaluation of the experience. It is not formed as the sum of all its details but as a retrospective memory. As such, this recollection may suffer from memory reconstruction biases. The most relevant moments for evaluating the experience are the general evaluation peaks and the final intensity of the experience.

Emotions strongly correlate with rational behavior, so both behaviors support each other. The capacity of machines permeates the need to deal with emotions to adapt to what is important. Machines are expected to recognize affective feedback, among other things, to better deal with humans. The recognition of emotions in a spontaneous environment has more significant complexities than that carried out deliberately due to noise and inadequacies in the environment for capturing relevant information.

Chapter 3

Theoretical Framework

3.1 User Experience Elements

The definition of user experience is essential in the context of this study to clearly define what is being evaluated and measured. We found several studies that define user experience differently. Park et al. [4] defined user experience as "everything that happens to us, from which we may obtain knowledge, feelings, and skills". Experience "includes all our routine activities, such as face-to-face relations and religious activities, as well as brand, product, or service experiences" [4].

The concept of user experience can be very wide. User experience can also be seen as "brand experience", another essential type of experience that includes all interactions with the corporation and its branded products and services. Factors that influence brand experience may be brand loyalty, brand awareness, attitude to brand, brand ethics, and experiences with products or services [4].

When we reduce the analysis to product and service experiences, the user experience can be defined as "an overarching experience that consists of all aspects of users' interaction with a product or service". The elements of user experience are usability, affect, and user value [4]. Figure 3.1 shows the UX experience model proposed by Park et al. [4]. The satisfaction dimension can also be considered as usability, affect, or user value, so, in this work, we consider it as user value. We can show the elements' definition and their dimensions as follows [4]:

- Usability: It's defined as "the effectiveness, efficiency, and satisfaction with which specified users can achieve specified goals in particular environments". Each dimension has several sub-elements, like simplicity, visibility, consistency, and error prevention.

- Affect: It's defined as "a neurophysiological state consciously accessible as the simplest non-reflective feelings evident in moods and emotions". In the context of UX, affect is "considered as an emotion that is a consequence of interaction with a product or service". It has sub-elements like delicacy, texture, color, and attractiveness.
- User Value: It's defined as "a subjective value that the user attaches to a product. The value may be related to how the user thinks the product is meaningful and significant in his or her life. This element of UX is correlated with symbolic association", which is "determined by what the product is seen to symbolize about its user or the social-cultural context of use". It has sub-elements like identity, confidence, fun, utility, expectation, and social value.

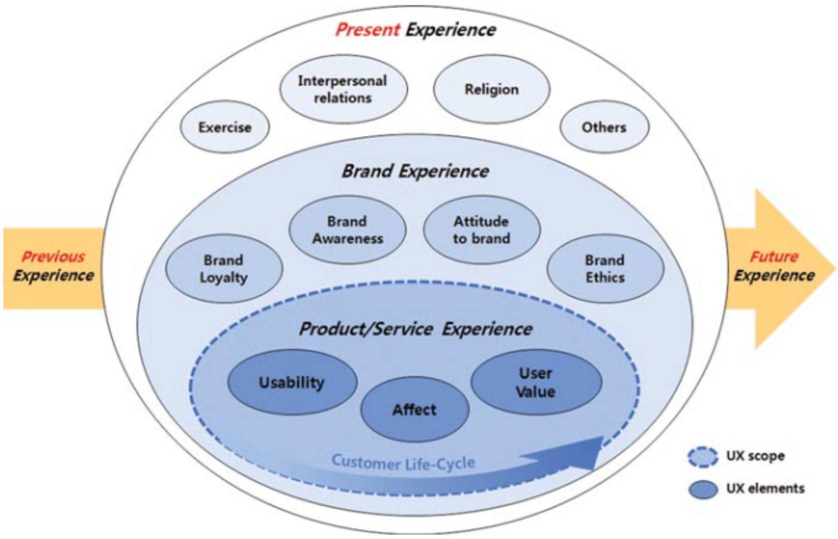

Figure 3.1: User experience model [4].

Alben [50] presented user experience as "all aspects of how people use an interactive product: how it feels, how understandable its operation is, how it serves people's purposes, and how it fits the context in which you are using it". The quality of the experience occurs when the experience occurs successfully and engagingly [50]. For the quality of the experience to occur, other elements of the user experience must be present. These elements are [50]:

- User understanding: how well did the design team identify user needs and reflect this in the product?
- Effective design process: Is the product design the result of a well-planned design process?
- Needed: What does the product achieve, and what is its social, economic, or environmental contribution?

- Learnable and Usable: Is the product easy to learn and use?
- Appropriate: Does the product serve users in an efficient way?
- Aesthetic experience: Is using the product aesthetically pleasing?
- Mutable: Does the design allow the product to change and evolve?
- Manageable: Does the product design support the entire context of use?

Desmet & Hekkert [7] also presented a framework for product experience. They defined it as “the entire set of affects that is elicited by the interaction between a user and a product, including the degree to which all our senses are gratified (aesthetic experience), the meanings we attach to the product (experience of meaning) and the feelings and emotions that are elicited (emotional experience)”. The framework is illustrated in Figure 3.2 [7].

- aesthetic experience: this level considers a product’s capacity to delight one or more of our sensory modalities.
- experience of meaning: this level considers cognition processes, like interpretation and memory retrieval, to recognize metaphors, assign personality, and assess the personal or symbolic significance of products.
- emotional experience: pleasant emotions pull us to products that are beneficial, whereas unpleasant emotions will push us from those that are detrimental to our well-being.

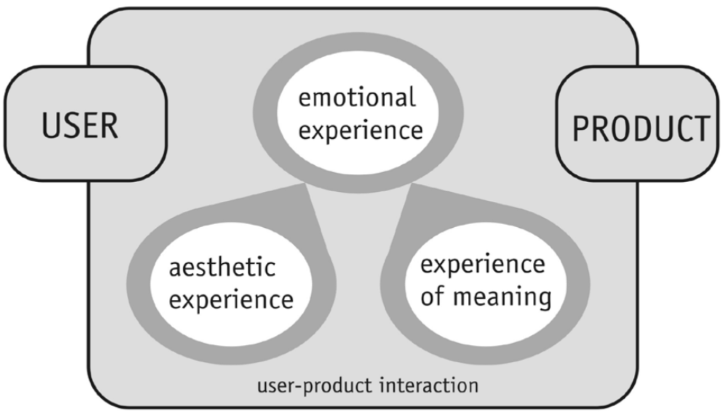

Figure 3.2: Framework of product experience [7].

3.1.1 System Usability Scale (SUS)

Usability can be defined as a “general quality of the appropriateness to a purpose of a particular artefact”. In this way, It’s impossible to specify the usability of a system without identifying its users and tasks, as well as the environment in which it will be used [41]. According to ISO 9241-11, usability should cover three elements [41].

- effectiveness: the ability of users to complete tasks using the system and the quality of the output of those tasks;
- efficiency: the level of resource consumed in performing tasks;
- satisfaction: user’s subjective reactions to using the system.

SUS is a simple, ten-item scale giving a global view of subjective usability assessments. SUS uses a five-point Likert scale [31]. In each item, the user should indicate the degree of agreement or disagreement [41]. The SUS questions are shown as follows [41].:

- 1. I think that I would like to use this system frequently
- 2. I found the system unnecessarily complex
- 3. I thought the system was easy to use
- 4. I think that I would need the support of a technical person to be able to use this system
- 5. I found the various functions in this system were well-integrated
- 6. I thought there was too much inconsistency in this system
- 7. I would imagine that most people would learn to use this system very quickly
- 8. I found the system very cumbersome to use
- 9. I felt very confident using the system
- 10. I needed to learn a lot of things before I could get going with this system

After the users evaluate the system, they should record an immediate response to the SUS questionnaire. The SUS score should consider all the items. To calculate it, we have to sum all the items’ contributions. The score position for an item ranges from 1 to 5, and each item’s score contribution ranges from 0 to 4. For items 1, 3, 5, 7, and 9, the score contribution is the scale position minus 1. The rest of the items have a score contribution of 5 minus the scale position. In the end, multiply the sum of the score by 2.5 to obtain the overall value of the SUS [41].

SUS is an effective tool to measure the usability of products and services. Besides, it isn't easy to understand what one individual evaluation with SUS means. The question "What is the absolute usability associated with any individual SUS score?" is not answered by SUS. To solve this problem, it's possible to add an adjective rating scale to the SUS [42].

Using one additional item to the SUS scale can give an overall understanding of the user's satisfaction with usability [42]. Despite that, all SUS items still remained important. Bangor et al. [42] proposed to include the following item [42]: 11. Overall, I would rate the user-friendliness of this product as: Worst Imaginable, Awful, Poor, Ok (So-So), Good, Excellent, Best Imaginable.

3.2 Related Work

This work focuses on evaluating the user experience of digital products. This research used the user experience concept developed by Park et al. [4]. The definition of UX and its elements was used as a working guideline to create the proposed model to measure the user experience. In the user experience discipline, we asses usability, affect, and user value. Then, we use these results to form a final qualitative UX index. Other authors helped us develop the proposed model, as we'll see.

Park et al. [4] study, presented in Chapter 3, showed UX as composed of usability, affect, and user value elements. This work used these elements as the fundamentals of UX measurement. Identifying positive, neutral, and negative interaction points can determine the automatic measurement of usability since the points the user has difficulty with are mapped by the negative emotions expected to be captured. The effect is identified with continuous emotion recognition, which is the emotion related to the product's use. The user value is captured with the overall satisfaction measurement proposed. All the three elements are necessary and complementary.

Brooke's study [41] was very helpful in understanding the need to identify the user's purpose to measure a system's usability. This question was included in the model with the SUS application. The SUS is used to evaluate the usability, and it should be added with the question proposed by Bangor et al. [42] to allow a better understanding of the SUS score.

Regarding user experience tools, the tool proposed in this work has a different approach from the ones listed in subsection 2.1.1. All the other tools are developed to support an evaluator in evaluating a website or digital product. To do so, these tools offer a lot of user data, suggested methods, or questionnaires that a UX expert should analyze. This tool provides a final user and user-friendly user experience indicator, i.e., any person can use this tool on a web/mobile digital product and understand its results. This tool offers a standalone evaluation as well as a long-term evaluation that considers both the user and the product.

Balbin et al. [10] stated that user satisfaction is decisive in the success of a business. Research was carried out to provide a device that can obtain, based on facial expressions, the satisfaction of the user enjoying a meal in a restaurant. To this end, an external camera, augmented reality glasses, and a personal computer were used. The *Affdex Software Developer's Kit* was also used to recognize emotions [10]. The authors identified the emotions of anger, disgust, fear, joy, sadness, and surprise. When more than one emotion was detected, the one with the greatest intensity was chosen. Emotions were classified as positive (joy), negative (anger, disgust, fear), and neutral (surprise). As surprise can be positive or negative, the second emotion was considered. Satisfaction was identified if the emotion was positive and dissatisfaction if the emotion was negative [10]. It was thus possible to measure user satisfaction based on emotion recognition, and, using chi-square tests, an association between user satisfaction and the taste of the food dishes tested was verified, which was the study's objective. The ease of use of the system and delivery of results in real-time were considered essential results of the study [10].

Chimienti et al. [11] used emotion recognition and behavioral analysis to create a personalized experience for users of an interactive movie selection system. The paper authors developed for the system a module for recommending films to watch to increase user involvement and improve satisfaction, and a behavioral analysis module to understand the user's needs when interacting with the system to enhance and personalize recommendations [11]. It was possible to improve the user experience and usability of the product through emotion recognition using recommendation and behavioral analysis modules. Initially, a first interaction allowed mapping of the user's emotions and behaviors. Next, the system was presented with the customizations from the mapping carried out in the previous phase [11].

In a complementary way, the personalized system used the mapping of the user's gaze, the mouse's position on a screen item, and the user's clicks. To recognize facial expressions, they used a facial expression model, "Fer2013", which has around 30 thousand images, including anger, disgust, fear, happiness, sadness, surprise, and neutral. This model offered a theoretical 96.7% accuracy. The *deep learning* algorithm used was the deep neural network "ResNet50", which was trained in 50 epochs and obtained 86% accuracy when using three of the seven possible emotions: happiness, sadness, and neutral [11].

Soleymani et al. [12] distinguished emotion recognition from sentiment analysis. While the first "is the automatic identification of an episodic emotional reaction, often from a single person", the second seeks "the automatic recognition of the polarity of opinions, that is, positive or negative". Emotions are short-term and belong to a single person, while opinions are shared by multiple people and are long-term. However, emotion recognition can be used to indicate the polarity of feeling [12].

Emotion recognition benefited from studies on identifying feelings using computational resources. This analysis, which was previously carried out in a primarily textual manner, began to be carried out in a multimodal way, involving text, image, and video. The use of multimodal approaches can generate greater reliability and robustness compared to the use of uni-modal methods due to complementary information present between the different channels [12]. Extracting multimedia information to summarize user ratings and opinions has become commercially available to aggregate opinions at scale and to obtain instant feedback. Companies using surveys or focus groups to get the same result is a much slower and more expensive activity [12].

Landowska [8] conducted a study to evaluate several emotion recognition techniques in usability tests. It was pointed out that "a study of fluctuations between positive, negative and neutral emotions throughout the application interaction process can provide valuable information about the overall user experience" [8]. It was observed that affect recognition has already been used to measure user experience in usability tests with four different scenarios: first impression test, task-based usability test, free interaction test, and comparative test [8].

It was verified in the studies by Kolakowska et al. [51] that important emotional states in the context of usability evaluation were frustration, empowerment, interest (excitement), boredom, disgust, engagement, and discouragement. It was observed that it was possible to map these emotional states based on Ekman's six basic emotions [52], happiness, sadness, anger, fear, disgust, and surprise, as presented in Figure 3.3 for frustration, boredom, interest (excitement) and empowerment [51] [8].

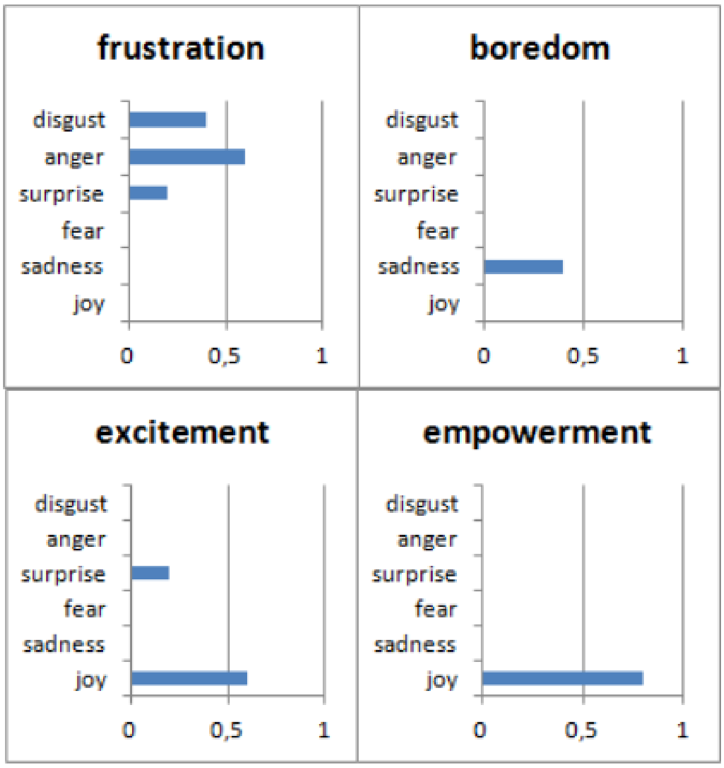

Figure 3.3: Mapping emotional states by Kolakowska et al. for Ekman according to Landowska [8].

Landowska [8], to evaluate emotion recognition techniques, brought usability test participants together in the same room. The techniques considered were [8]:

- questionnaire: application of a written questionnaire to identify the participant's emotional state;
- facial expression analysis: use of algorithms to analyze facial muscle movements to recognize emotions based on a previously obtained model of facial expressions;
- body posture analysis: analysis of body posture recognition based on gestures and poses obtained from the recorded video of the participant's posture;
- patterns of peripheral use: analysis of emotions based on information obtained from touching the keyboard and using the mouse;
- speech prosody: recognition of emotions based on changes in the participant's voice and intonation;
- sentiment analysis: recognition of emotions based on text analysis, often mined from public opinions;
- physiological measurements: use physiological information obtained through sensors to assess the participant's emotion.

The summary of the results in terms of emotion recognition accuracy and robustness to disturbances is presented in Table 3.1.

Emotion exploration technique used	Accuracy and granularity of emotion recognition	Robustness to disorders	Independence of human will	Interference in usability testing procedures
Quiz	Low	High	Low	None
Facial expression analysis	Medium to high	Low	Low to medium	None
Body posture analysis	Low	Medium	Low to medium	Low
Peripheral Usage Patterns	Low	Medium	Medium	None
Speech Prosody	High	Low	Medium	Medium to High
Sentiment Analysis	Medium to High	Medium	Low	Low
Physiological measurements	High	Medium to high	Very high	Medium to high

Table 3.1: Summary of the use of emotion recognition techniques in usability testing according to Landowska [8].

To compare and identify the difference between this work and related works identified, a summary is presented in Table 3.2 with the main characteristics of the works reported and proposed in this work.

Criteria / Study	**Balbin et al. [10]**	**Chimienti et al. [11]**	**Soleymani et al. [12]**	**Landowska [8]**	**This study**
Type	Individual experiment.	Individual experiment.	Systematic Literature Review.	Experiment in a room with several participants.	Individual experiment.
Goal	Automatically measure customer satisfaction when enjoying a meal in a restaurant.	Increase user satisfaction by creating a personalized experience of an interactive movie selection system.	Review developments in multimodal sentiment analysis across different domains.	Evaluate various emotion recognition techniques in usability tests in a call center system.	Automatically measure the user experience with a digital product.

Table 3.2 continued from previous page

Mode	- Task-based taste test; - Carrying out five tasks; - Collects the dominant emotion and evaluates its valence; - The valence of the emotion assessed measures test satisfaction.	- Task-based usability testing; - Carrying out six tasks; - Collects emotion according to valence; - Collected emotion fed recommendation module; - Test satisfaction is measured by administering questionnaires at the end of each task.	- Problem definition; - Analysis of findings in papers.	- Task-based taste test; - Carrying out tasks and filling out questionnaires; - Identification of all moments of satisfactory conditions for using a technique and inadequate environmental conditions for using the techniques, to allow mapping of usage challenges.	- Usability testing for new or known users to make real use of the product; - Collects emotion and user's screen continuously; - Continuous evaluation throughout the use of the product allows mapping positive, neutral and negative points of the product; - Possibility of evaluating user experience over several interactions with the same user and discovering possible changes in satisfaction levels over time. - Application of a questionnaire at the end of the interaction to complement and validate the evaluation.
Techniques	- Facial expressions.	- Facial expressions; - Mapping the User's Gaze; - Peripheral Usage Patterns (mouse).	- Distinction between concepts of affection, feeling, emotion, and opinion; - Identification of sentiment analysis approaches; - Identification of challenges and perspectives.	- Quiz; - Facial expressions; - Body Posture; - Peripheral Usage Patterns; - Speech prosody; - Analysis of feelings; - Physiological measurements.	- Facial expressions; - Sentiment analysis - Similarity analysis.

Tools	- Use of augmented reality; - Affdex Software Developer's Kit; - Use of 6 emotions: happiness, sadness, anger, fear, disgust, and surprise (considers the second strongest emotion).	- Recommendations module; - Behavioral analysis module; - Model Fer2013; - Deep Neural Network (ResNet50), trained in 50 epochs; - Use of 3 emotions: happiness, sadness and neutral; - Expressing Mixed Emotions, SUS and SUPR-Q questionnaires.	Not applicable.	- Each participant performed the test in an individual space in front of the computer; - Side and computer cameras recorded video and ambient sound; - Participants could interact with each other.	- Automatic user experience measurement module – Captures the user's screen; – Capture user video; – Identifies emotions; – Presents an overall evaluation grade. - Use of 4 emotions: joy, sadness, anger, and surprise. - Application of manual questionnaire.

Table 3.2: Summary of the main characteristics of related works and this study [10] [11] [12] [8].

3.3 Chapter Summary

This chapter presented the theoretical framework and the related work on user experience assessment using emotion recognition and its associated aspects, such as usability tests and satisfaction measurement. The main point is the definition of the user experience's elements suggested by Part et al. [4]: usability, affect, and user value. The usability measurement tool SUS was presented with its calculation procedure [41].

Several relevant related studies were identified, including the study by Balbin et al. of automatic satisfaction measurement using augmented reality and the *Affdex Software Developer's Kit*, the study by Chimienti et al. to develop a recommendation module for a movie system using emotion recognition and behavioral analysis, and Landowska's study to carry out usability tests using different emotion recognition techniques in a room with several participants.

Chapter 4

Systematic Literature Mapping

A systematic literature mapping (SLM) overviews a research area by classifying and counting contributions from the classification categories. Systematic literature mapping differs from systematic literature review (SLR). Generally speaking, SLM will uncover research trends, while RSL will aggregate evidence related to specific objectives [53]. In this work, a systematic literature mapping was conducted to map the state of the art related to the use of Artificial Intelligence tools and techniques related to emotion recognition, especially facial expression recognition, to identify techniques, benefits, and challenges associated with the development of a new tool that allows measuring user satisfaction in an automated way.

The systematic mapping of the literature and the remainder of this work were conducted by the first author and reviewed by the advisor. The SLM process involved the research planning, conducting, and reporting phases. Each stage was carried out according to the following set of activities:

- planning: definition of the research protocol, quality assessment, and data extraction form for research questions;
- conduction: involves the search, import, and selection of studies, as well as quality assessment and data extraction from papers;
- report: involves analyzing study data, answering research questions, and identifying threats to the answers.

The systematic literature mapping process used in this work is presented in Figure 4.1. The *Parsifal* platform [54] was used to facilitate each stage of the process.

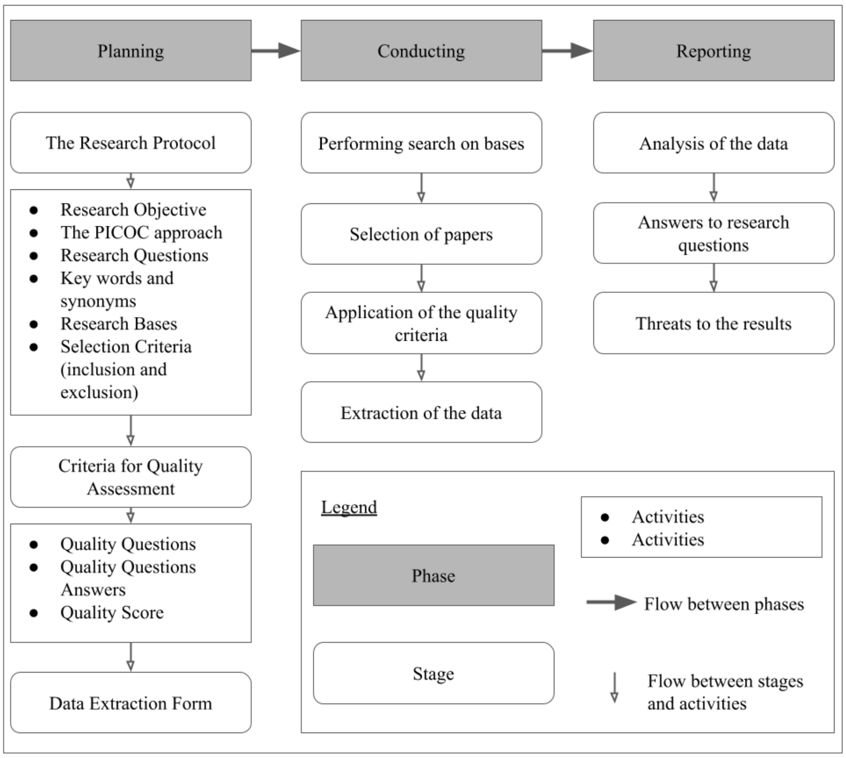

Figure 4.1: Systematic literature mapping used in this dissertation.

4.1 Planning

Planning the systematic literature mapping included defining the research protocol, quality assessment criteria, and data extraction form.

4.1.1 Research Protocol

The research protocol defined the elements necessary to define specific objectives for obtaining data, such as the research objective, the PICOC approach (*population, intervention, comparison, outcome, context*), and the definitions of research questions, as well such as selecting the appropriate bases for the query.

Research Objective

The research objective is to map the state of the art using Artificial Intelligence techniques related to emotion recognition, especially facial expression recognition. We sought to understand the main challenges and benefits of the techniques identified in the literature in the context of emotion recognition.

PICOC Approach

The elements of the PICOC approach were defined according to the model proposed by Petticrew & Roberts [55]. Among the five criteria of this approach, this mapping used three: population, intervention, and context. This is due to the more general nature of mapping, as seen previously. No research questions were created, and no keywords related to the other criteria were extracted. The definitions are presented in Table 4.1 [55].

Criteria	Definition
Population	Studies related to emotion recognition with artificial intelligence, starting in 2008 and focusing on facial expressions, associated with related aspects of user satisfaction assessment.
Intervention	Automated satisfaction measurement techniques for recognizing emotions visually using artificial intelligence.
Comparison	Manual satisfaction measurement techniques using customer satisfaction scores.
Outcome	Accuracy rate of automated measurement with artificial intelligence about manual measurement with scoring customer satisfaction.
Context	Benefits and challenges of visually recognizing emotions in measuring user satisfaction in human-computer interaction (HCI) of digital products.

Table 4.1: Presentation of the PICOC approach criteria applied to systematic literature mapping.

Due to advances in Artificial Intelligence, we must obtain the latest technologies and methods to recognize emotions visually. Therefore, we must take all relevant documents in this area. Consequently, we chose 2008 as the starting year of the SLM to cover the last fifteen years, as we considered it a sufficient period to find the most relevant papers. The research results are consistent with this temporal definition since the oldest paper selected from the subject of systematic literature mapping is from 2014.

Research Questions

It was then followed by defining the research questions and identifying the information necessary for the work. The research questions (RQ) are presented in Table 4.2.

ID	Research Question (RQ)
RQ.1	Are there studies in the literature related to emotion recognition in a visual way using Artificial Intelligence?
RQ.2	What are the visual emotion recognition techniques used in the literature related to automated satisfaction measurement using Artificial Intelligence?
RQ.3	What are the benefits and challenges of performing emotion recognition visually?

Table 4.2: Systematic Literature Mapping Research Questions.

Keywords and Synonyms

Before defining the search *string*, it was necessary to identify the keywords and their synonymous terms essential for the search to create a more robust search *string*. This survey is presented in Table 4.3.

Keyword	Synonym	PICOC
Emotion Recognition	Emotion Recognition Emotions Recognition Reconhecimento de Emoção	Population
User Satisfaction	Customer Satisfaction Satisfaction Measurement User Satisfaction Medição de Satisfação Satisfação do Cliente	Population
Intelligence Artificial	Análise de Sentimento Análise de Sentimentos Artificial Intelligence Facial Expression Recognition Reconhecimento de Expressão Facial Reconhecimento de Expressões Faciais Sentiment Analysis	Intervention

Table 4.3: Presentation of relevant keywords and synonyms in systematic literature mapping.

Search String

After mapping the keywords, it was possible to define the base of the *string* search. In mapping the PICOC approach, the aim was to identify papers that contained “emotion recognition” and “satisfaction measurement” and “Artificial Intelligence”, or any of the synonyms selected for these terms. This research was carried out in all fields of the paper.

Given the breadth of previous research and to favor papers that really dealt with "emotion recognition", a restriction on the presentation of this term in the title or abstract of the paper was added. Finally, another restriction was added, limiting the presentation of papers published from 2008 onwards to study papers with a greater chance of current relevance. The final search *String* was applied to all selected bases with minor adjustments to field names to suit the base syntax. The base version of the search string used is presented below:

("Emotion Recognition" OR "Emotions Recognition" OR "Reconhecimento de Emoção" OR "Reconhecimento de Emoções") AND ("User Satisfaction" OR "Customer Satisfaction" OR "Satisfaction Measurement" OR "Satisfação do Usuário" OR "Satisfação do usuário" OR "Medição de Satisfação") AND ("Sentiment Analysis" OR "Artificial Intelligence" OR "Facial Expression Recognition" OR "Análise de Sentimentos" OR "Análise de Sentimento" OR "Inteligência Artificial" OR "Reconhecimento de Expressões Faciais" OR "Reconhecimento de Expressão Facial") AND Title OR Abstract: ("Emotion Recognition" OR "Emotions Recognition" OR "Reconhecimento de Emoção" OR "Reconhecimento de Emoções") "filter": Publication Date: (01/01/2008 TO 12/31/2023).

Research Bases

Previous experiences by Petersen et al. were used to choose the base. [53] and Dyba et al. [56]. These study guides selected these bases using the relevance criterion in the context of *software*. Among the bases presented in the studies cited the primary bases that appeared in both *IEEE* and *ACM* were chosen, as well as the indexing bases for papers that appeared in these studies and that were accessible through CAPES, *Scopus* and *ISI Web of Science*. The research bases chosen to carry out the systematic literature mapping are presented in Table 4.4.

Search Base	**URL**
ACM Digital Library	http://portal.acm.org
IEEE Xplore	http://ieeexplore.ieee.org
ISI Web of Science	http://www.isiknowledge.com
Scopus	http://www.scopus.com

Table 4.4: Digital Bases Used in Systematic Mapping of Literature.

Selection Criteria (Inclusion and Exclusion)

The inclusion and exclusion criteria presented in Table 4.5 were used. The inclusion criteria were encoded in the *string* search, which favored the selection of papers with the subsequent ratification of the criteria's applicability.

ID	**Type**	**Criteria**
IC1	Inclusion	Publication must have occurred from 2008.
IC2	Inclusion	Publication deals with "emotion recognition" in a way related to "Artificial Intelligence" and "user satisfaction ".
IC3	Inclusion	Publication references "emotion recognition" in the title or abstract.
EC1	Exclusion	The central theme of the publication is not in the area of Computer Science.
EC2	Exclusion	The central theme of the publication is not "emotion recognition" in the visual modality (image or video).
EC3	Exclusion	Publication is not in English or Portuguese.
EC4	Deletion	Publication does not directly relate "recognition of emotions" with "user satisfaction".
EC5	Exclusion	Publication with less than six pages (Short Paper).
EC6	Deletion	Publication is a technical report, book chapter, master's dissertation, or doctoral thesis.

Table 4.5: Presentation of the inclusion and exclusion criteria used in systematic literature mapping.

The inclusion criteria used were motivated as follows:

- IC1: studies from the last 15 years were sought to contemplate the state of the art on the subject, but in such a way that it is possible to obtain information that is still relevant to guide new models of tools and experiments;
- IC2: criteria associated with the *string* search, created based on the *PICOC* [55] approach and the research questions;
- IC3: criterion added to favor papers that dealt with "emotion recognition" to avoid identifiable noise, the result of a vast number of studies, such as those that only cite papers related to recognition of emotions;

The exclusion criteria used were motivated as follows:

- EC1: exclude studies that fall outside the scope of systematic literature mapping.
- EC2: exclude studies that use emotion recognition in other modalities, such as voice, human-computer interaction elements, physiology, etc.

- EC3: exclude studies that cannot be analyzed due to language knowledge.
- EC4: exclude studies that are strictly technical or do not use emotion recognition for any purpose related to user satisfaction.
- EC5: exclude studies that are still ongoing or incomplete.
- EC6: excludes gray literature studies as they are not formally published and peer-reviewed.

4.1.2 Quality Assessment Criteria

It was also planned to apply a quality assessment to papers that passed the selection criteria. These criteria aimed to ensure that the selected papers answered the mapping questions.

Quality Issues

The quality questions presented in Table 4.6 were prepared. A quality question was associated with each research question. The questions were motivated by the following aspects:

- Q.1: when planning the quality of the work, it was observed that it was necessary to distinguish studies that aimed to measure user satisfaction using emotion recognition from those that merely mentioned user satisfaction throughout the text;
- Q.2: the objective of the question was to evaluate the degree of depth in which visual emotion recognition techniques were presented;
- Q.3: the objective of the question was to identify whether the benefits and challenges of recognizing emotions visually were discussed in the different areas of knowledge and whether use cases were presented.

ID	Quality Assessment Questions
Q.1	The study presented emotion recognition relationship in a visual way using Artificial Intelligence?
Q.2	The study presented emotion recognition techniques in a visual way?
Q.3	The study presented the current benefits and challenges of recognition of emotions visually?

Table 4.6: Presentation of questions for quality assessment of selected papers.

Quality Question Answers

The levels of compliance with the possible criteria for quality assessment were defined according to Table 4.7. A score is assigned to each answer.

Criteria Fulfillment Level	**Score**
Fully presented the quality criteria.	2
Partially presented the quality criteria.	1
Did not present the quality criteria.	0

Table 4.7: Presentation of levels of compliance with the criteria for quality assessment of selected papers.

Quality Score

The maximum score for meeting the quality criteria is 6 points, and the minimum is 0 points. A value of 3 points was established for the paper to reach the minimum level of quality.

This value was established to allow studies with one complete response and one partial or three partial responses to obtain a broader response based on multiple studies.

4.1.3 Data Extraction Form

The last step of the planning phase was the definition of the data extraction fields, which allowed the recording of answers to the research questions for the papers selected and approved in the quality criteria.

The research questions were used as unstructured text fields to record the information provided in the answers.

4.2 Conducting

The planned steps for identifying, selecting, and approving papers were carried out according to study selection and quality criteria. The workflow is presented in Figure 4.2.

4.2.1 Conducting Research in the Bases

The *string* search was applied to each database informed in the planning. Each database has a search syntax with particularities, including how to restrict the search by year. After inserting the search, the result offered by the database was stored in *bibtex* format

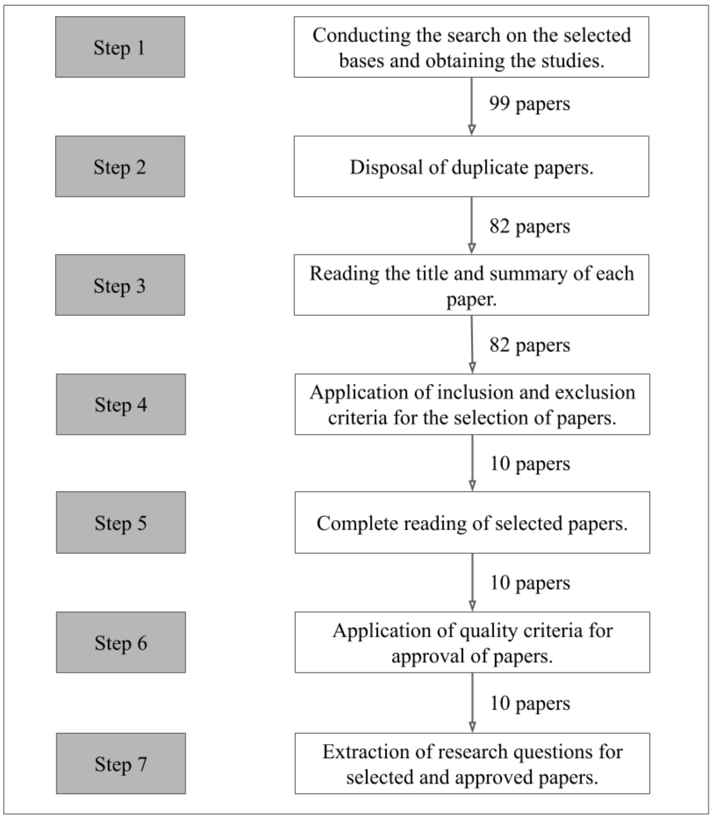

Figure 4.2: Systematic literature mapping workflow.

and imported into the *Parsifal* platform to be worked on. The total results found were 99 studies. Of these, 17 were eliminated for being duplicates, and 82 studies passed to the next phase.

4.2.2 Selection of Studies

Each study was opened individually to read the title and abstract. Based on this information, the inclusion and exclusion criteria were applied. Of the 82 papers, 72 were rejected, and 10 were selected for the next stage.

The selected papers represent 12% of the total papers brought without duplication by the search string in the selected databases. The reasons for rejection are shown in Figure 4.3.

4.2.3 Application of Quality Criteria

The ten selected studies were then read in full to be evaluated regarding quality criteria. After reading, a quality assessment was carried out. The authors considered all papers passed by the quality criteria. The selected studies are presented in Table 4.8

ID/Ref	Title	RQ that the study answers
S1 [12]	A survey of multimodal sentiment analysis	RQ.1 RQ.3
S2 [10]	Augmented reality aided analysis of customer satisfaction based on taste-induced facial expression recognition using Affdex Software Developer's Kit	RQ.1
S3 [11]	Behavioral analysis for user satisfaction	RQ.1 RQ.2
S4 [57]	Emotional valence from facial expression as an experience audit tool: an empirical study in the context of opera performance	RQ.1
S5 [58]	Gauging customer interest using skeletal tracking and convolutional neural network	RQ.1 RQ.2
S6 [59]	Recent trends in artificial intelligence for emotion detection using facial image analysis	RQ.1 RQ.2
S7 [60]	Retail managers' preparedness to capture customers' emotions: a new synergistic framework to explore unstructured data with new analytics	RQ.1 RQ.3
S8 [61]	Spatial augmented reality based customer satisfaction enhancement and monitoring system	RQ.1 RQ.2 RQ.3
S9 [62]	The relationship between human and smart TVs based on emotion recognition in HCI	RQ.1
S10 [8]	Towards emotion acquisition in IT usability evaluation context	RQ.1

Table 4.8: Studies selected in systematic literature mapping to answer research questions.

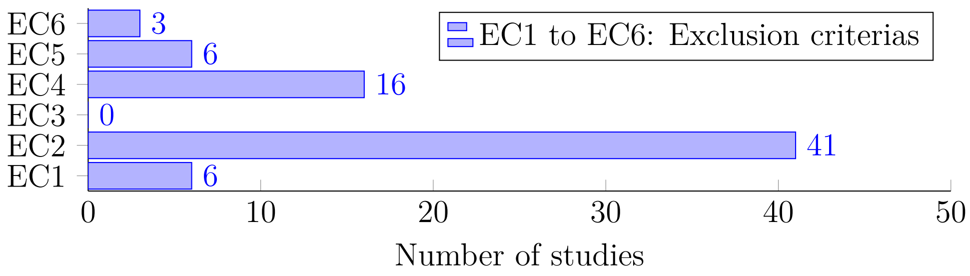

Figure 4.3: Paper rejection statistics by exclusion criteria.

4.2.4 Data Extraction

Data extraction was recorded while reading the studies to obtain information about the content of the studies to answer the research questions. All papers were read in full.

For each paper read, a summary was made with its objective, contextual information, and the central relevant aspect of the paper for systematic mapping, generally obtained from the results or conclusions. Furthermore, a mapping of the types of study of the papers was carried out, and each one was identified as:

- systematic literature review: a study that identified and reviewed other studies to present an overview of the subject in the literature;
- image capture experiment in a controlled environment: a study whose objective was to present the method and results of experiments carried out in an environment in which participants needed to follow well-defined instructions;
- image capture experiment in a spontaneous environment: a study whose objective was to present the method and results of experiments carried out in a free environment or in which participants had the autonomy to act flexibly and spontaneously.

4.3 Results

The selected studies adequately related elements of visual emotion recognition with Artificial Intelligence and user satisfaction and their related aspects. The ten papers approved in the selection and quality assessment were analyzed.

4.3.1 Data Analysis

Most selected studies referred to image capture experiments in a controlled environment, followed by image capture experiments in a spontaneous environment, and lastly, systematic reviews. The results of the studies selected by type of study are presented in Figure

4.4. This result is expected since conducting experiments in a spontaneous environment poses numerous challenges, such as insufficient lighting [8].

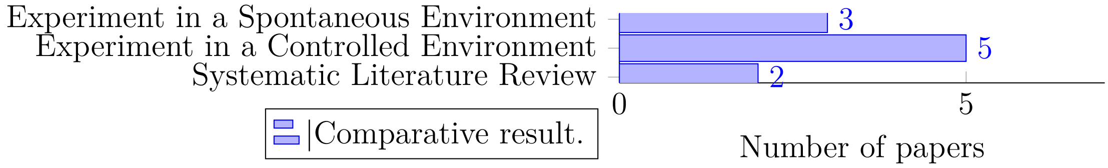

Figure 4.4: Papers selected by type of study.

4.3.2 Response to Research Questions

The results of the research questions are presented in the following sections.

RQ.1. Are there studies in the literature related to visual emotion recognition using Artificial Intelligence?

Emotions play a fundamental role in the user experience. They are considered essential reference points for user satisfaction and future purchase intention. Traditional emotion measurement uses self-assessment techniques, which, despite being easy and less expensive, can have limitations [57]. The answers provided in traditional methods may contain problems related to the subjectivity of the assessment, the person's unwillingness to respond, or some bias when remembering and explaining the answer. Emotion recognition using Artificial Intelligence is an approach that seeks to overcome these problems [57].

Ceccacci et al. [57] studied a way to automatically obtain satisfaction from users watching an opera. They invited people who agreed to participate in the experiment and positioned them in previously defined chairs as the target of cameras to collect images of these users' expressions during the presentation [57]. The images captured by the camera were processed by the computer, which assessed the person's gender, age, and emotion. One camera could capture data from up to 15 people at the same time. Ekman's six basic emotions [52] were evaluated, and their intensity was also measured. The predominant emotion and its valence were recorded [57].

In total, 132 people participated in the experiment, and it was possible to create a map of emotions per moment of the show. From the traditional satisfaction assessment work carried out with the participants, it was found that the level of emotional valence measured during the presentation contributed to determining the user's general level of satisfaction [57].

Agrawal et al. [58] studied a real-time solution for capturing user preferences in a fashion retail store. A recognition model of 8 emotions was used: neutral, anger, contempt, disgust, fear, happiness, sadness, and surprise [58]. The objective was to understand the public's reaction when passing in front of a window with mannequins on display. Recognition was carried out in two stages. One is to identify people's interest in the mannequin, and the second is to recognize the emotion presented. The accuracy obtained in recognizing emotions was around 92%. It was also possible to get the user's gender with an accuracy of 81% [58].

Rai Jain et al. [59] prepared a survey of various techniques for recognizing emotions through visual means, with their main characteristics and a list of references that use these techniques [59]. The study showed gaps in emotion recognition to detect mental health conditions. Furthermore, he pointed out that further studies are needed to identify the computational requirements of real-time emotion recognition [59].

Dampage et al. [61] proposed a study to increase user satisfaction in a restaurant using augmented reality and emotion recognition. The menu and dishes were designed three-dimensionally for users without special equipment. User satisfaction was measured using facial expression monitoring [61]. After presenting the 3D model of the dish, the designed system took photos of the participant every 1.5 seconds and sent them to the service server *Amazon Recognition.* The system returned a file that reported the analyzed emotions and confidence levels. The emotion with the highest level of confidence was recorded. The information was stored in a local bank to make it possible to associate the level of satisfaction with the dish and the time of year. The result showed that 87.5% of respondents were very satisfied with the three-dimensional experience, and 50% were very satisfied with the automatic satisfaction measurement [61].

Lee & Shin [62] proposed a TV system that identified the user's emotion and presented programs to be watched according to the detected emotion. The system worked gradually. When an emotion of happiness or sadness was identified, the contents migrated to funny or dramatic. Each user's preferred content was also obtained from the results on emotions [62]. The hypotheses were confirmed that "the *interface* for recognizing emotions from facial expressions on *smart TV* is the *interface* most interactive of all *interfaces*" and that "the individually optimized *interface* based on emotion is the most effective *interface* of all *interfaces*" [62].

A summary of the selected studies that focused on recognizing emotions visually using Artificial Intelligence is presented in Table 4.9.

ID	Title	Summary
S1 [12]	A survey of multi-modal sentiment analysis	The study identified recent developments in multi-modal sentiment analysis across different domains, including through emotion recognition in a visual form. The authors distinguished feeling, opinion, emotion, and affection. It was shown that the multimodal approach is more effective than the uni-modal one because it better uses complementary content between channels. Challenges in analyzing feelings and recognizing emotions were listed, such as the limitation to external manifestations and human cultural aspects, which can make a person express emotion differently from what they actually feel.
S2 [10]	Augmented reality aided analysis of customer satisfaction based on taste-induced facial expression recognition using Affdex Software Developer's Kit	The study proposed a solution for measuring restaurant customer satisfaction through emotion recognition. The authors identified anger, disgust, fear, joy, sadness, and surprise. An association was verified between user satisfaction and the taste of the food dishes tested, which was the study's objective. The ease of use of the system and delivery of results in real-time were considered important results of the study.
S3 [11]	Behavioral analysis for user satisfaction	The study proposed using an interactive movie selection system based on emotion recognition and behavioral analysis. Initially, a first interaction allowed mapping of the user's emotions and behaviors. Next, the system was presented with the customizations resulting from the mapping carried out in the previous phase. In a complementary way, the personalized system used the mapping of the user's gaze, the mouse's position on a screen item, and the user's clicks.

S4 [57]	Emotional valence from facial expression as an experience audit tool: an empirical study in the context of opera performance	The paper aimed to find a solution to the need for audience development in the artistic field using automatic emotion recognition in a spontaneous environment. It used facial recognition and gaze mapping technologies based on images obtained from participants using cameras in the environment. The system made it possible to identify Ekman's six basic emotions [52]. The technology captured the person's image, identified gender, age, and emotion, and then deleted the image without associating the extracted data with the person's name or identification.
S5 [58]	Gauging customer interest using skeleton tracking and convolutional neural network	The study aimed to perceive the public's reaction that passed in front of a fashion showcase with mannequins displayed. Recognition was carried out in two stages, one to confirm people's interest in the mannequin and another to identify emotions. The accuracy obtained in recognizing emotions was around 92%. It was also possible to obtain the client's gender with an accuracy of 81%.
S6 [59]	Recent trends in artificial intelligence for emotion detection using facial image analysis	The paper aimed to explore trends in emotion recognition from facial expressions. The study presented a large summary table with several visual emotion recognition techniques and a summary of the accuracy of techniques that used the Fer2013 dataset. The study concluded that models of simultaneous prediction of emotions and valence associated with intensity performed better than those that used just one or the other. The study also noted gaps in emotion recognition to detect mental health conditions.

S7 [60]	Retail managers' preparedness to capture customers ' emotions: a new synergistic framework to exploit unstructured data with new analytics	The study's objective was to understand retailers' demands for new tools that capture customer emotion. Customer behavior was found to be significantly influenced by affective experiences. Two studies related to emotion recognition were proposed—the first to create a machine capable of classifying generic images of people in a specific retail environment. The second study conducted interviews to understand how an emotion classifier system can add value to retail strategies and practices. It was found that this system could contribute to various retail store operations.
S8 [61]	Spatial augmented reality based customer satisfaction enhancement and monitoring system	The study's objective was to increase customer satisfaction in a restaurant using augmented reality and emotion recognition. The menu and dishes were designed three-dimensionally for customers without special equipment. Customer satisfaction was measured using facial expression monitoring. The result showed that 87.5% of respondents were very satisfied with the three-dimensional experience, and 50% were very satisfied with the automatic satisfaction measurement.
S9 [62]	The relationship between humans and based smart TVs on emotion recognition in HCI	The study aimed to compare interface types for a TV system. One of the *interfaces* identified the user's emotion and presented programs to watch according to the detected emotion. The system worked gradually. It has been observed that emotion recognition from facial expressions on *smart TV* is the most interactive*interface* and that the individually optimized *interface* based on emotion is the *interface* more efficient.

S10 [8]	Towards emotion acquisition in IT usability evaluation context	The study's objective was to evaluate several emotion recognition techniques in usability tests. The experiment was conducted in a room with a group of people who conducted individual usability testing. People could talk and interact with each other. Various emotion recognition techniques, such as facial expressions and voice, were applied to test the noise level and impact of environmental inadequacies.

Table 4.9: Summary of systematic literature mapping papers.

RQ.2. What are the visual emotion recognition techniques used in the literature related to automated satisfaction measurement using Artificial Intelligence?

Several techniques were identified in the literature, some of which were developed from data science algorithms modeled and trained for the work, and others used solutions already developed by third parties and pre-trained.

Ray Jain et al. [59] conducted a broad survey of emotion recognition techniques. The techniques identified with the most references in the literature were [59]:

- convolutional neural networks (CNN): deep learning technique for classifying images and videos. Uses *Multilayer algorithm Perceptron*(MLP) and requires relatively lower computational performance as it requires a smaller number of parameters;
- circumplex model of affect: is based on theories of emotional response to external stimuli. Emotions are represented using attributes of intensity and valence. They allow you to classify emotions, determine affective states, and interpret neuroimages;
- convolutional neural networks (DCNN): several convolutional neural networks are stacked to form a model. It uses more layers than a typical CNN and, therefore, guarantees a better fit than CNNs;
- support vector machines (SVM): supervised learning model based on classification and regression. Uses support vectors to separate data points into classes;
- VGG face: It is a DCNN with 22 layers and 37 units deep, trained with VGG data.

Furthermore, a set of data and systems aimed at recognizing facial expressions are still used to train the models [59]:

- AffectNet dataset: a collection of over 1 million candid images collected by search engines;
- CK/CK+ dataset: a set of 593 video sequences or 5876 posed images capable of distinguishing neutral expression from active expression;
- Facial Action Coding System (FACS): a system for classifying human facial expressions based on the facial and neck muscles involved. Each expression is formed by a set of action units, which represent parts of the expressions in a standardized way.
- FER 2013 dataset: set of around 30 thousand posed and candid images with different facial expressions labeled as anger, disgust, fear, happiness, sadness, surprise, or neutral.

Agrawal et al. [58] used a convolutional neural network (CNN) for emotion recognition. The network was based on the *ResNet 18* model, a convolutional network from *Microsoft* with 18 weighted layers. They also used the *Extended Cohn Kanade*, or CK+, dataset, which allowed cross-validation between action units (AU) of the *FACS*model and the image detected using *Active Appearance Models (AMM)* and a linear classifier *Support Vector Machine (SVM)* [58].

Chimienti et al. [11] used a deep convolutional neural network (DCNN), model *ResNet 50*, which was trained with 50 epochs with the FER2013 dataset. The network detected 86% accuracy based on this dataset.

Dampage et al. [61] used the service*AWS Amazon Rekognition*of emotion recognition. It uses *deep techniques learning*to analyze the images. Its support for facial expression analysis is continually increasing, and it has millions of images in its database, making it more accurate than locally trained models [61].

RQ.3. What are the benefits and challenges of performing emotion recognition visually?

Pantano et al. [60] proposed a study to understand retailers' demands for new tools that capture users' emotions. It was presented that understanding user experiences and engagement requires a deep insight into the emotions experienced by the user. It was found that affective experiences significantly influence user behavior. It was observed that users' hedonic responses can be evaluated as emotional reactions [60].

Two studies related to emotion recognition were proposed. The first to create a machine capable of classifying generic images of people in a specific retail environment. The classifier was trained with previously treated images to teach expressions referring to happiness and sadness. Validation of the classifier was more than 80% successful. After validation, the classifier was applied to photos voluntarily posted by users of a retail store [60].

The second study conducted interviews to understand how an emotion classifier system can add value to retail strategies and practices. It was found that this system could contribute to the following benefits [60]:

- space optimization: identifying places where people feel happiest can allow retail store managers to reorganize space occupancy;
- user profile: tracking users in retail stores can allow demographic mapping of users along with their emotions;
- purchasing behavior: it is possible to track a user during their purchases to measure their time spent in the store since entering the store, to allow managers to take actions so that they stay longer in the store and consume more ;
- promotions: identifying the emotional state of users in the store can allow managers to create promotional strategies for the times of the week when they are least satisfied;
- *user feedback*: collecting *feedback* from users about products or about the store installation itself is possible spontaneously with the use of emotion recognition;
- consumer relations: the use of emotions present in images on users' social networks can allow for improved communication between the retailer and the user;
- user privacy: concern for user privacy is critical in emotion recognition technologies. It must be at the center of developing any proposed operation using these technologies.

Dampage et al. [61] raised significant concerns about data privacy. With the use of the *Amazon service Rekognition*, they processed images and videos with the intention that the information, after processing, was eliminated.

Soleymani et al. [12] presented the challenge that, although extracting emotions from facial expressions is possible, the observed feelings may not correspond to people's real feelings. This is because people can act in a certain way to adapt to a cultural norm or even to express some issue of identity. Furthermore, people often do not express emotions, even if an event triggers them internally, and it is not possible to measure emotions that are not expressed [12].

Emotion recognition in the laboratory always requires human work to obtain and process the data. This demonstrates a limitation compared to the amount of data existing on the internet that could be the target of information extraction. Even in cases of data extraction by systems that scan the internet, there is the problem that the material produced and available is limited to a specific demographic context existing on the internet due to several factors, which has relevant ethical implications [12].

4.3.3 Threats to Results

Some issues that could impact the systematic mapping of literature were discussed and mitigated throughout the work:

- subjectivity in choosing the search keyword: Defining the search keyword was the first step towards creating the *string* search and was a determining factor in obtaining relevant studies. Given its importance, an inappropriate choice could compromise the entire work. To mitigate this problem, several synonyms of the keywords were selected, and the "or inclusive" (OR) function was used between the synonyms to create the *string* final search. It was possible to set up a search carried out between groups of keywords and synonyms, which allowed a broader search;
- inconsistency between planning definitions: the proposed systematic literature mapping sought to obtain information planned based on the research objective. There was a risk that the planning elements had been proposed out of step with the research objective. To mitigate this risk, we used the *Parsifal*[54] tool, which proposes a step-by-step guide for systematic mapping so that one planning stage is the input for the next stage. In addition, a consistency check was carried out on the alignment between the research planning stages;
- reproducibility of the work: transparency and the ability to reproduce a work are highly relevant factors for any study, as they allow anyone to verify the results. To maximize reproducibility, all steps followed in mapping were recorded in this study, from planning to results.

4.4 Chapter Summary

A systematic literature mapping (SLM) provides an overview of the research area and allows you to discover trends [53]. An SLM was carried out to map the state of the art related to the use of Artificial Intelligence tools and techniques related to emotion recognition, especially facial expression recognition, to identify techniques, benefits, and challenges associated with the development of a new tool that allows measuring the user satisfaction in an automated way.

Ten primary studies were identified in different areas of knowledge: restaurants, television systems, retail stores, artistic shows, and usability testing in a call center system. Two systematic literature reviews were also identified. The technique most used by primary studies is the convolutional neural network (CNN). The use of cloud services for emotion recognition was also verified. Benefits related to user *feedback* were reported, such as user profile mapping, and challenges were encountered for emotion recognition, such as user privacy and inadequacies of the capture environment.

Chapter 5

Proposed Model

After carrying out the systematic literature mapping, it was possible to obtain an overview for proposing the work to evaluate user experience by performing emotion recognition based on facial expressions and sentiment analysis. As mentioned in Section 1.4, this work aims to implement and validate a UX evaluation model through a tool that automatically determines a user's experience level when using a digital product and presents the positive, neutral, and negative points of using that product. Therefore, this chapter will describe the elements necessary to produce this solution.

The solution is based on Park et al. [4] studies and explores three elements of user experience: usability, affect, and user value. After identifying these three elements based on the emotional elements of the user's speech and visuals, the solution calculates the final UX level for the interaction with the user. The user can give personal feedback for each of these three elements. Thus, the application can compare the results and verify its assertivity.

5.1 Proposal Overview

The process enabled by the tool aims to conduct the user through the activities to generate the data necessary to evaluate emotional SUS, affect, and emotional satisfaction while the user interacts with a digital product. This can be done in several ways, such as through a guided usability test or without interfering with the user's usual work on this product. This approach uses a usability test with an open exploratory task and guidance for users to say what they feel or think (think-aloud protocol). The proposed model is shown in Figure 5.1

We carried out a first round of tests, and we observed that the first parameters defined to evaluate UX experience with a 5-point Likert [31] scale were not appropriate, as we show in Chapter 6. Then, we reduce the model to a 3-point scale.

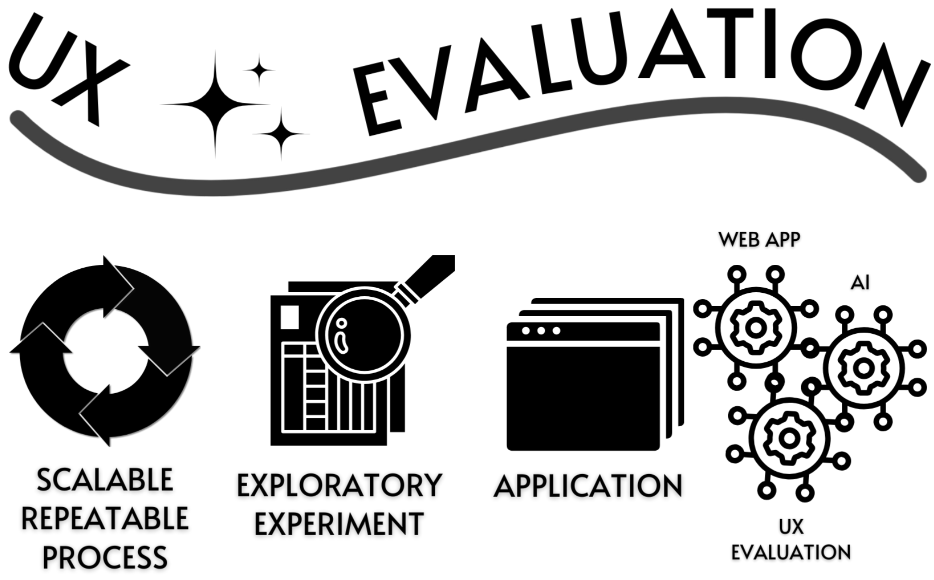

Figure 5.1: Proposed Model

5.1.1 Emotional Usability Evaluation

This procedure addresses the usability element of the UX experience. One of the most known tools for evaluating usability is the System Usability Scale (SUS), presented in Chapter 3. The user typically fills out the SUS questionnaire after using a digital product in a usability test. We created a way to evaluate SUS from the user's speech obtained due to the think-aloud protocol in the usability test. We analyze the speech by transcribing and separating the speech into sentences. Then, each sentence is evaluated by its similarity with SUS statements.

The evaluation generates a similarity score to indicate the proximity between one question from the SUS questionnaire and a sentence spoken by the user. A score of 0 means no similarity and a score of 100% means identical sentences. We test some configurations varying from 80% to 25%. Higher values increase the result's confidence but reduce the valid comparison results. Lower values result in more valid comparison results, but still with similarity. We perceived the size of sentences in a speech tends to be large, so it reduces the similarity with the given SUS sentences. In the first analysis, we define the score as 50%. After the first round of tests, we observed the sentences evaluated, allowing us to reduce that score to 25% and increase the number of analyzed sentences.

The users part of this experiment speak Portuguese as their native language. We obtained the Portuguese SUS sentences from Martins et al. [63], and we show it in

Portuguese in Table 5.1 and the English version in Table 5.2.

Portuguese SUS Sentences [63]	**Example of Synonyms Sentence**
1. Acho que gostaria de utilizar este produto com frequência.	Acredito que este produto seria útil para mim. Acho que este produto seria benéfico para mim. Acho que este produto seria produtivo para mim. Acho que este produto seria eficiente para mim. Acho que este produto seria agradável para mim. Acho que este produto seria conveniente para mim. Acho que este produto seria acessível para mim. Acho que este produto seria fácil de usar para mim. Acho que este produto seria flexível para mim. Acho que este produto seria personalizável para mim.
2. Considerei o produto mais complexo do que necessário.	Acho que o produto é difícil de entender. Acho que o produto é confuso. Acho que o produto é complicado. Acho que o produto é elaborado. Acho que o produto é intrincado.
3. Achei o produto fácil de utilizar.	Acho que o produto é intuitivo. Acho que o produto é simples. Acho que o produto é direto. Acho que o produto é claro. Acho que o produto é conciso.
4. Acho que necessitaria de ajuda de um técnico para conseguir utilizar este produto.	Acho que o produto é muito técnico. Acho que o produto é muito avançado. Acho que o produto é muito especializado. Acho que o produto é muito sofisticado. Acho que o produto é muito exigente.
5. Considerei que as várias funcionalidades deste produto estavam bem integradas.	Acho que o produto é coeso. Acho que o produto é harmonioso. Acho que o produto é consistente. Acho que o produto é integrado. Acho que o produto é sinérgico.
6. Achei que este produto tinha muitas inconsistências.	Acho que o produto é inconsistente. Acho que o produto é desorganizado. Acho que o produto é caótico. Acho que o produto é confuso. Acho que o produto é desordenado.
7. Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este produto.	Acho que o produto é fácil de aprender. Acho que o produto é intuitivo. Acho que o produto é simples. Acho que o produto é direto.

	Acho que o produto é claro.
8. Considerei o produto muito complicado de utilizar.	Acho que o produto é confuso. Acho que o produto é desorganizado. Acho que o produto é caótico. Acho que o produto é ineficiente. Acho que o produto é frustrante.
9. Senti-me muito confiante a utilizar este produto.	Acho que o produto é confiável. Acho que o produto é seguro. Acho que o produto é estável. Acho que o produto é robusto. Acho que o produto é durável.
10. Tive que aprender muito antes de conseguir lidar com este produto.	Acho que o produto é complexo. Acho que o produto é avançado. Acho que o produto é especializado. Acho que o produto é sofisticado. Acho que o produto é exigente.

Table 5.1: Portuguese sentences used to obtain baseline for SUS similarity score.

SUS Sentences	Example of Synonyms Sentence
1. I think that I would like to use this system frequently.	I believe this product would be useful for me. I think this product would be beneficial for me. I think this product would be productive for me. I think this product would be effective for me. I think this product would be enjoyable for me. I think this product would be convenient for me. I think this product would be affordable for me. I think this product would be easy to use for me. I think this product would be flexible for me. I think this product would be customizable for me.
2. I found the system unnecessarily complex.	I think the product is difficult to understand. I think the product is confusing. I think the product is complicated. I think the product is elaborate. I think the product is intricate.
3. I thought the system was easy to use.	I think the product is intuitive. I think the product is simple. I think the product is straightforward. I think the product is clear. I think the product is concise.
4. I think that I would need the support of a technical person to be able to use this system.	I think the product is very technical.

	I think the product is very advanced. I think the product is very specialized. I think the product is very sophisticated. I think the product is very demanding.
5. I found the various functions in this system were well-integrated.	I think the product is cohesive. I think the product is harmonious. I think the product is consistent. I think the product is integrated. I think the product is synergistic.
6. I thought there was too much inconsistency in this system.	I think the product is inconsistent. I think the product is disorganized. I think the product is chaotic. I think the product is confusing. I think the product is cluttered.
7. I would imagine that most people would learn to use this system very quickly.	I think the product is easy to learn. I think the product is intuitive. I think the product is simple. I think the product is straightforward. I think the product is clear.
8. I found the system very cumbersome to use.	I think the product is confusing. I think the product is disorganized. I think the product is chaotic. I think the product is inefficient. I think the product is frustrating.
9. I felt very confident using the system.	I think the product is reliable. I think the product is safe. I think the product is stable. I think the product is robust. I think the product is durable.
10. I needed to learn a lot of things before I could get going with this system.	I think the product is complex. I think the product is advanced. I think the product is specialized. I think the product is sophisticated. I think the product is demanding.

Table 5.2: English sentences used to obtain a baseline for SUS similarity score

As we showed before, the SUS questionnaire uses a five-point Likert scale [31]. The emotional SUS questionnaire is initiated with all questions set to the "Neither agree nor disagree" option, which means that if no user's sentence is similar to a SUS questionnaire question, that will be the considered value to calculate emotional SUS. Calculating SUS with these default values means we obtain 50 points, a neutral assessment.

To evaluate the emotional SUS, we must perform a similarity analysis between each SUS sentence and all the sentences extracted from the user's speech. This analysis results in a percentage that represents the similarity score. To allow better use of this questionnaire, in the first analysis, we considered the scores equal or greater than 70% as a "strongly agree" or "strongly disagree" with the SUS question and values between 50% (inclusive) and 70% (exclusive) as an "agree" or a "disagree" with the question. If the similarity is less than 50%, then the sentence is considered "Neither agree nor disagree" as a default value. After the first test round, we change these values to 25%, 75%, and 25%, as presented in Table 5.4.

After evaluating similarity, we need to determine the direction of similarity. As we observed in the SUS questionnaire presented in Tables 5.1 and 5.2, the sentences 1, 3, 5, 7, and 9 are positive, which means the greater its value, the more positive, and the sentences 2, 4, 6, 8, and 10 are negatives.

Thus, it's necessary to identify whether the user sentence is positive, neutral, or negative and whether the SUS sentence is positive, neutral, or negative. In the initial definition, if the effect matches both sentences, the correspondent value "strongly agree" or "agree" was used. If the effect doesn't match both sentences, the correspondent value "strongly disagree" or "disagree" was used. Neutral sentences were considered as "neither agree nor disagree". We presented the conversion schema in Table 5.3. After the first round of tests, we adjust these parameters as Table 5.4.

Similarity Score	**SUS Sentiment - User Sentiment**	**Likert Scale Equivalence**
>=70%	Positive - Positive or Negative - Negative	Strongly agree
>=50% & <70%	Positive - Positive or Negative - Negative	Agree
>=50%	Any - Neutral	Neither agree nor disagree
<50%	Any - Any	Neither agree nor disagree
>=50% & <70%	Positive - Negative or Negative - Positive	Disagree
>=70%	Positive - Negative or Negative - Positive	Strongly disagree

Table 5.3: Initial similarity score and sentiment analysis' valence conversion to Likert scale.

.

After identifying the emotional SUS points, it's necessary to transform this value into a qualitative scale. As we observed on a 1-100 scale, the neutral value ranges from 25 points in the sentiment analysis, as shown in table 5.20. Thus, we proposed a 30-point range to neutral value in the Emotional SUS Score. The initial conversion approach is shown in Table 5.5 and the final in Table 5.6.

Similarity Score	SUS Sentiment - User Sentiment	Likert Scale Equivalence
>=75%	Positive - Positive or Negative - Negative	Strongly agree
>=25% & <75%	Positive - Positive or Negative - Negative	Agree
>=25%	Any - Neutral	Neither agree nor disagree
<25%	Any - Any	Neither agree nor disagree
>=25% & <75%	Positive - Negative or Negative - Positive	Disagree
>=25%	Positive - Negative or Negative - Positive	Strongly disagree

Table 5.4: Final similarity score and sentiment analysis' valence conversion to Likert scale.

.

Emotional SUS Score	Emotional Usability Level
>=85	Very good
>=65 and <85	Good
>=35 and <65	Neutral
>=15 and <35	Poor
<15	Very poor

Table 5.5: Initial emotional usability scale.

Emotional SUS Score	Emotional Usability Level
>=65	Good
>=35 and <65	Neutral
<35	Poor

Table 5.6: Final emotional usability scale.

5.1.2 Affect Evaluation

This procedure addresses the affective element of the UX experience. The effect evaluation identifies the valence of the user's sentiment at a determinate time and shows the user's emotion if possible. It comprises emotion recognition originating from the facial expression mechanism and speech analysis. Whenever the application finds an emotion or sentiment, it is shown to the user as a time lime to identify its valence, intensity, and description, if available.

The application can identify the emotions of joy, sadness, surprise, and anger from facial expression analysis. Each has a valence to specify whether the emotion is positive, neutral, or negative. This mechanism is presented in Subsection 5.1.3. The application can also identify the affect of the user's speech analysis. It is done from the textual sentiment analysis of the user's speech, as we showed in Subsection 5.1.1.

The affect is shown as a positive, neutral, or negative sentiment. The application shows the instant affect distinguished into instant speech affect, instant video affect, and instant emotion. The positive and negative evaluation peaks are the highest and lowest instantaneous affect identified. The evaluation peaks consider both speech affect and video instant affect. The application uses the sentiment magnitude of speech affect to determine the peaks and the instant satisfaction plus the emotion duration to calculate the video affect peaks.

At the end of the experiment, the overall affect is evaluated. The formula for calculating it is the sum of the number of positive or negative evaluations divided by the total number of affect evaluations. The most frequent affect determines the final result. If more than 65% of the assessments have positive or negative valence, then the intensity is considered "positive" or "negative". Else, it is considered just "neutral". The final evaluation is presented in Table 5.7.

Affect's Frequency	**Affect's Valence**	**Final Affect**
>= 65%	Positive	Positive
>0 and <65%	Any	Neutral
>= 65%	Negative	Negative

Table 5.7: Final affect evaluation.

5.1.3 Emotional Satisfaction Evaluation

This procedure addresses the value element of the UX experience. Emotions, intensities, and valences determine the overall user satisfaction. Peaks of positive and negative emotions are also identified. This is done due to the identification by Kujala et al. [27] that user experience is evaluated as an overall evaluation of the experience, not a sum of each moment. Furthermore, this study identified that the final moments and evaluation peaks are decisive for reconstructing memory retrospectively and, consequently, for evaluating the experience.

Valence is an attribute that associates each emotion with positive, neutral, or negative semantics. It is possible to see valence as a sign of the intensity of the emotion. The value associated with the emotion's intensity will be positive for positive emotions. For negative emotions, the value related to the intensity of the emotion will be negative. This is done by transforming the valence into a +1 or -1 multiplier, as appropriate. Neutral emotions will have valence with a multiplier of 0.

Balbin et al. [10] presented a valence table in their work in which joy was considered positive, anger, disgust, and fear were deemed negative, and surprise was considered neutral. In this case, the valence of the second-highest emotion detected was considered. This work will use the same approach for joy, sadness, and anger. Still, the surprise will only be considered neutral without using the second emotion to identify the valence, according to Table 5.8.

Positive Emotions	**Negative Emotions**	**Neutral**
Joy	Sadness Anger	Surprise

Table 5.8: Valence of emotions used.

The AI component returns the following intensities for each emotion: *Unknown*, *Very_Unlikely*, *Unlikely*, *Possible*, *Likely* or *Very_Likely*. For calculation purposes, we need to transform qualitative intensity into quantitative. We initially associate a score for each intensity as shown in Table 5.9. After the first round of tests, we adjusted it to Table 5.10. The valence of satisfaction associated with the score varies according to Table 5.8. The initial defined level of satisfaction related to the score varied according to Table 5.12, and the final one varied according to Table 5.13.

The AI component also returns a general degree of confidence for the entire result, a percentage value. It will be used as a criterion for the quality of the result. As we did about the similarity score in Section 5.1.1, we consider 0.5 a quality score, so we only use evaluations greater than or equal to 0.5. In this case, if the confidence level is less than 50%, the identified emotion will be disregarded, and we will associate a confidence score of 0. If the degree of confidence is greater than 50%, it will be associated with a confidence score of 1. The confidence score is presented in Table 5.11.

Intensity	Intensity Score
Very_Likely	2
Likely	2
Possible	1
Unlikely	1
Very_Unlikely	0
Unknown	0

Table 5.9: Initial association of intensities with intensity score.

Intensity	Intensity Score
Very_Likely	1
Likely	0
Possible	0
Unlikely	0
Very_Unlikely	0
Unknown	0

Table 5.10: Final association of intensities with intensity score.

Confidence Score	Confidence Score
>=50%	1
<50%	0

Table 5.11: Association of confidence degree with confidence score.

We will measure satisfaction in a score calculation based on the intensity and valence of the emotion, as well as the degree of confidence in the result received. The instant satisfaction score will be calculated as the intensity score multiplied by the valence of the emotion and the instant confidence score. The formula that describes the calculation of instant satisfaction is:

$$IntensityScore * Valence * ConfidenceScore$$

The overall satisfaction score calculation will add the non-zero instant scores identified and divide by the total number of non-zero identifications. The final score will be rounded to the next integer if the fractional value obtained has the first decimal place greater than or equal to 5; otherwise, it will be rounded to its integer part. The formula that describes the calculation of the overall satisfaction score is:

$$round(\frac{\sum(NonZeroInstantSatisfactionScore)}{NonZeroInstantTotal})$$

As we initially defined, the overall satisfaction score was always one of the discrete values -2, -1, 0, 1, and 2. We associated the qualitative satisfaction level with each of these scores, respectively: "very dissatisfied", "dissatisfied", "neither satisfied nor dissatisfied", "satisfied" and "very satisfied". To facilitate the presentation and visualization of satisfaction, as well as standardize manual and automatic evaluation to an acceptable degree of precision, it was used a satisfaction score table, which varies from -2 to 2. It is shown in Table 5.12.

After the first round of tests, we refined the quality satisfaction levels and the satisfaction score table as Table 5.13.

Manual Alternative	Final Level	Corresponding Satisfaction
Strongly agree	+2	Very satisfied
Agree	+1	Satisfied
Neither agree nor disagree	0	Neither satisfied nor dissatisfied
Disagree	-1	Dissatisfied
Strongly disagree	-2	Very dissatisfied

Table 5.12: Initial manual review and final level conversion to satisfaction score.

Manual Alternative	Final Level	Corresponding Satisfaction
Agree	+1	Satisfied
Neither agree nor disagree	0	Neither satisfied nor dissatisfied
Disagree	-1	Dissatisfied

Table 5.13: Final manual review and final level conversion to satisfaction score.

5.1.4 User Experience Evaluation

The User Experience Evaluation is a qualitative index to translate the association of its elements evaluations. This evaluation considers the final index of each UX element. The application sums the given points independently of the other elements for each UX element evaluation. Initially, we proposed a UX element's score as shown in Table 5.14. After the first round of tests, we adjust the values as the Table 5.15. If, e.g., a task is evaluated as "Good", "Positive", and "Neither satisfied nor dissatisfied" for each element of Table 5.14, it receives 1+1+0 points, resulting in 2 points for the user experience score.

Emotional Usability	**Affect**	**Emotional Satisfaction**	**Points**
Very good	Not applicable	Very satisfied	+2
Good	Positive	Satisfied	+1
Neutral	Neutral	Neither satisfied nor dissatisfied	0
Poor	Negative	Dissatisfied	-1
Very poor	Not applicable	Very unsatisfied	-2

Table 5.14: First UX evaluation points.

Emotional Usability	**Affect**	**Emotional Satisfaction**	**Points**
Good	Positive	Satisfied	+1
Neutral	Neutral	Neither satisfied nor dissatisfied	0
Poor	Negative	Dissatisfied	-1

Table 5.15: Final UX evaluation points.

Finally, the final UX evaluation is based on the UX evaluation points. Each element contributes its points to the final score. The initial conversion of the final score to the UX Evaluation Scale is shown in Table 5.16. After the first round of tests, we adjusted as Table 5.17. In our previous example, we can evaluate the user experience score with a value of 2 as a "Good" user experience.

UX Final Score	**UX Evaluation Scale**
+4, +5	Very good
+2 or +3	Good
-1, 0, +1	Neither good nor poor
-2 or -3	Poor
-4, -5	Very poor

Table 5.16: Initial UX evaluation scale.

UX Final Score	UX Evaluation Scale
+2 or +3	Good
-1, 0, +1	Neither good nor poor
-2 or -3	Poor

Table 5.17: Final UX evaluation scale.

5.1.5 Proposed Model Activities

We present the proposed process for studying this model in Figure 5.2. The activities proposed in the work process are described below. Some activities can be carried out manually, and others automatically.

- The user accesses Google Meet: the user joins a remote meeting to interact with the test team or creates one on their own;
- The user accesses UXAPP: the user starts the UXAPP to register important data about the test;
- The user identifies themself: we identify the user to ensure that the person who accepted the terms is the same person who provided the data for processing;
- The user reads and accepts the service's terms of use: by accepting the terms of use, the user demonstrates that they understand which data will be used and in what way;
- The terms of use are stored: the storage of terms of use aims to ensure that data use authorizations are always available;
- The user informs experiment data: we should verify usability against a determined objective, check the user's initial emotional state, and verify whether the product is hedonic or instrumental.
- The user reads the orientations about the think-aloud protocol: this protocol is essential to identify Emotional SUS and helps to obtain facial expressions.
- The user accesses the digital product to be tested: the user opens the product that will be tested and makes it ready for use;
- The user shares screen and records the meet: the user indicates the screen where the product is located and starts recording the product screen and the computer camera;
- The user uses the product following think-aloud protocol: the user starts using the product according to the type of evaluation to be carried out;

- The user informs and registers the task's end: the user registers when the work on the product is finished;
- The user explains what they think and feel about the product: the user speaks what they felt using the product. It helps the application understand the user's affect and emotions;
- The user stops recording and screen sharing: the user stops screen sharing and ends screen and camera recording;
- The user informs whether they achieved the task goal: the user registers if the goal was achieved;
- The user answers the SUS questionnaire: this is the manual usability evaluation to compare with the automatic one;
- The user informs your emotional state: this is the manual affect evaluation to compare with the automatic one;
- The user informs the negative and positive peaks of satisfaction: This information is used to complement peaks of satisfaction automatically calculated.
- The user answers satisfaction level: this is the manual value evaluation to compare with the automatic one;
- UXAPP evaluates emotional usability: the application calculates Emotional Usability;
- UXAPP evaluates user's emotional state (affect): the application calculates Affect;
- UXAPP evaluates user's emotional satisfaction (user value): the application calculates Emotional Satisfaction;
- UXAPP evaluates digital product's User Experience: the application calculates UX level;
- UXAPP presents a video with emotional evaluation over time: the application shows the instant affect, emotion, and satisfaction in the video over time;
- UXAPP presents a report with manual and emotional evaluation: application creates a final report.

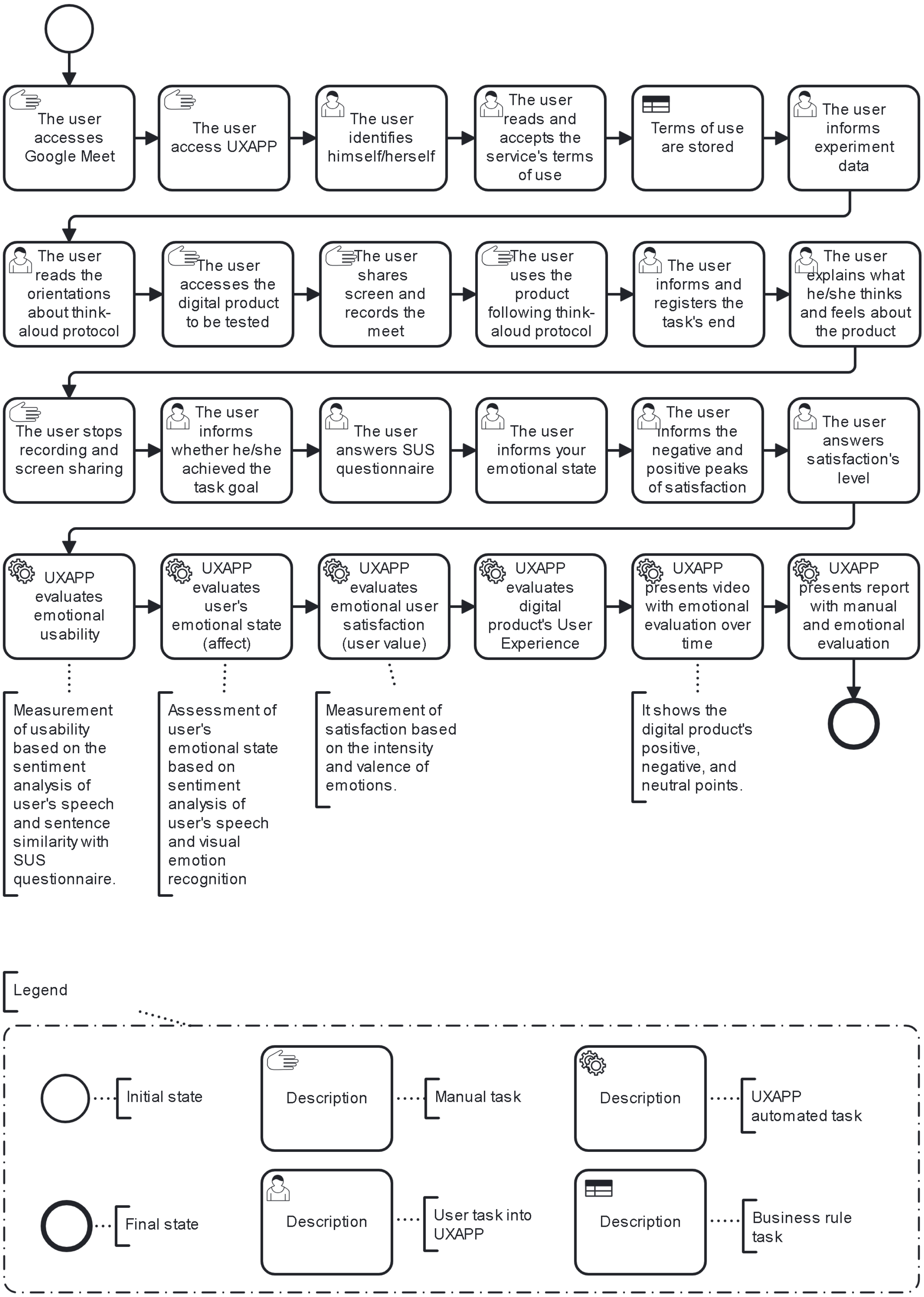

Figure 5.2: Working process of the proposed model.

5.1.6 Exploratory Experiment

Experimenting to test whether the automatic satisfaction level is in line with the manual satisfaction level must consider a minimum number of participants. Borsci et al. [64] studied the hypothesis of using five users in usability tests to be sufficient for the success of the evaluation and concluded that, although five users are enough in certain conditions, the use of ten users allows reaching more than 90% in estimated discovery probability for metrics: return on investment (ROI),*Good -Turing* (PGT-Norm), Monte Carlo (PMC),*Bootstrap Discovery Behavior* (PBDB) and average*P- Value* (PM). In this model, we suggest the use of 5 to 10 participants [64].

The restrictions for carrying out the test are presented in Table 5.18. We illustrated the experiment in Figure 5.3.

ID	**Criteria**	**Description**
R1	Procedure	It will be carried out remotely via teleconference.
R2	Procedure	The user is asked to say what they feel and do (think-aloud protocol).
R3	Procedure	The user is asked to inform their objective and whether they achieved it.
R4	Procedure	5 to 10 users to be recruited.
R5	Time	It should last between 15 and 60 minutes.
R6	Cost	It's free of expenses.

Table 5.18: Constraints for product usability testing.

5.2 Application

The application will be separated into three parts: *frontend*, *backend*, and the AI component. Details are presented in the following sections.

5.2.1 Architecture

The application that will test the digital product uses, as frontend, the Google Meet tool, as it already provides the necessary functionalities to capture the camera and screen when using the product, and an app created with Google Appsheet to collect test answers. The recorded video will be stored in Google Drive and made available for reading by the backend.

Next, as a backend, an API in Python stored and callable in Google Cloud Function will be used to process the video. Google Vision API performs emotion recognition and will return the processed information using a file in JSON format. We use Google Speech

Figure 5.3: Exploratory experiment illustration.

API to transcribe the speech and Google Natural Language API to identify the sentiment of the user's speech. An algorithm in Python to perform similarity analysis is also used.

Speech transcription can contain elements that can make it challenging to analyze feelings, such as repeated words, onomatopoeia, or lack of punctuation. With this, the transcription is submitted to the Gemini API so that the large language model (LLM) can generate a new transcription with the text separated into sentences and with the adjusted punctuation.

Then, the API creates the final video file and makes it available again on Google Drive to be accessed and used for analysis. All sending and receiving of information will be done using the HTTPS protocol, and access to all these solutions is carried out using just one internet browser, such as Google Chrome. Table 5.19 presents the list of technologies used. The architecture of the proposed model is shown in Figure 5.4.

We seek to propose a model that works in an integrated way, simple enough to be replicated in any organization. Despite being proprietary, the tools listed have free versions that allow anyone worldwide to implement the presented model at no cost. We believe other tools could be used instead of those shown, but this would need testing.

Google Vertex AI was chosen because it is a widely used environment for building computer vision applications and collecting insights from images and videos with APIs, AutoML, or pre-trained models. Using a validated environment proved more assertive than developing a new model with limited data resources about the amount of data

Technology	**URL**	**Description**
Google Appsheet	appsheet.com	Tool to create apps without coding.
Google Colab	colab.research.google.com	Tool to write and run python scripts without configuration, with GPU allocation and cloud storage.
Google Drive	drive.google.com	Tool for storing and editing files in the cloud with features for sharing files.
Google Gemini API	ai.google.dev	Gemini is a family of generative AI models developed by Google DeepMind that is designed for multimodal use cases.
Google Meet	meet.google.com	Tool for holding video conferences securely, with video recording, reduction features of noise and screen sharing.
Google Natural Language API	cloud.google.com/natural-language	An API for processing unstructured text using machine learning.
Google Speech to Text API	cloud.google.com/speech-to-text	An API for accurately convert speech into text based on Google's AI research and technology.
Google Vision API	cloud.google.com/vision	Environment for creating computer vision applications and collecting *insights* from images and videos with APIs, AutoML or pre-trained models .
Json	Not applicable	File format for fast and simple data exchange between systems.
Python	python.org	High-level programming language that allows the reuse of numerous libraries.
spaCy	spacy.io	A library to natural language processing with python. It has several trained pipelines for 25 languages, including Portuguese and English.

Table 5.19: List of technologies used in the proposed model.

used in cloud solutions. Furthermore, Choi et al. [65] presented that the Google Vertex AI AutoML model, in image prediction, "presented relatively favorable results, with an accuracy of 89.9%, precision of 94.2%, recall of 88.4 %, F1 score of 91.2%, and a log loss of 0.268".

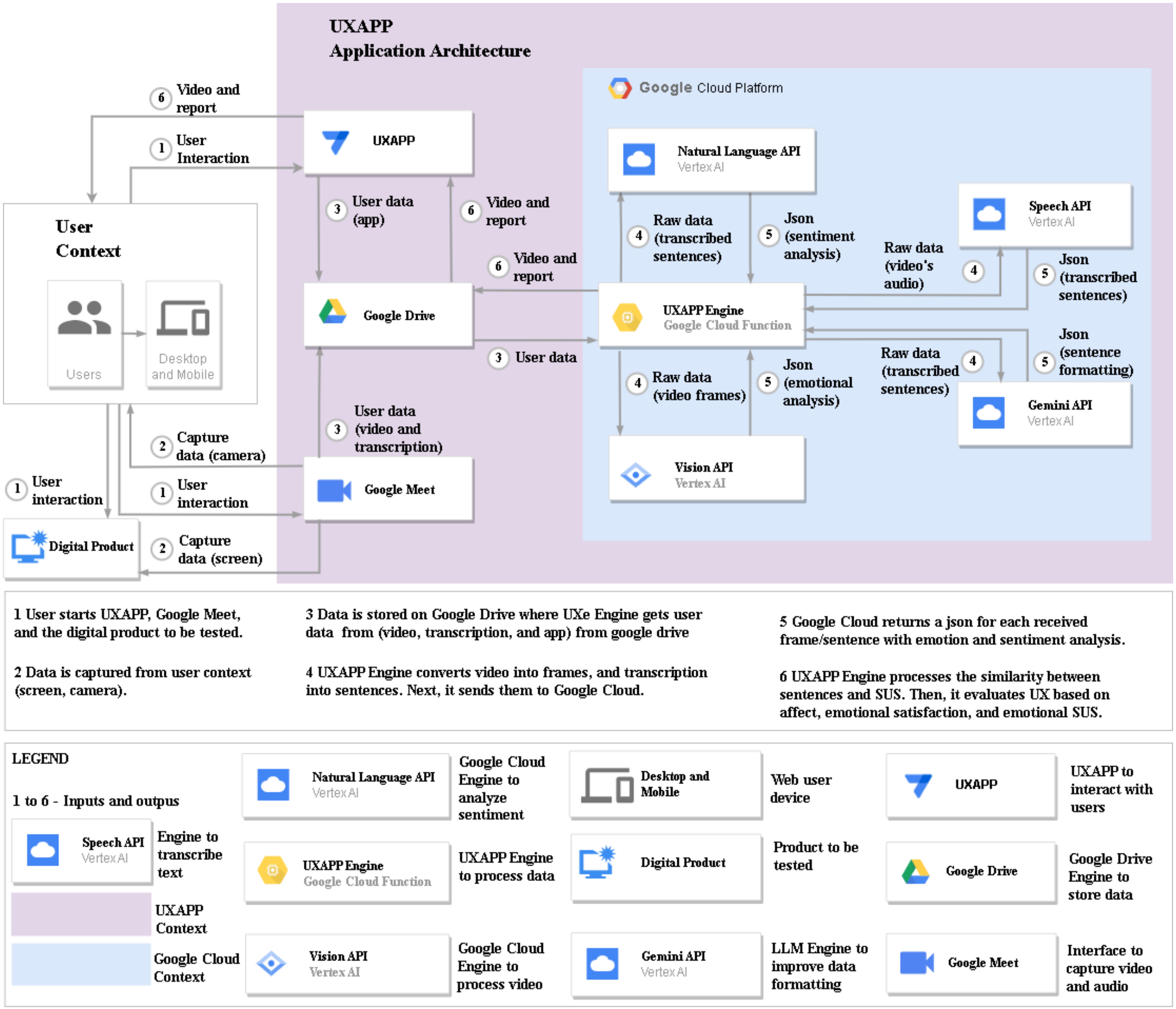

Figure 5.4: Proposed model architecture.

5.2.2 Overview

The description of each of the application components, *frontend*, *backend* and *AI component*, *frontend*, and their responsibilities are presented below.

UXAPP (Frontend)

This component is the main contact point with the user and is responsible for conducting the evaluation process with the user. The UXAPP allows the users to give their identification, presents and stores the terms of use, and collects the users' answers necessary for the UX evaluation and the manual evaluation verification.

This app was developed in the Google APPSHEET platform, which allows us to focus on the interactive points with the user. No code is necessary. All the information gathered from the user is stored in a Google Drive spreadsheet to facilitate the data exchange with the UXAPP Engine.

To maximize the user's success with the utilization of this APP, it was planned to be a step-by-step interactive approach. So, the user must finish the present activity to continue to the next one. It is also essential to give test instructions for them to use the app without the need for interaction with a usability test team.

Google Meet (Frontend)

This component is responsible for interacting with the user and capturing facial expressions and the user's speech, as shown in Figure 5.4. It also needs to capture the screen of the digital product as the user uses it. The two captures are taken simultaneously in an ideal scenario and generate a single video. This is because it is necessary to present the user's expression and the screen item the user is using.

Google Meet allows the user to control the beginning and end of the recording and select the screen on which the digital product will be manipulated. At the end of the recording, the video file is stored in Google Drive in a way that is accessible to other components.

Google Drive (Backend)

This component is responsible for storing all data received by the users and the one sent to them. Google Drive is widely known and can be integrated with several other tools, facilitating data control.

UXAPP Engine (Backend)

This component is responsible for managing the application's data flow and carrying out the necessary treatments so that the other components receive the data ready for use, as shown in Figure 5.4. Besides, this component carries out the similarity analysis between the SUS questionnaire questions and the sentences extracted from the user's speech.

It is first necessary to obtain the video file generated by the previous component, which is stored in the cloud. After loading, the video will be converted to frames. This is because applying facial recognition requires the receipt of individual frames to identify people and expressions. Each second of the video can contain dozens of frames. The 24 frames per second rate is very common, allowing for greater video fluidity. This, however, can generate an overload of processing and data transmission, which would be detrimental to the application.

Dampage et al. [61] had a similar need when taking photos of customers' facial expressions at a restaurant that presented the menu with augmented reality. Image processing was done by the service Amazon Rekognition, and pictures were taken and sent every 1.5 seconds. In this way, extracting and sending frames at the same rate may be sufficient to map the user's expression and not overload the application in cost or processing.

After sending a frame, the AI component returns a file in the format JSON, which presents the identified information. Among them, possible emotions are shown. The overall confidence level of the result is also given. It is possible to reassemble the video and add the emotion label and valence for each evaluated frame using this information.

After processing the satisfaction information, the application will insert this information into the final report to be made available to the end user in the UXAPP.

Regarding the sentiment analysis, this component sends audio to Google Speech API and receives the video's transcription. Then, it sends the text to Google Natural Language AI, as shown in Section 5.2.2. After treating the text, this component returns a JSON with the sentences and their analysis.

We use textual sentiment analysis to present the affect element and generate the video subtitles. Then, we send the transcription to Google Gemini API to organize the speech into sentences with punctuation and capitulation. The structured sentences are submitted to a similarity analysis with the SUS sentences.

This procedure compares each submitted sentence with all sentences in the SUS questionnaire. The analysis evaluates the similarity score as presented in Section 5.1.1. The base similarity algorithm is shown in Listing 5.1.

```
#Colab Script

#!python -m spacy download pt_core_news_lg
!python -m spacy download en_core_web_lg

import spacy

# Load the language model in English or Portuguese
#nlp = spacy.load("pt_core_news_lg")
nlp = spacy.load("en_core_web_lg")

#We set a SUS Sentence.
text1 = "I think that I would like to use this system frequently"
```

```

#We set a user sentence with a positive or negative sentiment analysis.
text2 = "Those with my email may be bad for privacy reasons, but I liked it because it
    makes it easier to fill out"

# Text processing
doc1 = nlp(text1)
doc2 = nlp(text2)

# Similarity calculation
similarity = doc1.similarity(doc2)

# Show results.
#print(f"A similaridade entre os textos é: {similarity}")
print(f"The similarity between the texts is: {similarity}")

#This process is carried out with each SUS sentence for all user sentences with
   positive or negative sentiment analysis.
#The similarity between the texts is 0.9085682970079145
```

Listing 5.1: Base Algorithm for Similarity Analysis

Vision API (AI Component)

The Google Vision API (Vertex AI) processes the frames of video received and returns a response in the format JSON with the identified emotions, as shown in Figure 5.4. No information is stored in this processing, so concerns about the user's privacy can be directed to the moments before or after processing.

The emotions detected are joy, sadness, anger, and surprise. For each emotion, the API returns its intensity. It also presents the overall confidence level of the result. This API lets you identify image quality information, such as underexposure or image blur issues. This *Google* service also returns security notes that allow you to determine whether the video file sent contains inappropriate content, such as adult, false, medical, violent, or obscene content. It is essential to make this available to the end user and identify whether the user made inappropriate use of the application and violated the terms of use.

Natural Language API (AI Component)

The component Google Natural Language AI detects text sentiment analysis from the user's speech. This component receives a text and returns a JSON file with the text's general sentiment and the sentiment analysis by the text's sentence. It also produces a score and a magnitude index. This score is within the interval -1.00 and 1.00, indicating the positive, neutral, or negative sentence's intention as Table 5.20 shows. Besides, this API returns the magnitude of the sentiment on a scale of 0.00 to 1.00. As we did about the similarity score in Section 5.1.1, in the first analysis, we consider 0.5 a quality score, so we only use evaluations with a magnitude greater than or equal to 0.5. After the first round of tests, we reduce this score to 0.25.

Natural Language AI Score	**Sentiment Analysis**
0.25 to 1.00	Positive
-0.25 to 0.25	Neutral
-1.00 to -0.25	Negative

Table 5.20: Sentiment Analysis Scores

Speech to Text API (AI Component)

The Speech to Text API converts an audio into a text document. This API can be configured to return sentences with time offsets and punctuation. This was used to generate the subtitle file integrated with the video and available to download through UXAPP. Each person's speech should be recorded in one separate channel in order to generate a better transcription.

The audio file is uploaded to a bucket in Google Cloud Storage, processed, and the results are returned or stored in another bucket. Google Speech to Text uses the "Chirp" model to transcribe the audio from several languages, including English and Portuguese.

Gemini API (AI Component)

The Gemini API is the newest multimodal Google AI generative model. It can generate images and text in various languages with a 32k text context, which helps deal with large text files. This API has a "gemini-pro" model for text only and "gemini-pro-vision" for interpreting text and images.

Gemini API receives a prompt and generates a response based on what is requested. We use the prompt presented in Listing 5.2 to generate the sentences from the transcribe.

```
prompt = f"While you neither invent words nor translate, and stay strict to the
    original text without removing the words from that text, separate the following
    text into sentences with no more than 32 tokens:\n{transcription}"
```

Listing 5.2: Prompt for Gemini API.

Gemini can be used with default configuration, or you could set the parameters to define the expected behavior:

- temperature: temperature controls the degree of randomness in token selection.
- max_output_tokens: tokens are selected from most probable to least until the sum of their probabilities equals the top_p value.
- top_p: tokens are selected from most probable to least until the sum of their probabilities equals the top_p value.
- top_k: a top_k of 1 means the selected token is the most probable.

We use the parameters presented in Listing 5.3 to define a more deterministic behavior.

```
parameters = {
     "temperature": 0.2
     "max_output_tokens": 16384
     "top_p": 0.2
     "top_k": 1
 }
```

Listing 5.3: Parameters for Gemini API.

The use of multimodal AI generative can potentially simplify several tasks in an application and deserves a more profound analysis in future works.

5.3 Chapter Summary

The application will consist of recording the participant's screen, face, and audio, processing the video, recognizing emotions, calculating usability, affect, and satisfaction, and generating a new video with UX elements information. Solutions provided by Google are used, including the Artificial Intelligence solution Google Vertex AI, which includes Vision API and Google Natural Language API, integrated through a Python API executed in Google Cloud Function.

Emotion measurement uses the intensity and valence of the identified emotion or sentiment and the overall confidence level of the result. The final UX evaluation is presented on a 5-point scale, from “very poor” to “very good” in the first analysis and on a 3-point scale in the second.

Chapter 6

Discussion

6.1 Implementation Overview

We implemented the proposed architecture to validate the model and executed an exploratory experiment with the proposed process.

6.1.1 Model Activities

The activities performed at this implementation followed the defined in Section 5.1.5. The model specifies a usability experiment to be carried out by the participants themselves, so we prepare a usability test script for the participants. We sent an e-mail to the participant with the items below. All files are available to download via the link in their title.

- Invitation to the test: we describe briefly the test, invite the participant, inform them about the terms of participation, especially about the use of data and the recordings, and give instructions to perform the test;
- Privacy Policy: the app's privacy policy, which describes how data is used;
- Terms of Use: the app's terms of use, which describes the conditions to use the app and the limitations of responsibilities;
- Frequently Asked Questions (FAQ): the app's FAQ, which describes the app and its works, especially about the emotion recognition mechanism, but also about the user experience evaluation model;
- Usability Test Script: a document with the presentation of the test, the think-aloud protocol recommendations, and the descriptions of four tasks. The tasks covered pre-execution in UXAPP, execution in the site to be tested, and pos-execution in

UXAPP again. This document guided the participants through all the steps they should take.

The organizer created a meet with the record option enabled and sent it to the participant. Most of the meetings occurred at night. In the meeting, the organizer thanked the participants for their time and instructed them about how to proceed with the test - instructions like being honest, thinking aloud, and expressing feelings or emotions.

Then, the organizer guided the participants through the UXAPP to create their profile, the experiment, and the first task. The participant registered their initial emotional state. The organizer described the task, as it was in the sent script, to guarantee the participant knew the objective. Then, the organizer closed his video and audio and started the meeting recording.

The participants carried out the task, introducing its objective and describing what they were doing, thinking, and feeling. In the end, the participant said they had finished the task or didn't find a way to execute it.

The organizer ended the recording, and the participants registered in the UXAPP the task success and their emotional final state. Next, the participants answered the evaluation questions involving a SUS questionnaire, a satisfaction evaluation, and two fields about the positive and negative sentiment peaks.

Then, the participant created the next task in UXAPP. At this time, the organizer didn't have to guide the participant in UXAPP's uses, but he was available for any doubts the participant could have. Besides, the organizer still described the next task to the participant. As the organizer started the recording, he added the task's record link to UXAPP and sent it to analyses.

6.1.2 Exploratory Experiment

The experiment was planned considering the restrictions presented in Section 5.1.6. The experiment's attributes are shown in Table 6.1.

We create the site uxapp.com.br to publicize UXAPP and the UX evaluation model. The same site was used to carry out the usability test. The home screen view of the UXAPP site is shown in Figure 6.1.

The site can be considered as a minimum viable product (MVP). This MVP comprises the macro-functionalities presented in Table 6.2.

The site design was planned to be pleasant, simple, and consistent. We use the definitions of Table 6.3 in our style guide. We used this style guide on the site and the app, substituting the IBM Plex Font Family for the Roboto Font Family in the app.

Attribute	Description	Motivation
Number of Users	9	It was enough to generate data to validate the model.
Number of Tasks per User	4	It was enough to cover the site's main services.
Rounds	1	We weren't evaluating user experience variation over time.
Expressing Feelings	Think-aloud Protocol	It was necessary to validate the user experience from speech.
Digital Product	UXAPP Site at uxapp.com.br	This site was created to describe UXAPP and allow the test to be carried out.
Modality	Online	It allowed users to carry out the test from anywhere.
Time	60 minutes	This time was necessary to explain the UXAPP usage and to carry out the tasks.
Cost	No Reward	It was necessary to avoid some test biases.

Table 6.1: Attributes of the experiment carried out.

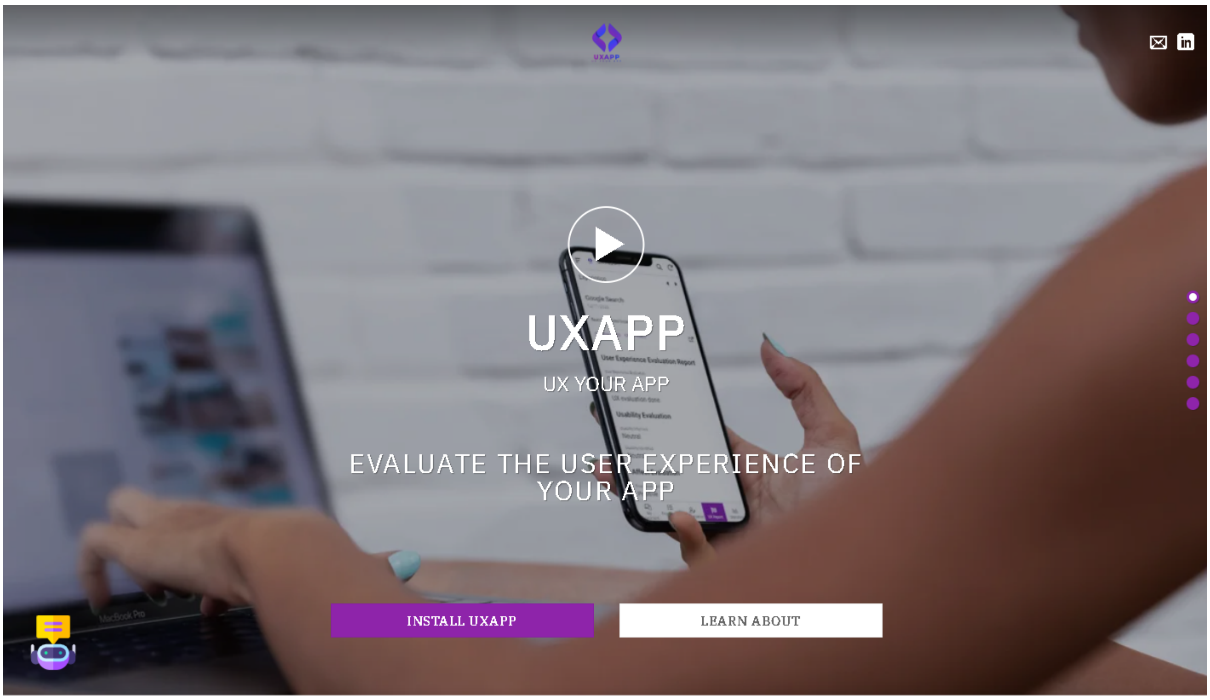

Figure 6.1: View of the home screen of the tested site uxapp.com.br

ID	**Name**	**Description**
F1	UXAPP Description	The site is a landing page that presents the UXAPP with a tagline and a portfolio of photos. Besides, a video in the page header explains what the UXAPP is and what it is for.
F2	Frequently Asked Questions (FAQ)	This functionality explains in detail what UXAPP is, the UX Evaluation Model, and the theoretical background behind the model.
F3	Install APP	A functionality to drive the user to the mobile or web install page.
F4	Request an Invitation	As the app is still under development, this functionality allows users to request trial access to the UXAPP.
F5	Chat with UXBOT	This functionality gives access to a Large Language Model (LLM) chat integration. It lets users obtain tips and tools, learn about design, or translate documents. This item was planned to be used only with registered users, but it was made available without it to perform the usability test.
F6	Contact Us	A simple message box to give a contact e-mail to the user.
F7	Share UXAPP	This functionality allows users to share the page with their fellows.

Table 6.2: Summary of macrofunctionalities of the current MVP of the UXAPP site.

Description	**Detail**	**Motivation**
Primary color	Purple with code #8E24AA	The color purple is associated with innovation and represents something rare. It is also associated with royalty in some contexts. It can be used to denote maturity to contrast with the childhood pink. In UXAPP, it denotes innovation and the design process.
Secondary color	Blue with code #114FEE	The color blue represents harmony and peace. It is the people's most preferred color. In UXAPP, this color means the innovation process can be done with harmony and reliability.
Headlines and Navigation Font	IBM Plex Sans	The SANS type is used here to bring impact and reduce visual noise to the headlines and navigation fonts. IBM Family Fonts is a widely known set of fonts with consistency. It's familiar to the users, and it has an open-source license.
Base and Alt Font	IBM Plex Serif	The SERIF type is used here to facilitate the readness of the site text paragraphs. IBM Family Fonts is a widely known set of fonts with consistency. It's familiar to the users, and it has an open-source license.
Logo	A diamond shape with the meeting of two boomerangs.	The diamond logo represents the convergency and divergency process of design. The meeting of two boomerangs represents complementary ideas. Some ideas grow and evolve, and others return to aggregate in a different way. The logo has primary and secondary colors to denote an innovation with a harmonious process.

Table 6.3: Style Guide definitions and its motivations.

On the other hand, some obstacles were added to the site to allow identifying whether the user shows some negative sentiment. These obstacles are presented in Table 6.4.

ID	Obstacle	Description
O1	Newsletter Popup	After 20 seconds of the user accessing the website, a popup with five fields appeared and asked them to be filled in to allow the user to continue. Despite this, there was an option to close the popup. It was shown at every page accessed or reloaded. Independently what the user answers, the form doesn't work.
O2	Fail to request invitation	The request invitation form triggers an error message: "The e-mail address entered is invalid. Please check and try again", independently of the user e-mail is correct.
O3	Link to nowhere	The "Learn about" panels had a link to a page with only other panels without any images or texts that made sense.
O4	Lack of translation	The site didn't offer a language other than English, so users with Portuguese as their native language could not see the site in their native language.

Table 6.4: Obstacles were intentionally added to the site.

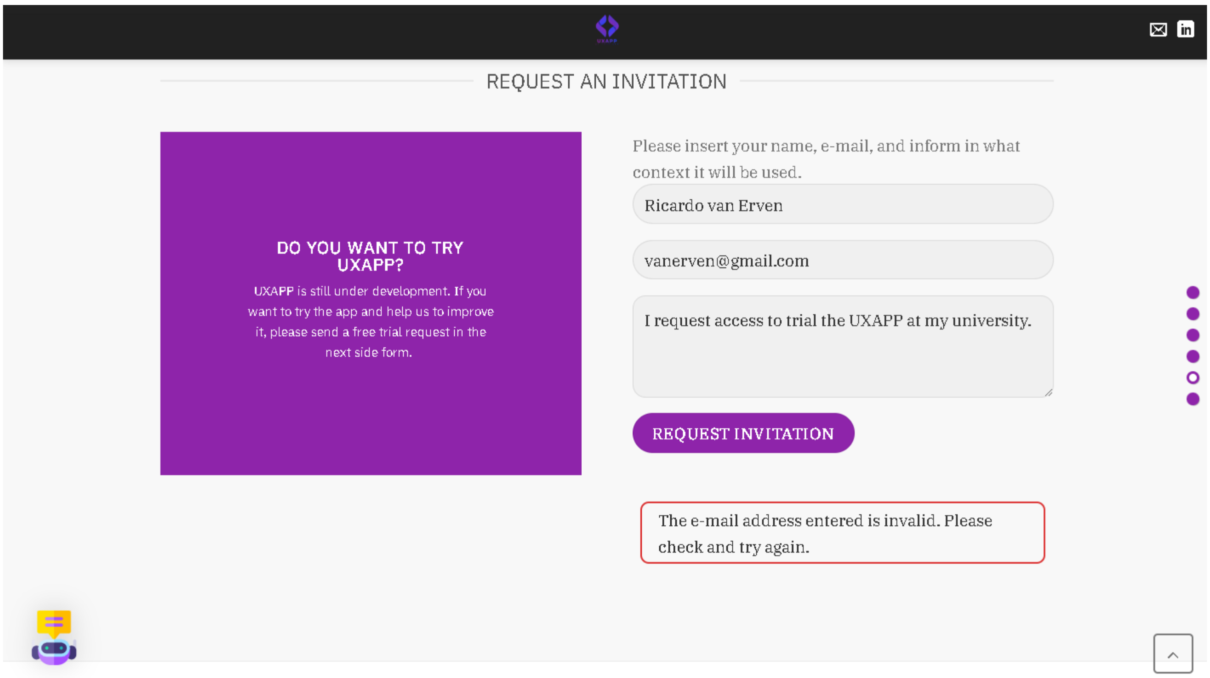

Figure 6.2: The error message of obstacle O2.

We prepare four tasks to execute with this site. The tasks are presented in Table 6.5. All experiment tasks were valid and received an ID, as shown in Table 6.6.

ID	Objective	Description
T1	Describe what UXAPP is and what it is for.	1) Look at this page and tell us what you think of it. Scroll down if you want, but don't click on anything yet. 1.1) What catches your attention? 1.2) Is it pleasant or not? 1.3) What can you do in it, and what is it for? 1.4) What emotions does it evoke? (e.g., joy, sadness, anger, surprise, neutrality) 1.5) What type of affection or feeling does it evoke? (e.g., positive, neutral, or negative) 2) Search for information about UXAPP. Now, you can click on the website. 2.1) Describe what UXAPP is. 2.2) Describe what UXAPP is for.
T2	Request an invitation to install UXAPP.	1) As the application is under development, an invitation is required to install it. 1.1) Identify on the page how you can obtain the invitation 1.2) Request that the invitation be sent to your email. Important: Try to carry out this task within the defined time of 3 minutes.
T3	Get a usability test script in German.	1) To perform usability testing, you need tips and tools, such as a usability testing script. The script you are preparing is for students at a German university. 1.1) Identify a way to obtain a usability test script on the website. 1.2) Get a script in German to test the Technische Universität München website with their students.
T4	Share the UXAPP website	1) There are friends who you believe may need UXAPP. 1.1) Identify a way to share UXAPP with your friends. 1.2) Share the UXAPP with some friends or share it to your email so you can forward it at another time. Important: Go until the website shows a sharing screen.

Table 6.5: Usability Test Tasks.

Experiement Task ID	User Number	Task Number	Status
ET1	1	1	Valid
ET2	1	2	Valid
ET3	1	3	Valid
ET4	1	4	Valid
ET5	2	1	Valid
ET6	2	2	Valid
ET7	2	3	Valid
ET8	2	4	Valid
ET9	3	1	Valid
ET10	3	2	Valid
ET11	3	3	Valid
ET12	3	4	Valid
ET13	4	1	Valid
ET14	4	2	Valid
ET15	4	3	Valid
ET16	4	4	Valid
ET17	5	1	Valid
ET18	5	2	Valid
ET19	5	3	Valid
ET20	5	4	Valid
ET21	6	1	Valid
ET22	6	2	Valid
ET23	6	3	Valid
ET24	6	4	Valid
ET25	7	1	Valid
ET26	7	2	Valid
ET27	7	3	Valid
ET28	7	4	Valid
ET29	8	1	Valid
ET30	8	2	Valid
ET31	8	3	Valid
ET32	8	4	Valid
ET33	9	1	Valid
ET34	9	2	Valid
ET35	9	3	Valid
ET36	9	4	Valid

Table 6.6: Experiment task IDs.

6.1.3 Application

UXAPP automatically evaluates user experience (UX) based on usability, affect, and user value. It uses emotion recognition obtained from audio and video recordings of the user while using a web or mobile digital product. The UXAPP was planned to support the usability test, collect data, and analyze it automatically to generate a UX report about the interaction. UXAPP can evaluate experiments with the user's native language in English or Portuguese.

To illustrate how UXAPP works, let's see an actual experiment. The user authorized this experiment publicity in a special request to present some data in this work and is the only one we asked for. UXAPP is shown in Figures 6.3, 6.4, 6.5 for task ID ET1, and has the functionalities presented in Table 6.7.

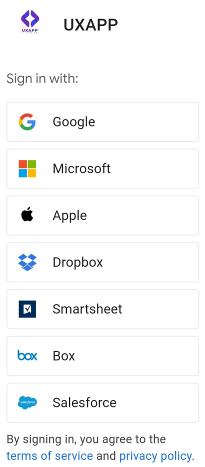

Figure 6.3: Login.

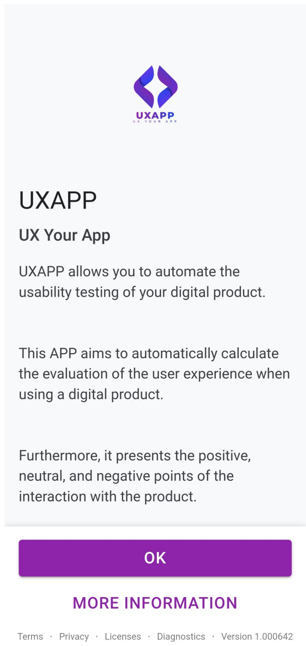

Figure 6.4: About.

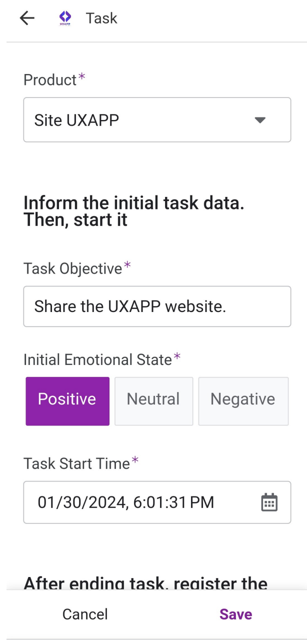

Figure 6.5: Task.

UXAPP uses Google components to perform all the analysis. In this way, we considered the processed data received as valid and correct within the confidence result score. We thought the capacity of these components necessary to recognize micro facial expressions. In this way, it can detect some expressions that are hard to identify without context.

After logging in, the user first creates their profile. Next, the user creates the experiment with the digital product's information to be tested. Next, the user creates the task and executes it with the recommendations of sharing the screen and recording the meeting. Then, the user registers the final task information and evaluates the interaction.

ID	Name	Description
APP1	My Experiment	Allows you to create a new experiment. It's the first step.
APP2	Experiment's Task	Allows you to create experiment tasks. It's the second step.
APP3	My Evaluation	Allows you to evaluate the tested product. It's the third step.
APP4	UX Report	Presents the result of automatically evaluating the user experience regarding product use.
APP5	Statistics	Presents statistical information about your created experiments.
APP6	Menu >Profile	Allows you to change your profile information.
APP7	Menu >Think-Aloud	Presents recommendations on proceeding according to the think-aloud protocol.
APP8	Menu >Feedback	Allows you to provide feedback to the UXAPP team.
APP9	Menu >Share	Allows you to share UXAPP with your acquaintances.
APP10	Menu >Add Shortcut	Allows you to add a shortcut to the application on the home screen.

Table 6.7: UXAPP's Functionalities

Finally, the user informs the link of the Google Meet recording, shares it with UXAPP, and sends the task to analyze. After that, the UX report is available.

6.2 Results

Four tasks were carried out for each participant in the usability test, resulting in 36 tasks. The user filled in the informed data without any direction from the organizer, except for explanations describing the fields. This means that the user fills out the form according to their perception. So, whether the users achieved the task objective or just believed they had successfully finished the task, they marked the task success as "yes". The same is valid for all other informed data.

To illustrate UXAPP results, let's see the task ID ET1. As we said before, the user authorized this experiment publicity with a special request to present some data in this work and is the only one we requested.

We show the "Task 1 - Describe what UXAPP is and what it is for´´, instant 0m32s before UXAPP analysis in Figure 6.6, after UXAPP analysis in Figure 6.7, and the Vision API emotion detection for the same instant in Figure 6.8 with confidence score in Figure 6.9. We use a black box to hide the user name.

UXAPP registered in the instant satisfaction, the affect from video, and the affect from speech. This is shown in the superior right black box. UXAPP also got the transcript of the user's speech and generated a subtitle, which we show in the inferior central black box.

The transcription is also used to evaluate SUS based on sentiment analysis and the similarity with SUS sentences. The transcript of Figure 6.7 says "I don't really like it when I open with a form, this actually leaves me (...)". As we can see, the affect (speech) is negative; besides, the user shows a joyful face.

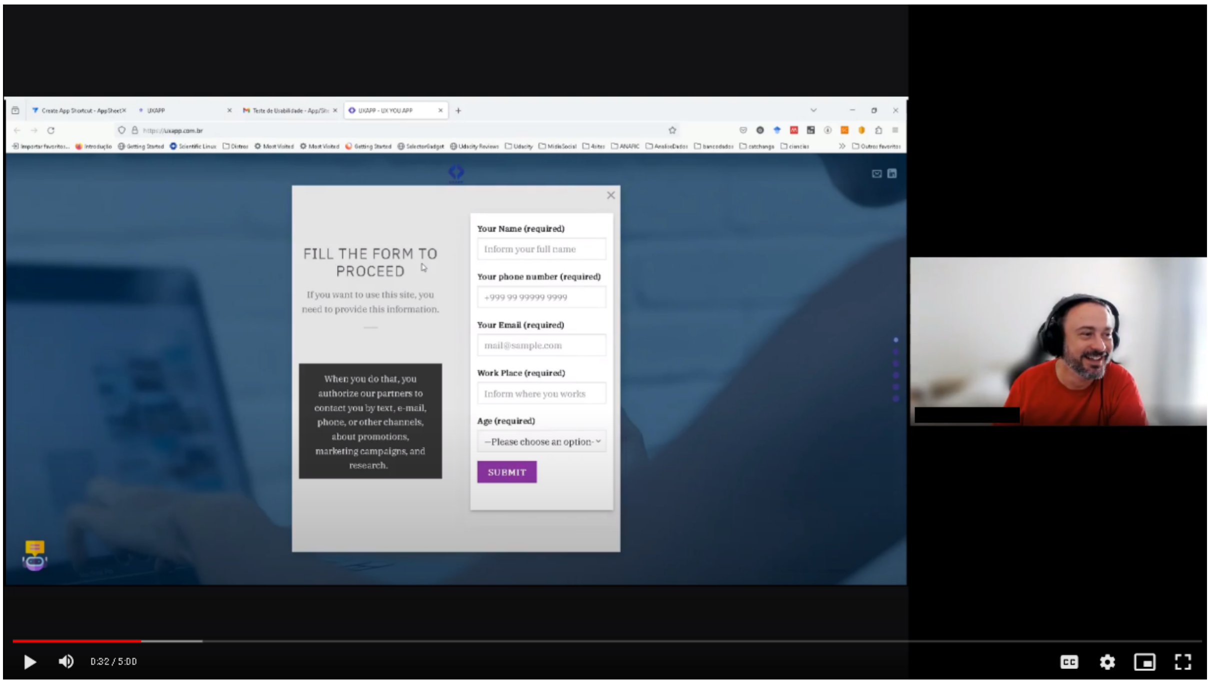

Figure 6.6: Task 1 recording before UXAPP analysis.

After this analysis is done for all videos, UXAPP generates the UX report. Figure 6.10 shows the card with the report of this task. In this card, the user can download the original task recording, the evaluated task video, the video's audio, and the video's subtitle in separate files.

Figure 6.11 shows the UX report header with task details and the user experience evaluation, which is "Good" in this case. This evaluation is based on an independent evaluation of usability, affect, and user value. We can also see the task status as "UX evaluation done".

Figure 6.12 shows the usability and the affect evaluations. The manual and calculated affects are "Positive". The user answers the manual one, and the automatic one is a score of overall positive and negative sentiments identified in the video and the user's speech. The usability evaluation calculates SUS from the manual user answers and the

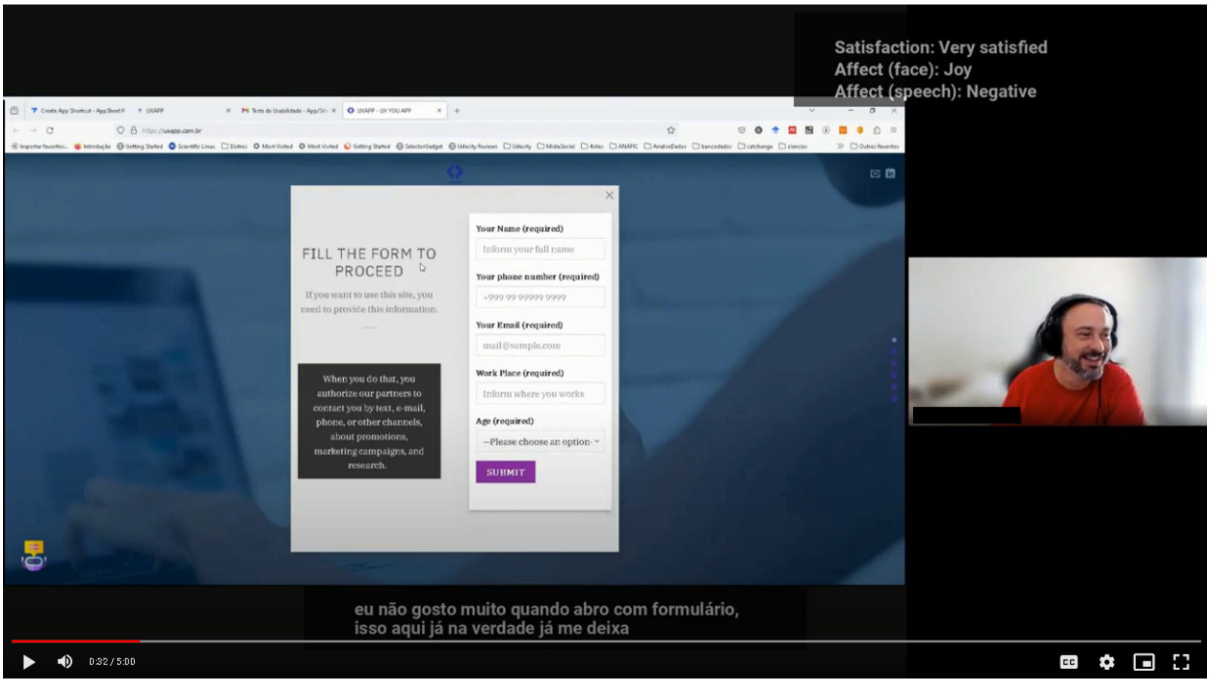

Figure 6.7: Task 1 recording after UXAPP analysis.

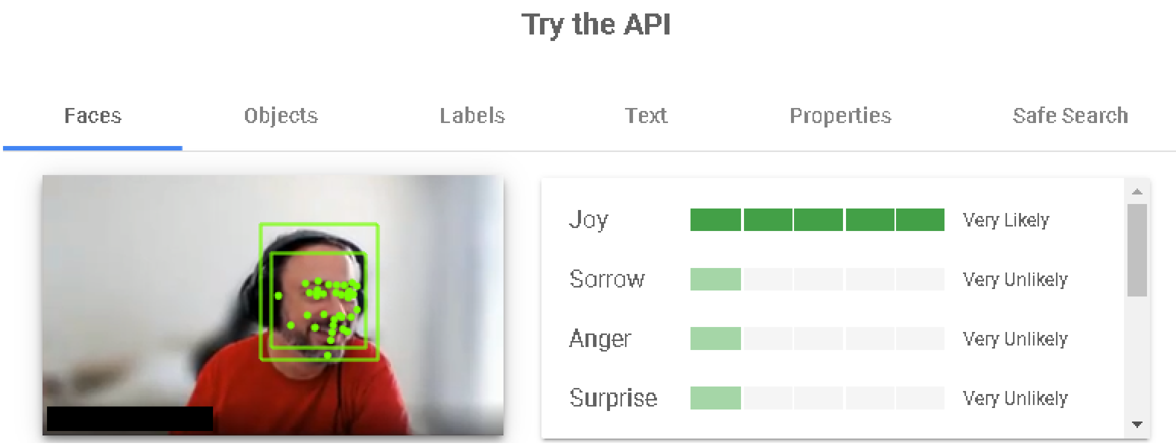

Figure 6.8: User image from task 1 recording in Vision API.

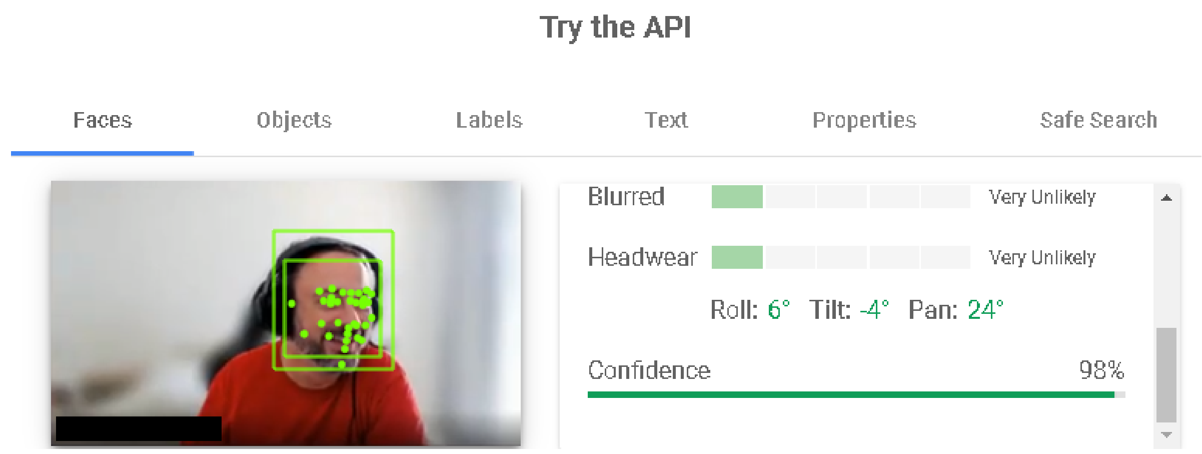

Figure 6.9: Confidence score of user image from task 1 recording in Vision API.

similarity and valence of user speech about SUS sentences. In this case, as we can see, both evaluations were "Good". This rating uses the usability scale defined in Chapter 5. Figure 6.14, in the "Additional Information" panel, shows the evaluated score for manual SUS of 82.5 and the evaluated by UXAPP of 72.5.

Figure 6.13 shows user value evaluation and additional information. The manual user value is a user answer to the question of the user's overall satisfaction with the product. The automatic user value is a score explained in Chapter 5. Both are considered as "Satisfied" in this case. The additional information field brings us more useful information to understand the UX evaluation. It has the user answers about positive and negative sentiment peaks, the calculated sentiment peaks (video and audio), all the sentiments identified, SUS scores, task success, and the task duration based on the informed task start and end. All this information are shown in Figures 6.14, 6.15, 6.16, 6.17, 6.18, 6.19, 6.20, 6.21, and 6.22.

We carried out the usability emotional evaluation based on the user's speech. After transcribing, we request Gemini API to separate the transcription into sentences. We perform a sentiment analysis in the sentences and then compare the similarity between each sentence and all the SUS sentences. To illustrate, let's see the subtitles of task 1 presented in Listing 6.1 and the corresponding sentences generated by Gemini API presented in Listing 6.2.

```
1
0:00:04,000240 --> 0:00:10,000480
hey, I'm going to start by carrying out the first task, so hey, I'm going to access
    the application, I'm going to see

2
0:00:10,000480 --> 0:00:14,000800
```

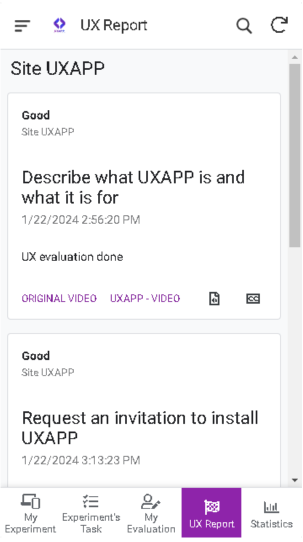

Figure 6.10: Main.

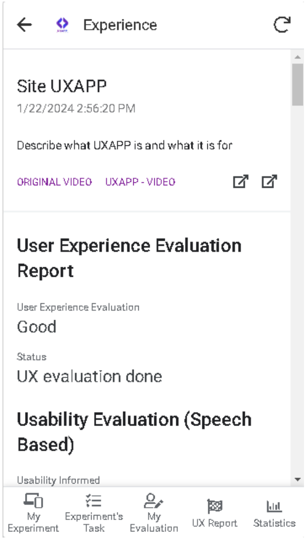

Figure 6.11: Screen 1.

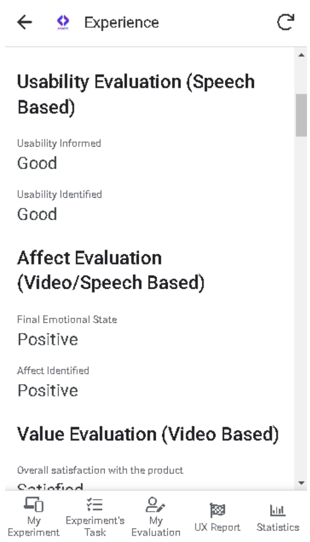

Figure 6.12: Screen 2.

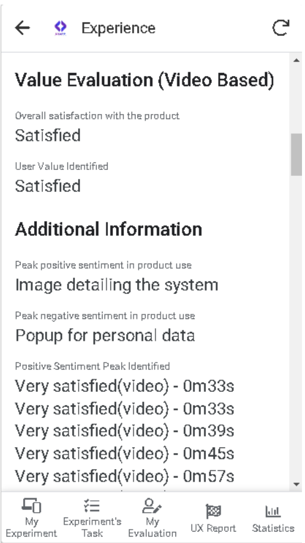

Figure 6.13: Screen 3.

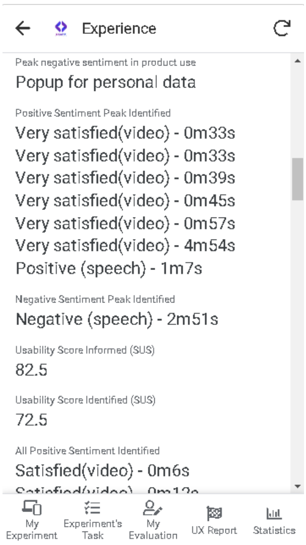

Figure 6.14: Screen 4.

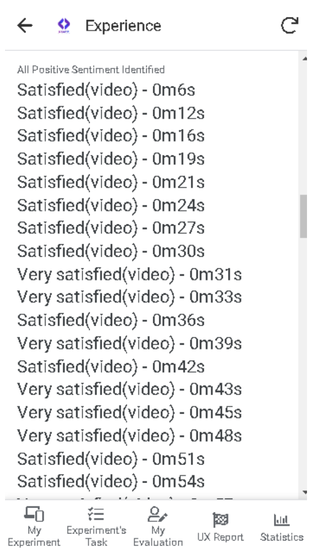

Figure 6.15: Screen 5.

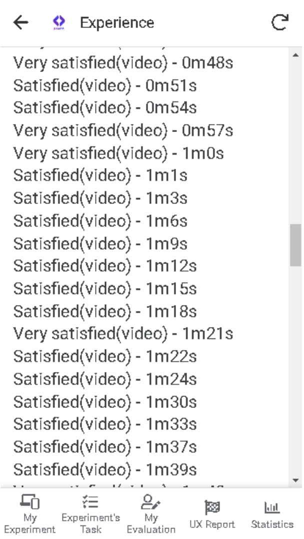

Figure 6.16: Screen 6.

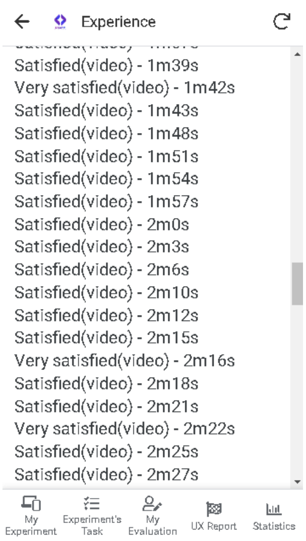

Figure 6.17: Screen 7.

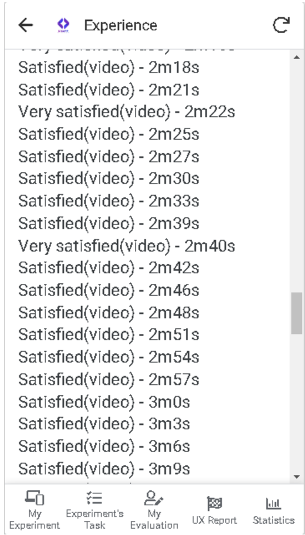

Figure 6.18: Screen 8.

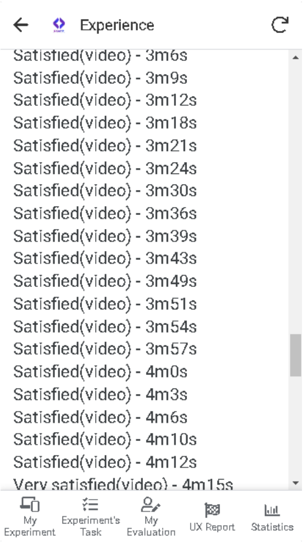

Figure 6.19: Screen 9.

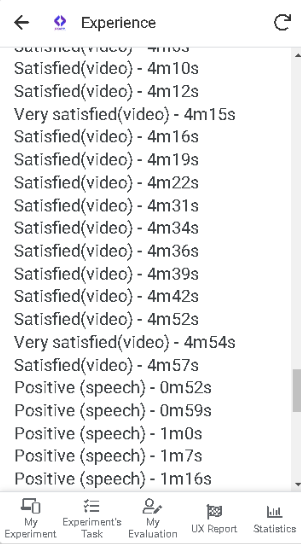

Figure 6.20: Screen 10.

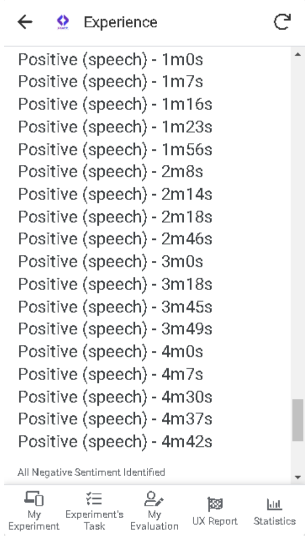

Figure 6.21: Screen 11.

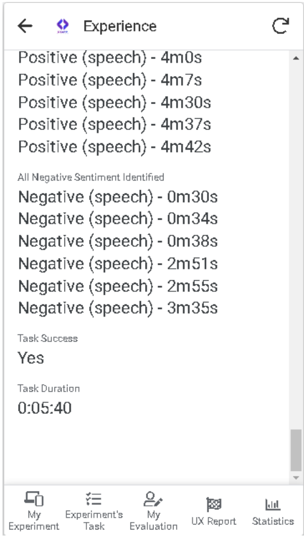

Figure 6.22: Screen 12.

```
what he attracts attention to, whether it is pleasant or not, eh, what he can do to her

3
0:00:14,000800 --> 0:00:20
and what it is for and what emotions it invokes, in addition to the type of affection and feeling

4
0:00:20 --> 0:00:29,000920
that it invokes, it is positive, neutral, negative, eh, okay, application, the...

5
0:00:30,000240 --> 0:00:34,000320
I don’t really like it when I open it with a form, this actually leaves me

6
0:00:34,000320 --> 0:00:38,000840
a little irritated because it forces me to kind of fill out a form without me even having

7
0:00:38,000840 --> 0:00:47,000080
accessed to the right of the page, so I won’t be able to scroll, so I’ll close

8
0:00:47,000080 --> 0:00:52,000280
```

```
okay, so, I, so, so, so the view of the page, right, the environment of the

9
0:00:52,000280 --> 0:00:59,000480
The page is nice, it’s clean, it’s good for well-distributed learning,

10
0:00:59,000480 --> 0:00:59,000920
It’s cool

11
0:01:00,000800 --> 0:01:07,000800
OK, I think it’s cool, ok, there are installation options here, find out more, so

12
0:01:07,000800 --> 0:01:16,000480
this is also good, so scrolling down the page, hey, the stickers are, I liked it

13
0:01:16,000480 --> 0:01:23,000320
of the stickers too, yeah, they’re cool stickers, yeah, apparently they explain well the activities that can

14
0:01:23,000320 --> 0:01:28
be done, okay, so here comes the description, this is also good because you already get

15
0:01:28 --> 0:01:29,000920
the first doubts right from the beginning,

16
0:01:30,000520 --> 0:01:34,000960
OK, so what would it be for, so the application can be scalable, it can help

17
0:01:34,000960 --> 0:01:43,000320
explore the experiments, eh, generate applications, apparently, so that would be it and create applications

18
0:01:43,000320 --> 0:01:56
web interconnecting with other features, features, right, so there are some references , that too
```

```

19
0:01:56 --> 0:01:59,000920
It’s interesting, it gives a stronger foundation, I think it’s cool too,

20
0:02:01,000600 --> 0:02:08,000160
the background part also talks about this, giving more details so if you click

21
0:02:08,000160 --> 0:02:14
here he already plays to the point ah there’s a little guide on the side too

22
0:02:14 --> 0:02:18,000160
This is interesting, okay, it’s cool to know where you are, so

23
0:02:18,000160 --> 0:02:24,000160
This here is also good, I think it’s cool, the sticker is always good, it’s generally good

24
0:02:24,000160 --> 0:02:28,000400
so we can better identify what it does, so apparently it is a software application

25
0:02:30,000120 --> 0:02:38,000200
which allows you to generate experiments for x and for building applications,

26
0:02:38,000200 --> 0:02:46,000200
so it was very clear at the beginning of the page, the little robot is here, oh allow me

27
0:02:46,000200 --> 0:02:51,000120
We can call and communicate with the chat person, it’s cool that he doesn’t

28
0:02:51,000120 --> 0:02:55,000760
I keep sending messages all the time here, ok, it’s a little annoying sometimes

29
0:02:55,000760 --> 0:02:59,000920
```

```
Sometimes when it keeps appearing here you have to keep closing it

30
0:03:00,000240 --> 0:03:06,000320
so this is positive, there is also a part here to get in touch, which is

31
0:03:06,000320 --> 0:03:18,000200
good eh ah it’s also here talking precisely about their bolt here separately

32
0:03:18,000200 --> 0:03:25,000640
of cool contacts so I liked the page it’s a page that wakes me up

33
0:03:25,000640 --> 0:03:29,000920
interest, I could look for more information to better understand how it

34
0:03:30,000120 --> 0:03:35,000280
It works, the only thing that irritates me a little is this issue of you opening

35
0:03:35,000280 --> 0:03:40,000120
the page already asks for your data, I think it could be somewhere else or

36
0:03:40,000120 --> 0:03:45,000640
have another reason, maybe sometimes when there are pages that include, for example

37
0:03:45,000640 --> 0:03:49,000600
Ah, I want to download a book or a folder, something, is that ok?

38
0:03:49,000600 --> 0:03:55,000040
suddenly but you kind of receive a prize there before you give the data

39
0:03:55,000040 --> 0:03:59,000920
It doesn’t look like a prize but it’s not as aggressive when it’s shown straight away.

40
0:04:00,000960 --> 0:04:07,000080
```

```
eh, but other than that, in general, on average, I found it very positive, so with
    that I

41
0:04:07,000080 --> 0:04:21,000280
I finish the first test, first first task ah it invokes eh tranquility, tranquility,
    in terms of

42
0:04:21,000280 --> 0:04:29,000920
feeling she will invoke tranquility, eh, no, perhaps the animation, but not much,

43
0:04:30,000240 --> 0:04:37,000360
But I think it’s more of a little bit of joy, but no no

44
0:04:37,000360 --> 0:04:42,000440
So much so, it seems very encouraging, it looks like an application that will help you
     a lot, consequently

45
0:04:42,000440 --> 0:04:49,000120
brings me a certain expectation, a good expectation, so excitement, let’s say,
    excitement, but not

46
0:04:49,000120 --> 0:04:59,000920
necessarily a euphoria, eh, I think that’s it.
```

Listing 6.1: Subtitles for task ID ET1. It was translated from Portuguese as it was.

```
- Hey, okay, I’ll start by doing the first task.
- I’m going to access the application, see what catches my attention.
- If it is pleasant or not, what can you do with it.
- What it is for and what emotions it invokes.
- In addition to the type of affection and feeling it invokes.
- It is positive, neutral, negative.
- Eh, okay, the application, the... I don’t really like it when I open it with a form.
- This actually makes me a little irritated.
- Because it forces me to kind of fill out a form.
- Without me even accessing the page correctly, okay.
- So I won’t be able to scroll, so I’ll close it, okay.
- Hey, so the view of the page, right?
- The page environment is pleasant.
```

14 - It is clean and is suitable for well-distributed learning.
15 - That’s cool, okay, I think it’s cool.
16 - Ah, you already have the installation options here and find out more.
17 - So that’s also good, so scrolling down the page.
18 - Eh, the stickers are, I liked the stickers too.
19 - Okay, they’re cool stickers, okay.
20 - Apparently they explain well the activities that can be done.
21 - Okay, so here comes the description, that’s good too.
22 - Because you clear up your first doubts right from the beginning.
23 - Okay, so what would it be for, so the application can be scalable.
24 - It can help explore experiments, eh, generate applications.
25 - Apparently, that would be it and create web applications.
26 - Interconnecting features with other characteristics, right?
27 - Eh, so there are some references, that’s also interesting.
28 - It gives a stronger foundation, I think it’s cool too.
29 - And the background part he also talks about giving more details.
30 - So if you click here it will play for the point.
31 - Ah, there’s a little guide on the side, that’s interesting too.
32 - It’s nice to know where you are, so this is also good.
33 - I thought it was cool, the sticker is always good, okay.
34 - It’s generally good for us to better identify what he does.
35 - So apparently it’s an application of x that allows you to eh.
36 - Generate experiments for x and for building applications.
37 - So it was already very clear at the beginning of the page.
38 - The little robot here allows us to call and communicate, right?
39 - With the chat person, it’s nice that he doesn’t spend all the time sending messages here.
40 - Ah, it’s a little annoying sometimes when it keeps appearing here.
41 - You have to keep closing, so this is also positive.
42 - Here’s the part to get in touch, which is good.
43 - Oh, here we are also talking about their bolt and here’s the cool contacts part, so.
44 - I liked the page, it’s a page that sparks my interest.
45 - And I could look for more information to better understand how it works.
46 - The only thing that irritates me a little is that when you open the page it asks for your details.
47 - I think it could be in another place or have another reason, maybe eh.
48 - Sometimes when there are some pages that say, for example, oh, I want to download a book or a folder, something, then all of a sudden everything is fine.
49 - But you kind of receive a prize there before you give the data, right.
50 - It looks like a prize but is not as aggressive when showing it straight away.
51 - Eh, but other than that, in general, on average, I found it very positive.
52 - And with that I finish the first test, first first task, ah, it invokes tranquility.
53 - Tranquility, in terms of feeling it will invoke tranquility.
54 - Eh, no, maybe the animation, but not much, but I think it’s more than eh.
55 - It’s a bit of joy, but not that much, it seems quite encouraging.

```
- If it looks like an application that will help you a lot, consequently it gives me a
    certain expectation, a good expectation, then animation.
- Let’s put it this way, excitement, but not necessarily euphoria, eh, I think that’s
    it.
```

Listing 6.2: Sentences obtained from Gemini API for task ID ET1. It was translated from Portuguese as it was.

We can now see a summary of the usability test. Table 6.8 shows the task success and duration. In total, we recorded 5h 35m 31s of task execution. The average task duration is 9m 19s. The average processing time for analyzing each task is 24m 37s. The percentage of task success is 47.22%, and unsuccess is 52.78%. The task success for the experiment’s task is shown in Figure 6.23.

Task 1 depended on an understanding of the site description. It was a challenge for some users because it was in English and contained technical subjects, despite the user using the app what the website talked about. We considered this lack of understanding as an external problem. Some users had difficulty because part of the site does not fully load. As the site has several images and they were not treated to keep quality while having a small size, the low-speed internet of these users generates an external problem. This task had a success rate of 44.44%.

Task 2 was impossible to finish because of an intentional error in e-mail validation. Despite this, two users considered they successfully finished the task by considering they found the feature and input their data. This task had a success rate of 22.22%.

Task 3 required previous knowledge about chat interaction with generative IA. Some users try to find a document to download and navigate across the site to see it. One user tried to use the UXBOT, but there was a problem external to the task with the bot. This task had a success rate of 33.33%.

Task 4 was generally considered easy, and only one user didn’t finish it. It happens because the user tried to answer the intentionally problematic newsletter form. This task had a success rate of 88.88%.

As we can see, the majority of results were task unsuccess. No user considered they had finished all four tasks. Two users considered they had success only in the last task. One user considered they had success in no task.

6.2.1 First Round of Analysis

The first round of analysis was made with initial parameters defined in Chapter 5. The overall analysis brings a significant match rate in each UX element analysis. We found, however, two main issues with these results. First, we identified some emotions that didn’t

ID	Task Success	Duration
ET1	Yes	0:05:40
ET2	Yes	0:06:37
ET3	No	0:10:55
ET4	Yes	0:03:33
ET5	No	0:18:06
ET6	No	0:08:44
ET7	Yes	0:08:59
ET8	Yes	0:06:05
ET9	Yes	0:10:48
ET10	No	0:05:15
ET11	Yes	0:09:32
ET12	Yes	0:04:18
ET13	No	0:19:02
ET14	No	0:06:38
ET15	Yes	0:07:51
ET16	Yes	0:05:42
ET17	No	0:21:46
ET18	No	0:06:34
ET19	No	0:08:18
ET20	No	0:07:36
ET21	Yes	0:09:44
ET22	No	0:05:28
ET23	No	0:08:58
ET24	Yes	0:05:16
ET25	Yes	0:17:45
ET26	Yes	0:05:27
ET27	No	0:08:08
ET28	Yes	0:06:20
ET29	No	0:13:24
ET30	No	0:07:16
ET31	No	0:11:30
ET32	Yes	0:03:05
ET33	No	0:23:04
ET34	No	0:07:42
ET35	No	0:09:47
ET36	Yes	0:10:38

Table 6.8: Summary of the task success and duration.

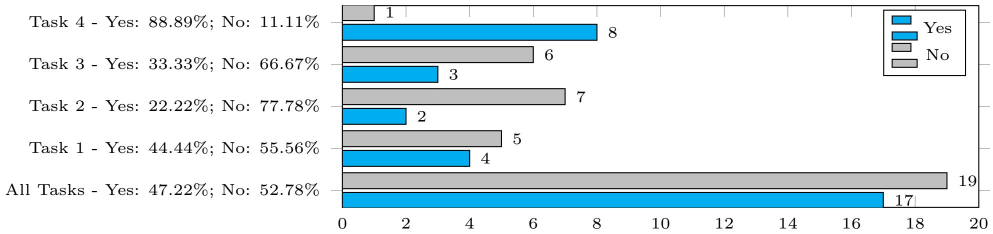

Figure 6.23: Comparative task success rate.

correspond to the real ones. This is related to excessive positive sentiments identified in the video.

The second issue was the large scale chosen to represent the usability, user value, and UX evaluation. UXAPP could evaluate usability and the other elements by distinguishing between good or poor without an exact match with the user's answer. We realized the precision rate of a 5-point scale was unnecessary to identify positive, neutral, and negative points of digital product use.

Emotional Usability Evaluation for the First Analysis

We tabulate the results of the SUS evaluation and SUS score. We check the percentage of tasks in the UXAPP answer that match the user's answers, the percentage of tasks with a mismatch between the user's answer and UXAPP but with both answers in the same direction (positive or negative SUS), and the percentage of tasks with a mismatch between the user's answer and UXAPP and with answers in different directions. We consider the same direction when both values exceed 50 or are below 50. The usability data is presented in Table 6.9. We can see the match percentage in Figure 6.24.

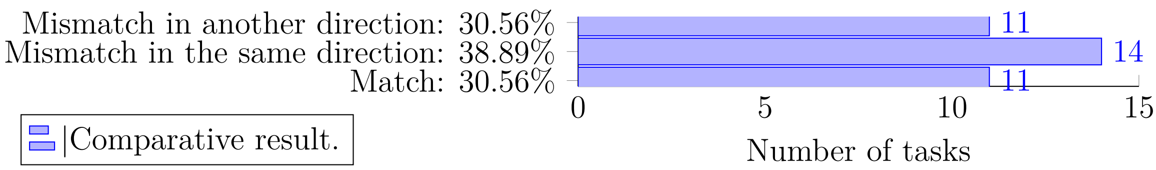

Figure 6.24: Distribution of usability evaluation match for the first analysis.

As we see, UXAPP directly matches the user perception for 30.56% of the results. UXAPP also gives the right direction to 38.89% of the results, which results in 69.45% of close results.

As we said before, we believed the 5-point scale precision didn't deliver the right message. So, we reduce the evaluation scale to 3 points in the second analysis.

ID	Usability Informed	Usability Identified	SUS Score (Informed)	SUS Score (Identified)
ET1	Good	Good	82.5	72.5
ET2	Neutral	Neutral	40	37.5
ET3	Poor	Very poor	20	6.1
ET4	Very good	Good	97.5	67.5
ET5	Poor	Poor	20	15
ET6	Poor	Poor	30	16.1
ET7	Good	Good	72.5	65
ET8	Good	Neutral	67.5	50
ET9	Poor	Neutral	32.5	60
ET10	Neutral	Very poor	52.5	11.1
ET11	Good	Neutral	80	52.5
ET12	Good	Very poor	80	10
ET13	Neutral	Very poor	60	5
ET14	Neutral	Very poor	55	6.1
ET15	Good	Neutral	80	37.5
ET16	Good	Neutral	70	55
ET17	Neutral	Neutral	40	40
ET18	Neutral	Poor	45	32.5
ET19	Neutral	Poor	47.5	16.1
ET20	Neutral	Poor	45	15
ET21	Neutral	Neutral	50	50
ET22	Neutral	Good	62.5	82.5
ET23	Poor	Poor	32.5	21.1
ET24	Very good	Poor	85	15
ET25	Very good	Neutral	85	52.5
ET26	Neutral	Very poor	40	6.1
ET27	Neutral	Very poor	62.5	5
ET28	Neutral	Good	62.5	7.3
ET29	Neutral	Poor	50	17.5
ET30	Neutral	Poor	13.2	20
ET31	Poor	Poor	22.5	32.5
ET32	Good	Good	75	20.3
ET33	Neutral	Very poor	62.5	12.5
ET34	Neutral	Poor	25.2	14.1
ET35	Poor	Poor	27.5	27.5
ET36	Very good	Poor	87.5	25

Table 6.9: First analysis of usability evaluation data.

Affect Evaluation for the First Analysis

We tabulate the results of the affect evaluation. We check the percentage of tasks with a match between UXAPP and the user's answer. The affect data is presented in Table 6.10. We can see the match percentage in Figure 6.25. We consider "same direction" when one value is "Neutral" and the other positive or negative and "different direction" when one value is positive and the other negative.

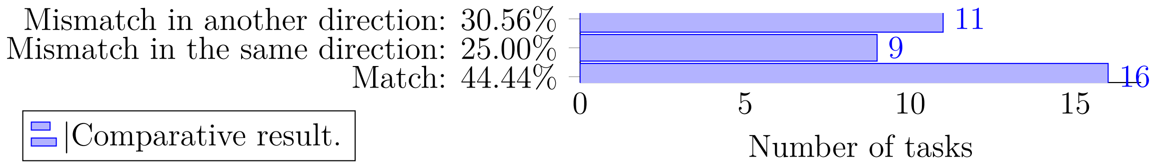

Figure 6.25: Distribution of affect evaluation match for the first analysis.

As we see, UXAPP directly matches the user perception for 44.44% of the results. UXAPP also gives the right direction to 25.00% of the results, which results in 69.44% of close results. We also can see the results regarding sentiment changes versus task success. We consider "stable" the tasks that haven't changed between the initial and final emotional states. Otherwise, we assume that a "change" happens. This approach is shown in Figure 6.26

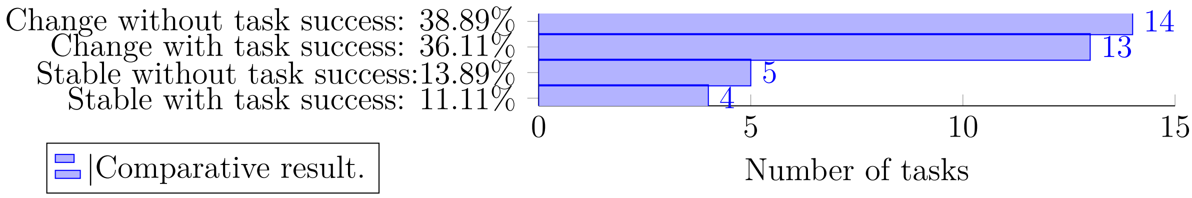

Figure 6.26: Distribution of sentiment changes according to task results.

We expected the "stable with task success" to have a positive final state, as task success motivates the user. In the same way, if there was no change, the initial state also should be positive. Indeed, 3 of 4 tasks in this state have a positive initial emotional state. We show these tasks in Figure 6.27.

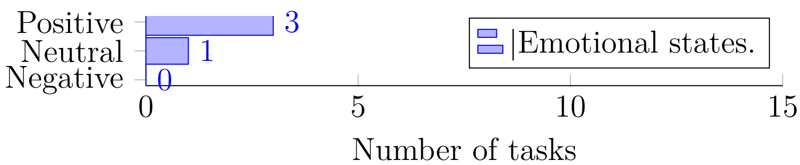

Figure 6.27: Tasks in the state "stable with task success" per initial emotional state.

Similarly, we expected that the tasks in the state "change without task success" to have a negative final emotional state due to the bad experience of the task. Indeed, 12

ID	Initial Emotional State	Final Emotional State	Affect Identified	Task Success
ET1	Positive	Positive	Positive	Yes
ET2	Positive	Negative	Positive	Yes
ET3	Positive	Negative	Positive	No
ET4	Positive	Neutral	Positive	Yes
ET5	Negative	Negative	Positive	No
ET6	Negative	Negative	Positive	No
ET7	Negative	Positive	Positive	Yes
ET8	Neutral	Positive	Positive	Yes
ET9	Neutral	Negative	Positive	Yes
ET10	Negative	Negative	Positive	No
ET11	Negative	Positive	Positive	Yes
ET12	Positive	Neutral	Positive	Yes
ET13	Neutral	Neutral	Negative	No
ET14	Neutral	Neutral	Negative	No
ET15	Neutral	Positive	Positive	Yes
ET16	Neutral	Neutral	Neutral	Yes
ET17	Neutral	Negative	Neutral	No
ET18	Neutral	Negative	Negative	No
ET19	Neutral	Negative	Negative	No
ET20	Neutral	Negative	Negative	No
ET21	Positive	Neutral	Neutral	Yes
ET22	Positive	Negative	Positive	No
ET23	Positive	Negative	Positive	No
ET24	Positive	Positive	Positive	Yes
ET25	Positive	Neutral	Positive	Yes
ET26	Positive	Negative	Positive	Yes
ET27	Positive	Neutral	Positive	No
ET28	Positive	Positive	Positive	Yes
ET29	Neutral	Negative	Negative	No
ET30	Positive	Neutral	Negative	No
ET31	Positive	Negative	Negative	No
ET32	Negative	Positive	Positive	Yes
ET33	Positive	Negative	Positive	No
ET34	Positive	Negative	Neutral	No
ET35	Neutral	Negative	Negative	No
ET36	Negative	Positive	Negative	Yes

Table 6.10: Affect evaluation data for the first analysis.

of 14 tasks in this state have a negative final emotional state. We show these tasks in Figure 6.28.

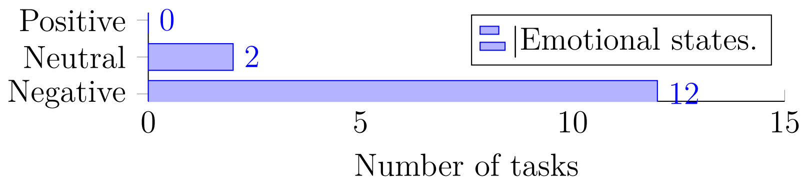

Figure 6.28: Tasks in the state "change without task success" per final emotional state.

These data allow us to conclude there's a stronger relation between emotional state and task success or unsuccess. The greater the task's unsuccessful rate, the more negative the final emotional state rate. In the same way, in tasks without emotional change, the greater the task success rate is, the more positive the emotional initial state is.

UXAPP identified positive and negative sentiments from speech and video. This is shown in Figure 6.29. Attracts attention to the numerous positive video times identified what is discussed in Subsection 6.2.1.

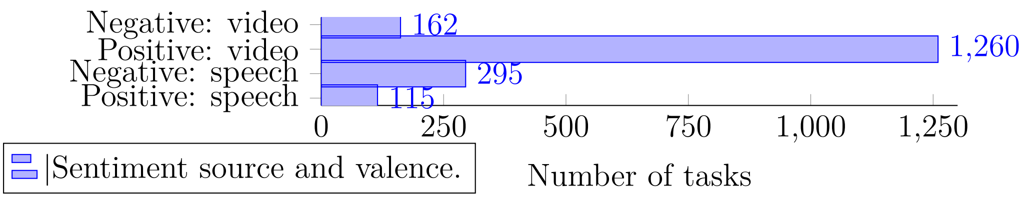

Figure 6.29: Positive and negative sentiments identified in video and speech for the first analysis.

Emotional Satisfaction Evaluation for the First Analysis

We tabulate the results of the user value evaluation. We check the percentage of tasks with a match between UXAPP and the user's answer. The user value data is presented in Table 6.11. We can see the match percentage in Figure 6.30. We consider "same direction" when one value is "Neither satisfied nor dissatisfied" and the other "Very satisfied" / "Satisfied" or "Very unsatisfied" / "Dissatisfied", and "different direction" when one value is "Very satisfied" / "Satisfied" and the other "Very unsatisfied" / "Dissatisfied".

We observe an equilibrated distribution over the results. UXAPP matched 33.33% of the results and got a close result of other 33.33%.

Watching some videos, we could realize some emotions detected didn't correspond to the real ones. So, we increased the emotion recognition confidence score to 0.8 and ran the analysis for the task ET9 again. This task had 118 positive emotion times identified.

ID	Manual Geral Satisfaction	User Value Identified
ET1	Satisfied	Satisfied
ET2	Dissatisfied	Satisfied
ET3	Dissatisfied	Satisfied
ET4	Very satisfied	Satisfied
ET5	Dissatisfied	Satisfied
ET6	Dissatisfied	Satisfied
ET7	Satisfied	Satisfied
ET8	Satisfied	Satisfied
ET9	Neither satisfied nor dissatisfied	Satisfied
ET10	Very unsatisfied	Satisfied
ET11	Very satisfied	Satisfied
ET12	Satisfied	Satisfied
ET13	Neither satisfied nor dissatisfied	Satisfied
ET14	Dissatisfied	Dissatisfied
ET15	Very satisfied	Satisfied
ET16	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET17	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET18	Dissatisfied	Neither satisfied nor dissatisfied
ET19	Dissatisfied	Neither satisfied nor dissatisfied
ET20	Dissatisfied	Neither satisfied nor dissatisfied
ET21	Satisfied	Satisfied
ET22	Neither satisfied nor dissatisfied	Satisfied
ET23	Dissatisfied	Satisfied
ET24	Very satisfied	Satisfied
ET25	Satisfied	Very satisfied
ET26	Neither satisfied nor dissatisfied	Very satisfied
ET27	Satisfied	Satisfied
ET28	Satisfied	Satisfied
ET29	Neither satisfied nor dissatisfied	Dissatisfied
ET30	Dissatisfied	Dissatisfied
ET31	Very unsatisfied	Neither satisfied nor dissatisfied
ET32	Very satisfied	Neither satisfied nor dissatisfied
ET33	Dissatisfied	Satisfied
ET34	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET35	Very unsatisfied	Neither satisfied nor dissatisfied
ET36	Satisfied	Neither satisfied nor dissatisfied

Table 6.11: User value evaluation data for the first analysis.

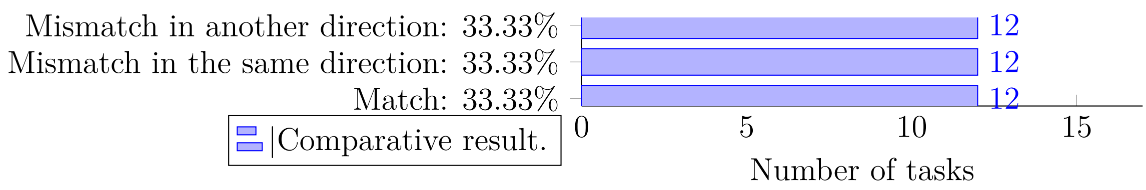

Figure 6.30: Distribution of user value evaluation match for the first analysis.

After the confidence score was adjusted, we got 119 positive emotion times. So, we undo this change and take some video prints to test directly in the Vision API site with the trial option. We verified that the evaluation points table wasn't driving the right emotion because we were optimistic in interpreting the emotion recognition scale ("unknown" to "very likely"). After analysis, we defined that only "very likely" can be used confidently to define an emotion. We changed the code, reduced the scale to 3 points, and reran the analysis, which we show in Subsection 6.2.2.

To illustrate the issue, we show images of the task ID ET1, in which the UXAPP indicates the user expresses joy in Figure 6.31. The image width resolution should be bigger than 1600 px to improve emotion recognition. So we need to crop the user image into the frame, resize it to 1920 px, and send it to Vision API. We show the resized and cropped user image of the joy picture in Figure 6.32. After a few seconds, another image of the same task, in which UXAPP indicates neutrality, as shown in Figure 6.33.

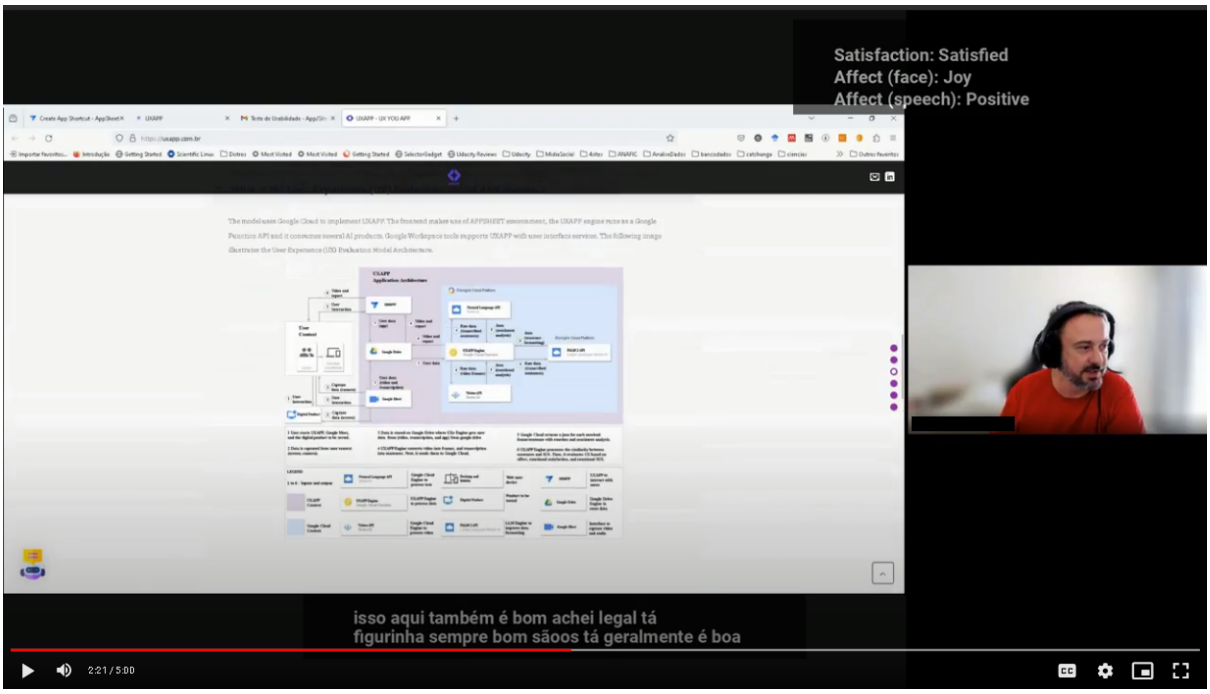

Figure 6.31: UXAPP shows joy.

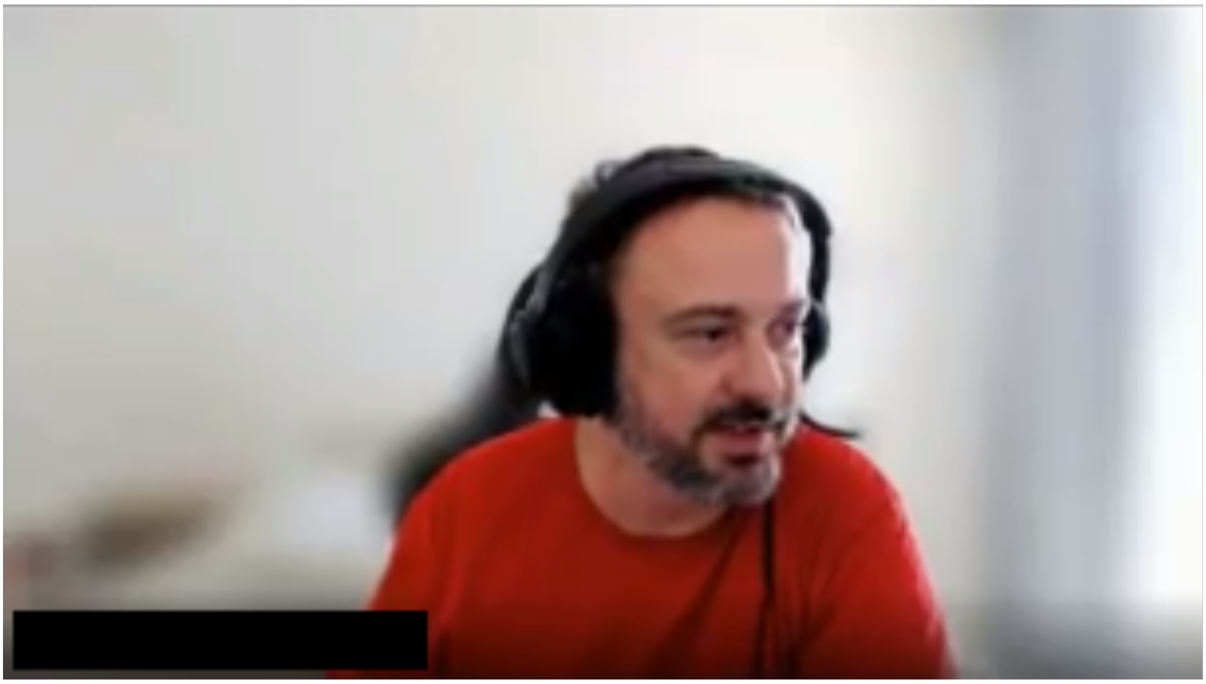

Figure 6.32: Resized and cropped user image of the joy figure.

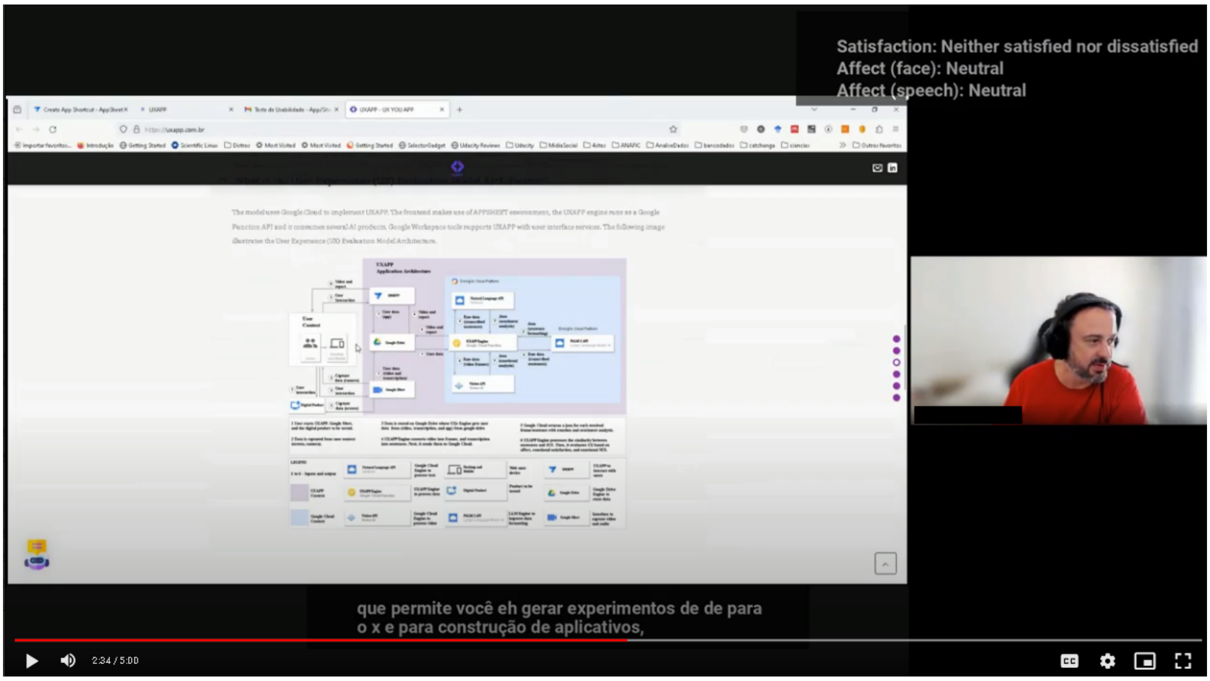

Figure 6.33: UXAPP shows neutrality.

User Experience Evaluation for the First Analysis

We tabulate the results of the user experience evaluation. The data is presented in Table 6.12. We show the distribution of the values in Figure 6.34.

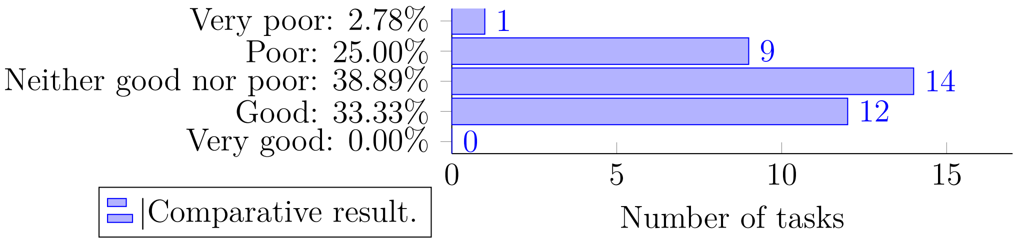

Figure 6.34: Distribution of user value evaluation match for the first analysis.

As we can see, the positive values sum 33.33%, and the negative values sum 27.78%, a difference of 5.55% between them. When we compare it with the task success rate, we observe the similarity in the results. The task success rate, shown before, is 47.22%, and the unsuccess rate is 52.78%, a difference of 5.56%. This analysis shows a balance between the positive and the negative sides, which appears in UXAPP's UX evaluation and the task results, as expected.

We tabulate the user manual evaluation and calculate the UX evaluation based on the UXAPP approach with the user input data. This data is presented in Table 6.13. The UX evaluation is shown in Figure 6.35. There is a significant proximity between the UX evaluation based on the user input and the UXAPP analysis. We remember, however, that these results are influenced by the excessive number of positive sentiments identified in the videos.

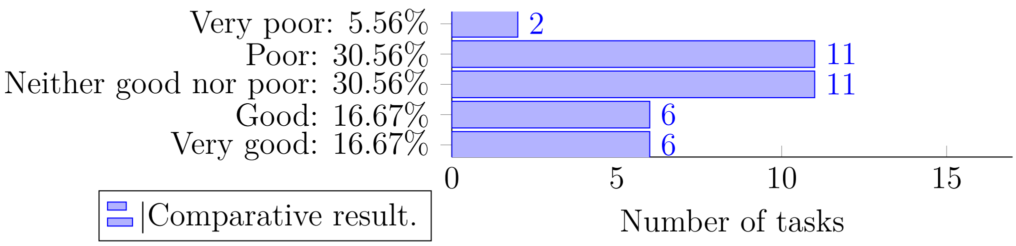

Figure 6.35: Distribution of UX evaluation with user input data for the first analysis.

6.2.2 Second Round of Analysis

The second round of analysis was made with the final parameters defined in Chapter 5. The overall analysis brings the most consistent result regarding usability and satisfaction analysis.

ID	User Experience Evaluation
ET1	Good
ET2	Good
ET3	Neither good nor poor
ET4	Good
ET5	Neither good nor poor
ET6	Neither good nor poor
ET7	Good
ET8	Good
ET9	Good
ET10	Neither good nor poor
ET11	Good
ET12	Neither good nor poor
ET13	Poor
ET14	Very poor
ET15	Good
ET16	Neither good nor poor
ET17	Neither good nor poor
ET18	Poor
ET19	Poor
ET20	Poor
ET21	Neither good nor poor
ET22	Good
ET23	Neither good nor poor
ET24	Neither good nor poor
ET25	Good
ET26	Neither good nor poor
ET27	Neither good nor poor
ET28	Good
ET29	Poor
ET30	Poor
ET31	Poor
ET32	Good
ET33	Neither good nor poor
ET34	Neither good nor poor
ET35	Poor
ET36	Poor

Table 6.12: User experience evaluation data for the first analysis.

ID	Usability Informed	Final Emotional State	Overall Satisfaction
ET1	Good	Positive	Satisfied
ET2	Neutral	Negative	Dissatisfied
ET3	Poor	Negative	Dissatisfied
ET4	Very good	Neutral	Very satisfied
ET5	Poor	Negative	Dissatisfied
ET6	Poor	Negative	Dissatisfied
ET7	Good	Positive	Satisfied
ET8	Good	Positive	Satisfied
ET9	Poor	Negative	Neither satisfied nor dissatisfied
ET10	Neutral	Negative	Very unsatisfied
ET11	Good	Positive	Very satisfied
ET12	Good	Neutral	Satisfied
ET13	Neutral	Neutral	Neither satisfied nor dissatisfied
ET14	Neutral	Neutral	Dissatisfied
ET15	Good	Positive	Very satisfied
ET16	Good	Neutral	Neither satisfied nor dissatisfied
ET17	Neutral	Negative	Neither satisfied nor dissatisfied
ET18	Neutral	Negative	Dissatisfied
ET19	Neutral	Negative	Dissatisfied
ET20	Neutral	Negative	Dissatisfied
ET21	Neutral	Neutral	Satisfied
ET22	Neutral	Negative	Neither satisfied nor dissatisfied
ET23	Poor	Negative	Dissatisfied
ET24	Very good	Positive	Very satisfied
ET25	Very good	Neutral	Satisfied
ET26	Neutral	Negative	Neither satisfied nor dissatisfied
ET27	Neutral	Neutral	Satisfied
ET28	Neutral	Positive	Satisfied
ET29	Neutral	Negative	Neither satisfied nor dissatisfied
ET30	Neutral	Neutral	Dissatisfied
ET31	Poor	Negative	Very unsatisfied
ET32	Good	Positive	Very satisfied
ET33	Neutral	Negative	Dissatisfied
ET34	Neutral	Negative	Neither satisfied nor dissatisfied
ET35	Poor	Negative	Very unsatisfied
ET36	Very good	Positive	Satisfied

Table 6.13: User input data used to calculate UX evaluation for the first analysis.

Emotional Usability Evaluation for the Second Analysis

We tabulate the results of the SUS evaluation and SUS score in the second analysis. We check again the percentage of tasks in the UXAPP answer matches the user's answers, the percentage of tasks with a mismatch between the user's answer and UXAPP but with both answers in the same direction (positive or negative SUS), and the percentage of tasks with a mismatch between the user's answer and UXAPP and with answers in different directions. Again, we consider the same direction when both values are greater than 50 or lower than 50. The usability data is presented in Table 6.14. We can see the match percentage in Figure 6.36.

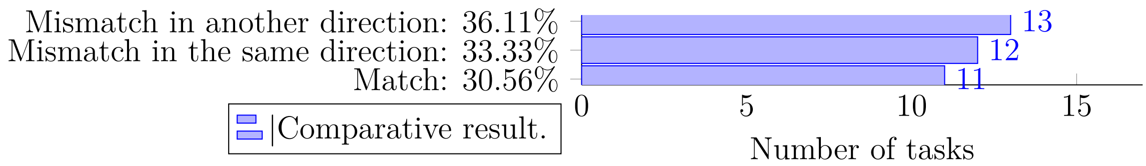

Figure 6.36: Distribution of usability evaluation match of the second analysis.

Due to the reduction of a 5-point scale to a 3-point scale, we expected an increase in the percentage of matches. The match result was the same, slightly changing the other percentages.

One cause of the unexpected result is how we calculate the SUS similarity. To evaluate the score, we selected the sentence that was most similar to the selected SUS sentence among all sentences from the speech. If we sum the contribution of all sentences with similarity, making a ponderated calculation, this procedure could be more precise.

Another cause is the complex procedure of emotional SUS. First, the user needs to say something related to the SUS questions for UXAPP to have something to evaluate. If the user thinks something about the digital product without speaking it, then we can't do any evaluation. We can figure out some scenarios.

In one first scenario, the user says something negative, thinks something positive without speaking it, and ultimately evaluates the task as positive. In this case, the SUS evaluated will be in a different direction from the actual evaluation.

In another scenario, the user says something similar to one SUS sentence but not for all SUS sentences. In this case, the direction of the UXAPP evaluation will be the same as the user evaluation, but it won't match. Still, as the initial UXAPP evaluation is neutral (value 3), the UXAPP evaluation score will be close to the neutral value of 50.

Besides, the analysis must clean empty words of speech and correct punctuation. When people are required to think aloud, some try to speak all the time. Words are spoken slowly, and it doesn't make much sense when they are transcripted, like we see in

ID	Usability Informed	Usability Identified	Usability Score Informed (SUS)	Usability Score Identified (SUS)
ET1	Good	Good	82.50	65.00
ET2	Neutral	Neutral	40.00	47.50
ET3	Poor	Poor	20.00	17.50
ET4	Good	Good	97.50	65.00
ET5	Poor	Poor	20.00	25.00
ET6	Poor	Neutral	30.00	37.50
ET7	Good	Good	72.50	75.00
ET8	Good	Neutral	67.50	35.00
ET9	Poor	Good	32.50	90.00
ET10	Neutral	Poor	52.50	20.00
ET11	Good	Neutral	80.00	47.50
ET12	Good	Poor	80.00	25.00
ET13	Neutral	Poor	60.00	10.00
ET14	Neutral	Poor	55.00	17.50
ET15	Good	Poor	80.00	17.50
ET16	Good	Neutral	70.00	55.00
ET17	Neutral	Neutral	40.00	35.00
ET18	Neutral	Poor	45.00	27.50
ET19	Neutral	Poor	47.50	17.50
ET20	Neutral	Poor	45.00	22.50
ET21	Neutral	Poor	50.00	27.50
ET22	Neutral	Neutral	62.50	52.50
ET23	Poor	Neutral	32.50	40.00
ET24	Good	Poor	85.00	15.00
ET25	Good	Neutral	85.00	55.00
ET26	Neutral	Poor	40.00	17.50
ET27	Neutral	Poor	62.50	20.00
ET28	Neutral	Good	62.50	70.00
ET29	Neutral	Neutral	50.00	37.50
ET30	Neutral	Poor	45.00	17.50
ET31	Poor	Poor	22.50	30.00
ET32	Good	Good	75.00	80.00
ET33	Neutral	Poor	62.50	20.00
ET34	Neutral	Poor	57.50	22.50
ET35	Poor	Neutral	27.50	45.00
ET36	Good	Neutral	87.50	45.00

Table 6.14: Usability evaluation data of second analysis.

subtitle ID 8, “okay, so, I, so, so, so the view of the page, right, the environment of the (...)”.

We also perceived responding to the SUS questionnaire as the moment the user spent the most time in the UXAPP activities. To make the process repeatable, it’s necessary to find a more straightforward way to measure the usability, and it’s essential to research other approaches.

Considering the exposure, it would be interesting to consider the possibilities of large language models (LLM) in understanding user speech with their context and translating it to a usability evaluation. Other usability metrics can be helpful to contribute to the analysis, like the task’s duration and success.

Affect Evaluation for the Second Analysis

We tabulate the results of the affect evaluation of the second analysis. We check the percentage of tasks with a match between UXAPP and the user’s answer. The affect data is presented in Table 6.15. We can see the match percentage in Figure 6.37. Again, we consider the “same direction” when one value is “Neutral” and the other positive or negative and “different direction” when one value is positive and the other negative. We included the number of positive and negative sentiments identified per task and their differences. We presented it in Figure 6.38.

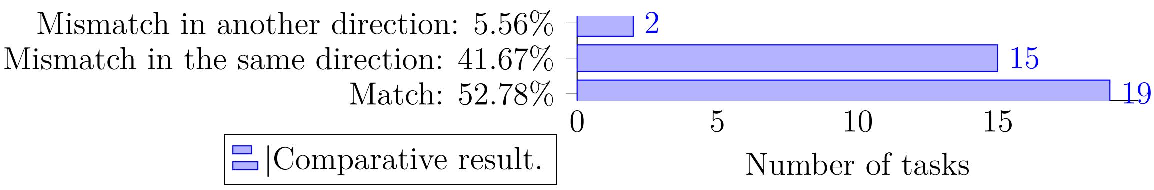

Figure 6.37: Distribution of affect evaluation match of second analysis.

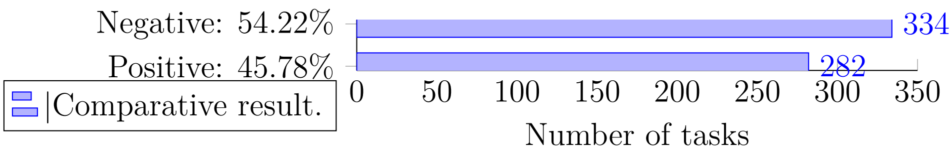

Figure 6.38: Distribution of positive and negative affect identification in the second analysis.

The percentage of positive and negative sentiment identified is close to the number of successful and unsuccessful tasks, demonstrating the consistency of affect data.

This analysis shows a significant increase in “match” and “mismatch in the same direction”. It happens due to the adjustment in the video emotion recognition to consider

ID	Initial Emotional State	Final Emotional State	Affect Identified	Task Success	Identified Positive Sentiments Number	Identified Negative Sentiments Number	Difference
ET1	Positive	Positive	Positive	Yes	27	7	20
ET2	Positive	Negative	Negative	Yes	1	17	-16
ET3	Positive	Negative	Negative	No	1	20	-19
ET4	Positive	Neutral	Positive	Yes	9	2	7
ET5	Negative	Negative	Negative	No	2	19	-17
ET6	Negative	Negative	Negative	No	2	10	-8
ET7	Negative	Positive	Positive	Yes	7	1	6
ET8	Neutral	Positive	Negative	Yes	1	5	-4
ET9	Neutral	Negative	Neutral	Yes	9	16	-7
ET10	Negative	Negative	Negative	No	3	12	-9
ET11	Negative	Positive	Neutral	Yes	18	12	6
ET12	Positive	Neutral	Neutral	Yes	5	7	-2
ET13	Neutral	Neutral	Negative	No	5	31	-26
ET14	Neutral	Neutral	Negative	No	2	14	-12
ET15	Neutral	Positive	Positive	Yes	9	4	5
ET16	Neutral	Neutral	Neutral	Yes	4	7	-3
ET17	Neutral	Negative	Negative	No	5	11	-6
ET18	Neutral	Negative	Negative	No	2	7	-5
ET19	Neutral	Negative	Negative	No	2	5	-3
ET20	Neutral	Negative	Negative	No	0	4	-4
ET21	Positive	Neutral	Negative	Yes	5	13	-8
ET22	Positive	Negative	Neutral	No	3	5	-2
ET23	Positive	Negative	Neutral	No	6	9	-3
ET24	Positive	Positive	Neutral	Yes	3	4	-1
ET25	Positive	Neutral	Positive	Yes	35	4	31
ET26	Positive	Negative	Positive	Yes	26	4	22
ET27	Positive	Neutral	Positive	No	19	10	9
ET28	Positive	Positive	Positive	Yes	24	4	20
ET29	Neutral	Negative	Negative	No	2	8	-6
ET30	Positive	Neutral	Negative	No	0	4	-4
ET31	Positive	Negative	Neutral	No	12	11	1
ET32	Negative	Positive	Positive	Yes	2	0	2
ET33	Positive	Negative	Neutral	No	23	24	-1
ET34	Positive	Negative	Negative	No	4	9	-5
ET35	Neutral	Negative	Negative	No	1	9	-8
ET36	Negative	Positive	Neutral	Yes	3	5	-2

Table 6.15: Affect evaluation data of second analysis.

only "very likely" results. This hugely reduced the false-positive and false-negative affect identified in the video. This is shown in Figure 6.39.

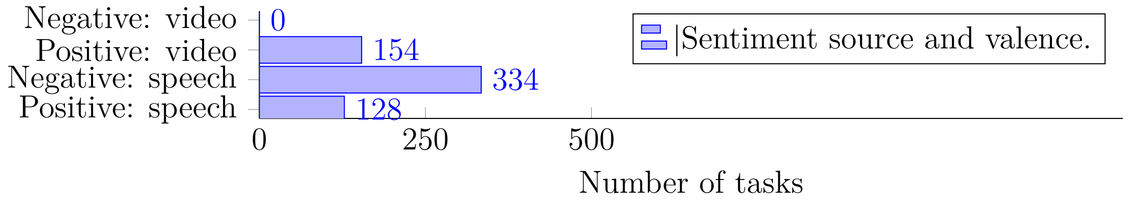

Figure 6.39: Positive and negative sentiments identified in video and speech of second analysis.

Instead of 1422 video sentiments identified, we observe only 154. As the affect analysis counts the number of positive and negative sentiments identified from audio and video, it results in a gain of consistency in evaluated affect.

One crucial point is that this procedure counts the positive and negative affects identified by Google's IA engine. In this way, it represents all the sentiments demonstrated by the user during the test. Attracts attention to the mismatch between the final emotional state and the affect identified. Remembering Chapter 2, we would expect the user users to see the positive and negative peaks and the final moments in the task to evaluate their own final emotions.

So, it's likely that tasks with great differences between positive and negative affects identified have the user's final emotional state as the affect identified by UXAPP. We identified and analyzed the tasks with more than 10 points of difference: ET1, ET2, ET3, ET5, ET13, ET14, ET25, ET26, and ET28. At first, we looked for a relation between the final emotional state and the affect identified. Figure 6.40 shows the match between these variables. We observed the same proportion of matches and mismatched from the result with all tasks, but we expected all or almost all results to match. Then, we realized that the affect identified corresponded with the task success or unsuccess, as shown in Figure 6.41.

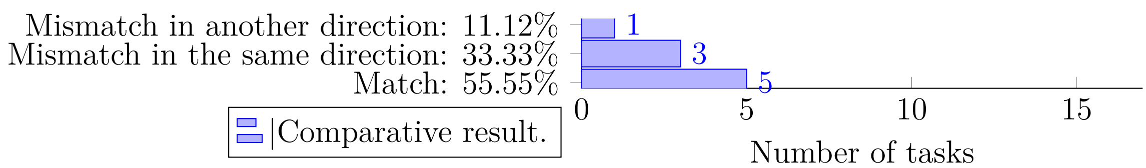

Figure 6.40: Emotional final state vs Affect identified for tasks with great differences between positive and negative affect.

This analysis confirmed the affect element can drive the UX experience. The only mismatch in this analysis is the task ET2. ET26 matches but with inconsistency. Tasks

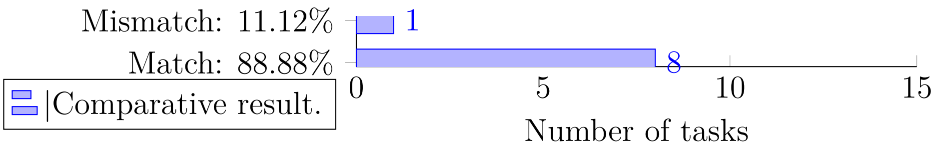

Figure 6.41: Task result vs Affect identified for tasks with great differences between positive and negative affect.

ET2 and ET26 are of type 2, in which the user must request an invitation to use UXAPP. As we said before, the task of type 2 had an intentional issue in the e-mail validation that prevented the user from finishing the task. So, the only possible result for this task is unsuccess. Despite this, the users of these two tasks' IDs marked it successful.

Reviewing these videos, we observed in ET2 that the user had a bad experience trying to fill out the problematic newsletter form and could not send the request. As a result, the objective of the task was not achieved. In this case, the affect is consistent (negative), and the task success informed is inconsistent (positive).

In ET26, we have a different situation. The task was marked as successful, and the UXAPP affect identified was positive. In the video, the user didn't finish the task and expressed a sentiment of frustration. Despite that, the user presented a positive facial expression at this moment and during the task execution. This expression, however, didn't represent joy, and it can be interpreted as a way to be sympathetic while saying something negative about the digital product. In this case, the actual experience is negative, and the task was unsuccessful. So, both the task success and the affect identified are inconsistent.

We also can expect unsuccessful tasks to have a negative final emotional state despite the affect identified. We have 19 tasks with unsuccess marked. The distribution of the final emotional state for these tasks is presented in Figure 6.42. To enable the comparison, we show the same distribution for affect identified in Figure 6.43.

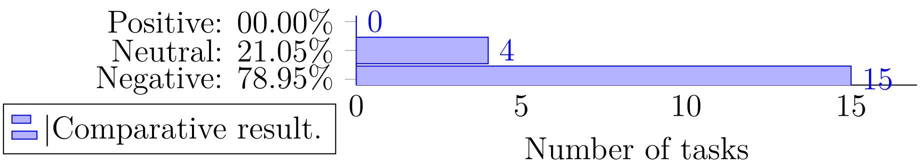

Figure 6.42: Distribution of final emotional state in second analysis for unsuccess tasks.

Despite the similar results, the neutral final emotional states didn't match with the neutral affect identified, so there was a mismatch in the same direction in these cases. Indeed, most users set their final emotional state as negative despite the affect identified when the task is unsuccessful. We observe, however, that some users stood in a neutral

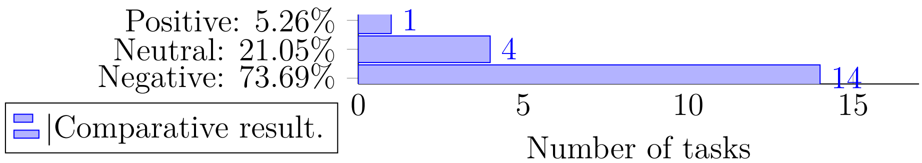

Figure 6.43: Distribution of affect identified in second analysis for unsuccess tasks.

emotional state. If we consider that as the user's emotional response and the affect identified as the noticeable user's answer to the experience, it means the user's emotional state can be dissociated from the final noticeable user experience. Even though we observed users having bad experiences, the user still can feel neutral or even happy. Most users repeat their initial emotional state in all tasks despite failure or success in the tasks, as shown in Figure 6.44.

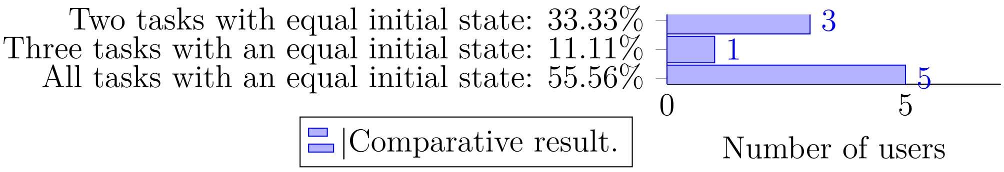

Figure 6.44: Distribution of the number of users that repeat an initial emotional state.

To illustrate, one user expressed an optimistic outlook on life and a desire to maintain it during testing. Another user had a contusion while playing their favorite sport and was not emotionally fine. The other one stays neutral in all interactions, describing punctually their emotional state when success or unsuccess happens.

After this analysis, we can state that the user's emotional state is related to the user experience of a digital product immediately before, during, and immediately after using the product. Still, the emotional state is a more profound and more complex concept than the noticeable user experience of a digital product.

Emotional Satisfaction Evaluation for the Second Analysis

We tabulate the results of the user value evaluation for the second analysis. We check the percentage of tasks with a match between UXAPP and the user's answer. The user value data is presented in Table 6.16. We can see the match percentage in Figure 6.45. We consider the "same direction" when one value is "Neither satisfied nor dissatisfied" and the other "Satisfied" or "Dissatisfied", and the "different direction" when one value is "Satisfied" and the other "Dissatisfied". As we reduce the 5-point scale to a 3-point scale, we considered users' answers with the values "Very satisfied" and "Very unsatisfied" as "Satisfied" and "Dissatisfied", respectively.

ID	Manual Geral Satisfaction	User Value Identified
ET1	Satisfied	Satisfied
ET2	Dissatisfied	Neither satisfied nor dissatisfied
ET3	Dissatisfied	Neither satisfied nor dissatisfied
ET4	Satisfied	Neither satisfied nor dissatisfied
ET5	Dissatisfied	Satisfied
ET6	Dissatisfied	Satisfied
ET7	Satisfied	Satisfied
ET8	Satisfied	Neither satisfied nor dissatisfied
ET9	Neither satisfied nor dissatisfied	Satisfied
ET10	Dissatisfied	Neither satisfied nor dissatisfied
ET11	Satisfied	Satisfied
ET12	Satisfied	Satisfied
ET13	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET14	Dissatisfied	Neither satisfied nor dissatisfied
ET15	Satisfied	Neither satisfied nor dissatisfied
ET16	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET17	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET18	Dissatisfied	Neither satisfied nor dissatisfied
ET19	Dissatisfied	Neither satisfied nor dissatisfied
ET20	Dissatisfied	Neither satisfied nor dissatisfied
ET21	Satisfied	Neither satisfied nor dissatisfied
ET22	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET23	Dissatisfied	Satisfied
ET24	Satisfied	Neither satisfied nor dissatisfied
ET25	Satisfied	Satisfied
ET26	Neither satisfied nor dissatisfied	Satisfied
ET27	Satisfied	Satisfied
ET28	Satisfied	Satisfied
ET29	Neither satisfied nor dissatisfied	Neither satisfied nor dissatisfied
ET30	Dissatisfied	Neither satisfied nor dissatisfied
ET31	Dissatisfied	Satisfied
ET32	Satisfied	Neither satisfied nor dissatisfied
ET33	Dissatisfied	Satisfied
ET34	Neither satisfied nor dissatisfied	Satisfied
ET35	Dissatisfied	Neither satisfied nor dissatisfied
ET36	Satisfied	Satisfied

Table 6.16: User value evaluation data for the second analysis.

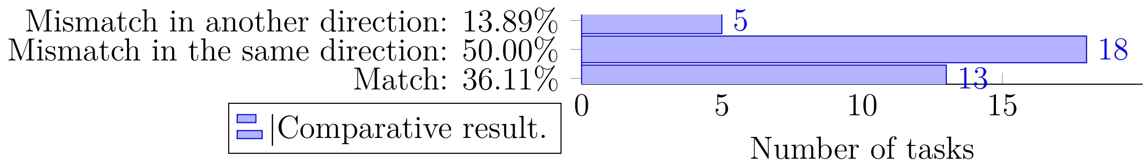

Figure 6.45: Distribution of user value evaluation match for the second analysis.

We observe the same movement occurred in Subsection 6.2.2. The reduction of scale and adjustment in emotion intensity interpretation increases the “match” and the “mismatch in the same direction” results. It represents a gain in the consistency of user value evaluation. This analysis confirmed the user value element can drive the UX experience.

Similarly to affect, we can expect unsuccessful tasks to be dissatisfied as user value despite the identified affect. The distribution of user value for unsuccess tasks is presented in Figure 6.46. To enable our comparison, we show the same distribution for user value identified in Figure 6.47.

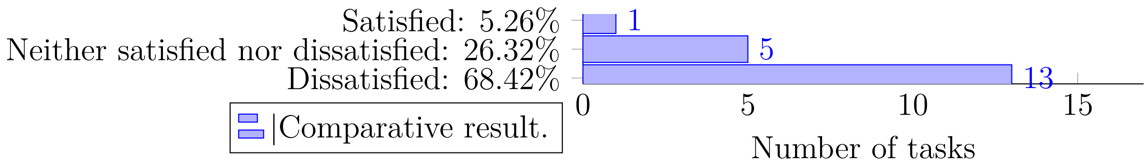

Figure 6.46: Distribution of overall user value in second analysis for unsuccess tasks.

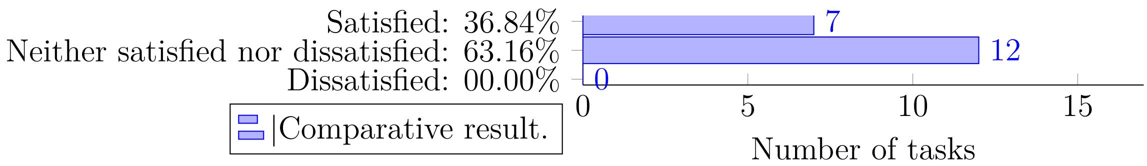

Figure 6.47: Distribution of user value identified in second analysis for unsuccess tasks.

In the same way as affect evaluation, the user value dissatisfaction with unsuccessful tasks was independent of the user value identified. The “satisfied” result attracts attention. In this case, the task required to obtain a usability test script in German. The user marked their initial state as positive and their final state as neutral. The affect identified for this task was positive. Watching the video, the user stated there was no way to achieve the task objective. We can interpret that the user was satisfied with their perception the task was not achievable.

UXAPP identified no negative emotion about the user value evaluation, which explains the satisfied results. Again, we observe a mismatch between the noticeable user experience of digital products and the internal user emotion represented by the user’s satisfaction.

Considering all the second analyses over user value, we can state there is a relationship between user satisfaction and the noticeable user experience of digital products. However, the concept of user satisfaction is deeper and more complex than the noticeable user experience of digital products.

User Experience Evaluation for the Second Analysis

We tabulate the results of the user experience evaluation for the second analysis. The data is presented in Table 6.17. We can see the distribution of the values in Figure 6.48.

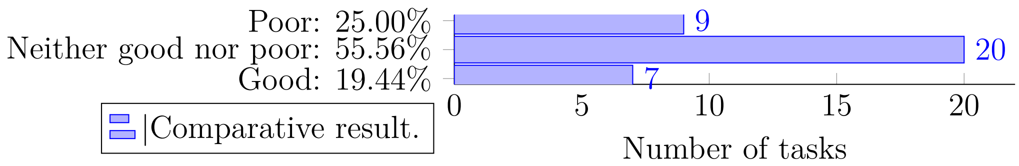

Figure 6.48: Distribution of user value evaluation match for the second analysis.

As we can see, there was a reduction of the "Good" and "Poor" percentages in favor of "Neither good nor poor". This happens due to adjustments in emotion intensity interpretation from the videos.

We tabulate the user manual evaluation for the second analysis and calculate the UX evaluation based on the UXAPP approach with the user input data. This data is presented in Table 6.18. The UX evaluation is shown in Figure 6.49.

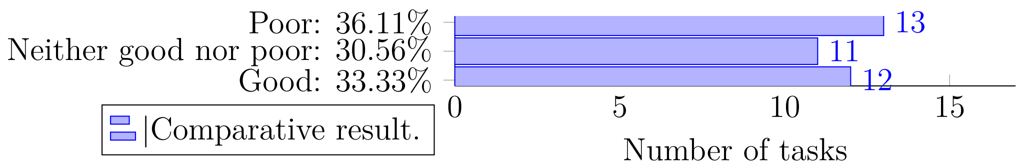

Figure 6.49: Distribution of UX evaluation with user input data for the second analysis.

The user registered a more equal distribution of the evaluation than the UXAPP evaluation. It means the user showed less emotion than they felt. However, the UXAPP evaluation captured the essence of the user evaluation, with a "Poor" percentage a little more significant than the "Good" one.

The comparison between the user experience obtained from UXAPP analysis and user input data is shown in Table 6.19 and Figure 6.50.

The UXAPP analysis obtained a 50% direct match with the user answer. Due to the issues related to the noticeable user experience exposed throughout this work, we see this result in an optimistic way. It's also a good result of the value of 2.78% in the mismatch in another direction, which could lead to a contradictory interpretation of a usability test.

ID	User Experience Evaluation
ET1	Good
ET2	Neither good nor poor
ET3	Poor
ET4	Good
ET5	Neither good nor poor
ET6	Neither good nor poor
ET7	Good
ET8	Neither good nor poor
ET9	Good
ET10	Poor
ET11	Neither good nor poor
ET12	Neither good nor poor
ET13	Poor
ET14	Poor
ET15	Neither good nor poor
ET16	Neither good nor poor
ET17	Neither good nor poor
ET18	Poor
ET19	Poor
ET20	Poor
ET21	Poor
ET22	Neither good nor poor
ET23	Neither good nor poor
ET24	Neither good nor poor
ET25	Good
ET26	Neither good nor poor
ET27	Neither good nor poor
ET28	Good
ET29	Neither good nor poor
ET30	Poor
ET31	Neither good nor poor
ET32	Good
ET33	Neither good nor poor
ET34	Neither good nor poor
ET35	Neither good nor poor
ET36	Neither good nor poor

Table 6.17: User experience evaluation data for the second analysis.

ID	Usability Informed	Final Emotional State	Overall Satisfaction
ET1	Good	Positive	Satisfied
ET2	Neutral	Negative	Dissatisfied
ET3	Poor	Negative	Dissatisfied
ET4	Good	Neutral	Satisfied
ET5	Poor	Negative	Dissatisfied
ET6	Poor	Negative	Dissatisfied
ET7	Good	Positive	Satisfied
ET8	Good	Positive	Satisfied
ET9	Poor	Negative	Neither satisfied nor dissatisfied
ET10	Neutral	Negative	Dissatisfied
ET11	Good	Positive	Satisfied
ET12	Good	Neutral	Satisfied
ET13	Neutral	Neutral	Neither satisfied nor dissatisfied
ET14	Neutral	Neutral	Dissatisfied
ET15	Good	Positive	Satisfied
ET16	Good	Neutral	Neither satisfied nor dissatisfied
ET17	Neutral	Negative	Neither satisfied nor dissatisfied
ET18	Neutral	Negative	Dissatisfied
ET19	Neutral	Negative	Dissatisfied
ET20	Neutral	Negative	Dissatisfied
ET21	Neutral	Neutral	Satisfied
ET22	Neutral	Negative	Neither satisfied nor dissatisfied
ET23	Poor	Negative	Dissatisfied
ET24	Good	Positive	Satisfied
ET25	Good	Neutral	Satisfied
ET26	Neutral	Negative	Neither satisfied nor dissatisfied
ET27	Neutral	Neutral	Satisfied
ET28	Neutral	Positive	Satisfied
ET29	Neutral	Negative	Neither satisfied nor dissatisfied
ET30	Neutral	Neutral	Dissatisfied
ET31	Poor	Negative	Dissatisfied
ET32	Good	Positive	Satisfied
ET33	Neutral	Negative	Dissatisfied
ET34	Neutral	Negative	Neither satisfied nor dissatisfied
ET35	Poor	Negative	Dissatisfied
ET36	Good	Positive	Satisfied

Table 6.18: User input data used to calculate UX evaluation for the second analysis.

ID	User Experience Evaluation - UXAPP	User Experience Evaluation - User
ET1	Good	Good
ET2	Neither good nor poor	Poor
ET3	Poor	Poor
ET4	Good	Good
ET5	Neither good nor poor	Poor
ET6	Neither good nor poor	Poor
ET7	Good	Good
ET8	Neither good nor poor	Good
ET9	Good	Poor
ET10	Poor	Poor
ET11	Neither good nor poor	Good
ET12	Neither good nor poor	Good
ET13	Poor	Neither good nor poor
ET14	Poor	Neither good nor poor
ET15	Neither good nor poor	Good
ET16	Neither good nor poor	Neither good nor poor
ET17	Neither good nor poor	Neither good nor poor
ET18	Poor	Poor
ET19	Poor	Poor
ET20	Poor	Poor
ET21	Poor	Neither good nor poor
ET22	Neither good nor poor	Neither good nor poor
ET23	Neither good nor poor	Poor
ET24	Neither good nor poor	Good
ET25	Good	Good
ET26	Neither good nor poor	Neither good nor poor
ET27	Neither good nor poor	Neither good nor poor
ET28	Good	Good
ET29	Neither good nor poor	Neither good nor poor
ET30	Poor	Neither good nor poor
ET31	Neither good nor poor	Poor
ET32	Good	Good
ET33	Neither good nor poor	Poor
ET34	Neither good nor poor	Neither good nor poor
ET35	Neither good nor poor	Poor
ET36	Neither good nor poor	Good

Table 6.19: Comparative results of user experience data for the second analysis.

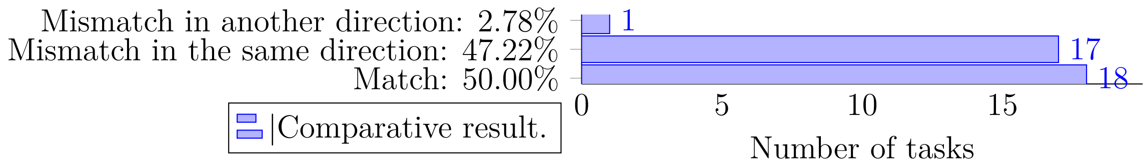

Figure 6.50: Distribution of UX evaluation match for the second analysis.

We can now state that UX evaluation from user input data differs from UXAPP evaluation. While the first one connects with the complexity of the more profound thoughts and feelings of an individual, the second one measures the noticeable user experience. In an analogy with an iceberg, the first is the submerged part of the iceberg, and the second is the visible part. Both are important and complementary.

Again, all the data before allows us to conclude the elements' evaluations determine the right direction in usability, affect, and satisfaction. Each UX element contributes in your way to the final evaluation, and the sum of their contribution makes the final UX evaluation score more balanced.

UXAPP Satisfaction Assessment

One week after the end of the tests, users were asked to answer anonymously about their satisfaction with UXAPP. We did ask two questions:

- Question 1 - How satisfied are you with using UXAPP to help test the site? (only rate the APP, not the website): the form made available a 5-point scale from "Very unsatisfied" to "Very satisfied". This answer was required.
- Question 2 - Why did you give this rating: this question made a short answer field available. This answer was required.

About question 1, we received the answers presented in Figure 6.51

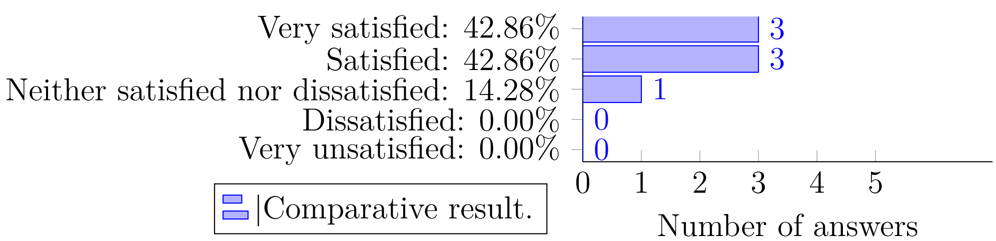

Figure 6.51: Distribution of user satisfaction score with UXAPP.

About question 2, the answers of the users are presented in Table 6.20.

Comment	Valence
Covers the main needs for testing the product	Positive
The system is handy for evaluating the website.	Positive
I think it gives a good idea for carrying out usability testing, and the combination with the SUS questionnaire is really cool! But it is not very clear, at first glance, what action should be taken at each moment. After using it for a couple of tests, it becomes clearer.	Positive and negative
Addressed several test questions.	Neutral
I found it interesting and innovative.	Positive
As a result of helping to improve the application	Positive
I couldn't use it.	Negative

Table 6.20: Comments of user satisfaction score with UXAPP.

6.3 Chapter Summary

The usability test with nine participants allowed us to validate the UXAPP model and adjust parameters to simplify the model and make it more consistent. Each UX element was analyzed separately, as well as the UX evaluation itself.

In a comparative analysis, UXAPP presented a similar result with the user input data, but these two metrics are different and didn't work in the same way. We observe that the UXAPP model can evaluate the noticeable user experience and can be complemented with manual evaluation performed by the user.

Chapter 7

Conclusions

Digital transformation has enabled the creation of new business models to generate more value for companies. For consumers, access to new technologies has increased their expectations regarding the service, including digital transactions. Using better technologies made it possible to extract usage data that boosted the services provided. [1], [13], [1].

The measurement of organizational performance began to be carried out from the point of view of demand, considering aspects of quality and user satisfaction. The primary metric for measuring performance in a non-financial way is measuring user satisfaction. User satisfaction plays a relevant role in identifying consumption aspects, such as future purchase intention, consumer loyalty, and end-user retention [14] [3] [39] [57].

Several aspects, however, can impact the measurement of user satisfaction. Among them, it was observed that user satisfaction is subjective and affected by cultural elements, is limited by technical barriers to product use, and can change over time. The use of traditional self-assessment methods to measure satisfaction can lead to problems related to subjectivity, the unwillingness of participants to respond, or biases of various types in responding [15] [57].

User satisfaction is an attribute of user experience (UX) composed of usability, affect, and user value elements [4] [5]. To measure UX experience, we need to evaluate all its components. This work sought to understand whether it could automatically measure user experience by identifying a digital product's positive, neutral, and negative points. The recognition of emotions based on the recognition of facial expressions would be an alternative to this form of measurement.

Therefore, a systematic literature mapping was carried out to identify the state of the art, obtain an overview, and discover trends related to the recognition of emotions through facial expressions using Artificial Intelligence (AI). The selected studies differed by type: two were systematic literature reviews, three were experiments in a spontaneous environment, and five were in a controlled environment. The questions related to mapping asked which studies dealt with visual emotion recognition and Artificial Intelligence (AI), what the techniques for carrying out this type of emotion recognition were, and what the benefits and challenges related to this approach were.

Based on the information identified in the systematic literature mapping, a UX evaluation model was proposed to perform emotion recognition using Artificial Intelligence. It was decided to integrate existing *Google* solutions for the end user and in the cloud, which carried out the responsibilities of components frontend, backend, and AI, with the execution of a UXAPP engine to analyze the AI results.

The application receives the user's camera and screen, as well as the input data into UXAPP. The application delivers a video with instant satisfaction and affect evaluations throughout the video, as well as a UX report with the UX evaluation score, all the positive and negative points with sentiment detection, and additional information to help the user understand the experience.

An experiment in a controlled environment was carried out. This approach is based on a usability test with nine participants and four tasks per participant. The test is oriented towards talking about actions and feelings. A manual user experience assessment was carried out at the end of the experiment. After the test, UXAPP analyzed the task data to elaborate the UX report. For an average duration of 9m 19s, UXAPP analyzed the task for an average time of 24m 37s.

The UXAPP evaluation gave us a consistent and balanced result related to the user input data result, but it can't be confused with the evaluation carried out by the user.

The UXAPP UX evaluation directly matches the user's evaluation in 50% of the tasks and gives a close answer in others 47.22% of the results. When comparing the user input data analysis and the UXAPP analysis, we verified that the user showed less emotion than they felt. We stated that UX evaluation from user input data differs from UXAPP evaluation. While the first one connects with the complexity of the more profound thoughts and feelings of an individual, the second one measures the noticeable user experience. In an analogy with an iceberg, the first is the submerged part of the iceberg, and the second is the visible part. Both are important and complementary.

Our work identified several issues in the UX experience evaluation. First, we identified external barriers that impacted the user experience, like bad internet connection, which prevented the site from being fully loaded, lack of knowledge about the idiom of the tested site, and technical language to describe the tested site functionalities. Second, we observe the noticeable experience is not necessarily what the user registered in the task. Some users demonstrated emotions with a valence and registered an emotional state with the opposite valence. Third, even after some users had a bad experience, we observe they didn't change their initial emotional state, which means the user's emotional state differs from the noticeable user experience using a digital product.

The UXAPP usability analysis matches the user's evaluation in 30.56% of the tasks and gives a close answer in others 33.33% of the results. We observed that the SUS approach consumes much time registering the user evaluation. The UXAPP usability analysis process is very complex due to the need to identify SUS sentences' occurrences with similarity and the same valence of the user's speech. A better approach could be using large language models (LLMs) or AI generative multimodal models to determine the usability of directly interpreting the user context.

The UXAPP affect analysis matches the user's evaluation in 52.78% of the tasks and gives a close answer in others 41.67% of the results. UXAPP also identified 616 points of positive or negative sentiments in video and speech for all tasks. We observed the identified affect corresponded with the task result for tasks with significant differences between positive and negative affect. It means the affect identified drives the noticeable user experience correctly. We stated that the user's emotional state is related to the user experience of a digital product immediately before, during, and immediately after using the product. Still, the emotional state is a more profound and more complex concept than the noticeable user experience of a digital product.

The UXAPP user value analysis matches the user's evaluation in 36.11% of the tasks and gives a close answer in 50.00% of the results. This result behaves similarly to the affect evaluation. We observed a mismatch between the noticeable user experience of digital products and the internal user emotion represented by the user's satisfaction. We stated there is a relationship between user satisfaction and the noticeable user experience of digital products. However, the concept of user satisfaction is deeper and more complex than the noticeable user experience of digital products.

We registered 5h 35m 31s of tasks' duration. Based on our experience, we estimate that a person needed at least four times this task duration to process all the information presented in the UX report, including finding all sentiment points and distinguishing positive and negative sentiment peaks.

A week after the usability test, the users assessed their satisfaction with UXAPP using a scale from 1 (very unsatisfied) to 5 (very satisfied). The weighted arithmetic mean of the UXAPP satisfaction assessment was 4.29 on a scale of 5 points.

The UX evaluation process and experiments worked to extract the necessary data for the experiment. The results of UXAPP demonstrate that the UX evaluation model can drive the user experience.

Two main difficulties were identified during the work. The first relates to obtaining the correct emotional input from the user. To UXAPP analysis works as expected, as we saw, we needed the users to express their emotions with a high degree of fidelity. We believe factors like cultural questions, the presence of other persons with the user, the moment the user carries out the test, and the user's expectation of what the organizer needs to obtain can impact the noticeable emotions. UXAPP didn't intend to map the user's emotional states and their more profound experience. Still, it has to obtain useful improvement points of the app and determine the direction of a digital product user experience improvement.

Another difficulty observed is to adjust the application parameters to get a consistent result. We realize two rounds of analysis with some parameter variations to reach the last results. Each analysis has parameters like confidence degree, similarity degree, and emotion intensity, and it's necessary to map these parameters to the expected result in the application. If you set a higher confidence degree, e.g., you may lose essential results; if you define a lower degree, you may include invalid results in your analysis.

This work provides an essential benefit of mapping the state of the art related to studies measuring user satisfaction using emotion recognition through facial expressions. The availability of this mapping allows new studies to be proposed to identify a lack of technologies, algorithms, and studies about the need for applying these resources in new areas or areas still not met by them.

Besides, this work delivers a tested model to evaluate user experience from emotional recognition. The UXAPP model allows us to scale the user experience evaluation of digital products to a new degree and obtain continuous feedback to improve the development of new digital products and services.

To expand the work, we suggested that new systematic mappings of literature about emotion recognition in other contexts, such as the medical field or the public security area, be elaborated. Another interesting possibility is expanding the evaluation of emotion recognition solutions beyond the recognition of facial expressions and speech sentiments with the use of voice tone, mouse, keyboard, gaze mapping, and other ways, which tends to increase the degree of accuracy of the model.

We expect the user experience measurement procedures will be facilitated with the use of emotion recognition tools, and this makes it possible to identify the main points of product improvement and friction in digital services so that these products and services can be evolved to deliver more value to service providers and society.

7.1 Future Work

Future work related to this study focuses on studying new approaches to evaluate UX elements and understanding how to aggregate new emotion recognition models to obtain a more sophisticated and consistent user experience measurement result.

Another exciting possibility is applying the UX evaluation model in other contexts like medical or virtual environments, like metaverse or augmented realities, which can help develop new areas of knowledge by unifying the resolution of user problems and needs with a technological and innovative approach.

Another interesting future work is developing a solution that measures user experience more naturally, automatically captures the user's video and the product usage screen after their consent and transparently handles this for the user. This would allow the application to reach a larger audience of users interested in checking the user experience of digital products.

The benefits of using emotion recognition tools found in this work are related to the automatic obtaining of feedback from the user, such as mapping the user profile, identifying purchasing behavior, optimizing store space, and improving consumer relations. Challenges were also encountered, such as protecting the user's and their data's privacy, the failure to recognize emotions for cultural reasons, and the failure to capture images due to environmental inadequacies.

References

[1] Verhoef, Peter C., Thijs Broekhuizen, Yakov Bart, Abhi Bhattacharya, John Qi Dong, Nicolai Fabian, and Michael Haenlein: *Digital transformation: A multidisciplinary reflection and research agenda.* Journal of Business Research, 122:889–901, January 2021, ISSN 01482963. viii, 1, 124

[2] Favoretto, Camila, Glauco Henrique de Sousa Mendes, Moacir Godinho Filho, Maicon Gouvea de Oliveira, and Gilberto Miller Devós Ganga: *Digital transformation of business model in manufacturing companies: challenges and research agenda*, February 2022. ISSN 08858624. viii, 1

[3] Birch-Jensen, A., I. Gremyr, J. Hallencreutz, and Rönnbäck: *Use of customer satisfaction measurements to drive improvements.* Total Quality Management and Business Excellence, 31(5-6):569–582, April 2020, ISSN 14783371. viii, 1, 4, 124

[4] Park, Jaehyun, Sung H. Han, Hyun K. Kim, Youngseok Cho, and Wonkyu Park: *Developing elements of user experience for mobile phones and services: Survey, interview, and observation approaches*, July 2013. ISSN 10908471. viii, xvii, 2, 19, 20, 23, 29, 51, 124

[5] International Organization for Standardization (ISO): *Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts.* Technical report, International Organization for Standardization (ISO), 2018. viii, 2, 124

[6] Silva Franco, R.Y. da, R.S. do Amor Divino Lima, R. do Monte Paixão, C.G.R. dos Santos, and B.S. Meiguins: *UXmood-A sentiment analysis and information visualization tool to support the evaluation of usability and user experience.* Information (Switzerland), 10(12), 2019. xvii, 12, 13

[7] Desmet, Pieter and Paul Hekkert: *Framework of product experience.* International Journal of Design, 1(1):57 – 66, 2007. `https://www.scopus.com/inward/record.uri?eid=2-s2.0-70350501467&partnerID=40&md5=815e17103a2bf4bf2d816dc796e35bb9`, Cited by: 824. xvii, 21

[8] Landowska, Agnieszka: *Towards emotion acquisition in IT usability evaluation context.* In *ACM International Conference Proceeding Series*, volume 29-30-Jun-2015. Association for Computing Machinery, June 2015, ISBN 9781450336017. xvii, xx, 25, 26, 27, 29, 39, 41, 46

[9] Li, L.: *Study on the interactive relationship between customer's emotional response and the brand trust - In the view of online shopping*. In *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2013*, pages 245–248, 2013, ISBN 9781479905300. xx, 2

[10] Balbin, Jessie R., Charmaine C. Paglinawan, Mary Josanne A. De Castro, Jared Kobe C. Llamas, Mikka Ellah T. Medina, John Jomel O. Pangilinan, and Flordeliza L. Valiente: *Augmented Reality Aided Analysis of Customer Satisfaction based on Taste-Induced Facial Expression Recognition Using Affdex Software Developer's Kit*. In *ACM International Conference Proceeding Series*, pages 204–209. Association for Computing Machinery, March 2019, ISBN 9781450361309. xx, 24, 27, 29, 39, 43, 59

[11] Chimienti, Michela, Ivan Danzi, Vincenzo Gattulli, Donato Impedovo, Giuseppe Pirlo, and Davide Veneto: *Behavioral Analysis for User Satisfaction*. In *Proceedings - 2022 IEEE 8th International Conference on Multimedia Big Data, BigMM 2022*, pages 113–119. Institute of Electrical and Electronics Engineers Inc., 2022, ISBN 9781665459631. xx, 24, 27, 29, 39, 43, 47

[12] Soleymani, Mohammad, David Garcia, Brendan Jou, Björn Schuller, Shih Fu Chang, and Maja Pantic: *A survey of multimodal sentiment analysis*. Image and Vision Computing, 65:3–14, September 2017, ISSN 02628856. xx, 24, 25, 27, 29, 39, 43, 48, 49

[13] Vial, Gregory: *Understanding digital transformation: A review and a research agenda*, June 2019. ISSN 09638687. 1, 124

[14] Bititci, U., P. Garengo, V. Dörfler, and S. Nudurupati: *Performance Measurement: Challenges for Tomorrow*. International Journal of Management Reviews, 14(3):305–327, 2012. 1, 124

[15] Alawneh, Ali, Hasan Al-Refai, and Khaldoun Batiha: *Measuring user satisfaction from e-Government services: Lessons from Jordan*. Government Information Quarterly, 30(3):277–288, July 2013, ISSN 0740624X. 2, 124

[16] Cowie, R., E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor: *Emotion recognition in human-computer interaction*. IEEE Signal Processing Magazine, 18(1):32–80, 2001. 3

[17] Kasiran, Z. and S. Yahya: *Facial expression as an implicit customers' feedback and the challenges*. In *Computer Graphics, Imaging and Visualisation: New Advances, CGIV 2007*, pages 377–381, 2007, ISBN 9780769529288. 3

[18] Hallencreutz, Jacob and Johan Parmler: *Important drivers for customer satisfaction–from product focus to image and service quality*. Total Quality Management and Business Excellence, 32(5-6):501–510, 2021, ISSN 14783371. 3, 4

[19] Tian, X., W. Hou, and K. Yuan: *A study on the method of satisfaction measurement based on emotion space*. In *9th International Conference on Computer-Aided Industrial Design and Conceptual Design: Multicultural Creation and Design - CAIDCD 2008*, pages 39–43, 2008, ISBN 9781424432912. 4

[20] Picard, R.W.: *Affective Computing: From laughter to IEEE.* IEEE Transactions on Affective Computing, 1(1):11–17, 2010. 4

[21] Anderson, Eugene W, Claes Fornell, and Donald R Lehmann: *Customer Satisfaction, Market Share, and Profitability: Findings from Sweden.* Journal of Marketing, 58(3):53–66, 1994. https://doi.org/10.1177/002224299405800304. 5

[22] Erven, Ricardo Cordeiro Galv
btxfnamespacelong ao Sant'Ana van and Edna Dias Canedo: *Measurement of user's satisfaction of digital products through emotion recognition.* In *Proceedings of the XXII Brazilian Symposium on Software Quality*, SBQS '23, page 62–71, New York, NY, USA, 2023. Association for Computing Machinery, ISBN 9798400707865. https://doi.org/10.1145/3629479.3629488. 6

[23] Erven, Ricardo Cordeiro Galvão Sant' Ana van, Demétrius de Almeida Jubé, Helen Reis Santos, Sergio Antonio Andrade de Freitas, and Edna Dias Canedo: *Gamification project to receive continuous feedback in the context of the evolution of public service for lawyers.* In *2023 IEEE Frontiers in Education Conference (FIE)*, pages 1–8, Oct 2023. 7

[24] Erven, Ricardo Cordeiro Galvão SantAna van, Pollyanna C. O. Dias, Demétrius de Almeida Jubé, George Marsicano Corrêa, and Edna Dias Canedo: *Avaliação de conhecimento em engenharia de requisitos no startup gov.br.* In Antonelli, Leandro, Márcia Lucena, and Roxana L. Q. Portugal (editors): *Anais do WER23 - Workshop em Engenharia de Requisitos, Porto Alegre, RS, Brasil, Agosto 15-17, 2022.* LFS (UFRN, Brasil), 2023. https://doi.org/10.29327/1298356.26-11. 7

[25] De Almeida Jubé, Demétrius, Célio Castro Wermelinger, Ricardo Cordeiro Galvão Sant'Ana Van Erven, and Fernando Narciso Bertolaccini De Souza E Edison Ishikawa: *Optimization of vehicle routing in the distribution of electronic voting machines for elections.* In *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, June 2023. 8

[26] Gahler, Markus, Jan F. Klein, and Michael Paul: *Customer Experience: Conceptualization, Measurement, and Application in Omnichannel Environments.* Journal of Service Research, May 2022, ISSN 15527379. 9

[27] Kujala, Sari, Virpi Roto, Kaisa Väänänen-Vainio-Mattila, Evangelos Karapanos, and Arto Sinnelä: *UX Curve: A method for evaluating long-term user experience.* Interacting with Computers, 23(5):473–483, 2011, ISSN 09535438. 10, 11, 59

[28] McCarthy, John, Peter Wright, Jayne Wallace, and Andy Dearden: *The experience of enchantment in human-computer interaction.* Personal and Ubiquitous Computing, 10(6):369–378, October 2006, ISSN 16174909. 10

[29] Karapanos, Evangelos, Marc Hassenzahl, and Jean Bernard Martens: *User Experience over Time.* In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3561–3566, New York, NY, USA, 2008. Association for Computing Machinery, ISBN 9781605580128. https://doi.org/10.1145/1358628.1358891. 10, 11

[30] Setchi, Rossitza and Obokhai K. Asikhia: *Exploring user experience with image schemas, sentiments, and semantics.* IEEE Transactions on Affective Computing, 10(2):182–195, April 2019, ISSN 19493045. 11, 12

[31] Likert, Rensis: *A technique for the measurement of attitudes.* Archives of psychology, 1932. 11, 22, 51, 55

[32] Roto, Virpi, Heli Väätäjä, and Tiina Koponen: *Developing practical tools for user experience evaluation: a case from mobile news journalism. Developing Practical Tools for User Experience Evaluation-A Case from Mobile News Journalism.* VTT Symposium (Valtion Teknillinen Tutkimuskeskus), 258:240 – 247, 2009. `https://www.scopus.com/inward/record.uri?eid=2-s2.0-72049112287&partnerID=40&md5=d864997536c88e4cb67af7884cd6a321`, Cited by: 9. 11

[33] Lachner, Florian, Philippd Naegelein, Robert Kowalski, Martin Spann, and Andreas Butz: *Quantified UX: Towards a common organizational understanding of user experience.* In *ACM International Conference Proceeding Series*, volume 23-27-October-2016. Association for Computing Machinery, October 2016, ISBN 9781450347631. 11

[34] Desolda, Giuseppe, Andrea Esposito, Rosa Lanzilotti, and Maria F Costabile: *Detecting Emotions Through Machine Learning for Automatic UX Evaluation.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12934 LNCS:270 – 279, 2021. 12

[35] Ntoa, Stavroula, George Margetis, Margherita Antona, and Constantine Stephanidis: *Uxami observer: An automated user experience evaluation tool for ambient intelligence environments.* Advances in Intelligent Systems and Computing, 868:1350 – 1370, 2018. 12

[36] Oliveira, S., A. Cristo, M. Geovane, A. Xavier, R. Silva, S. Rocha, L. Marques, G. Gomes, B. Gadelha, and T. Conte: *UXNator: A Tool for Recommending UX Evaluation Methods.* In *International Conference on Enterprise Information Systems, ICEIS - Proceedings*, volume 2, pages 336–343, 2023, ISBN 9789897586484. 12

[37] Alves, Rui, Pedro Valente, and Nuno Jardim Nunes: *The state of user experience evaluation practice.* In *Proceedings of the NordiCHI 2014: The 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 93–102. Association for Computing Machinery, October 2014, ISBN 1595930361. 12

[38] Bartikowski, Boris and Sylvie Llosa: *Customer satisfaction measurement: Comparing four methods of attribute categorisations.* Service Industries Journal, 24(4):67–82, July 2004, ISSN 02642069. 13

[39] Wilson, A.: *Attitudes towards customer satisfaction measurement in the retail sector.* International Journal of Market Research, 44(2):213–222, 2002. 13, 124

[40] Hackl, Peter and Anders H. Westlund: *On structural equation modelling for customer satisfaction measurement.* Total Quality Management, 11(4-6):820–825, 2000, ISSN 09544127. 13, 14

[41] Brooke, John: *Sus: A quick and dirty usability scale.* Usability Eval. Ind., 189, November 1995. https://www.researchgate.net/publication/228593520. 14, 22, 23, 29

[42] Bangor, Aaron, Philip Kortum, and James Miller: *Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale.* Journal of Usability Studies, 4:114–123, 2009. 14, 23

[43] Coelho, P.S. and S.P. Esteves: *The choice between a five-point and a ten-point scale in the framework of customer satisfaction measurement.* International Journal of Market Research, 49(3):313–339, 2007. 14

[44] Sengupta, K. and M. Zviran: *Measuring user satisfaction in an outsourcing environment.* IEEE Transactions on Engineering Management, 44(4):414–421, 1997. 14

[45] Bailey, James E and Sammy W Pearson: *DEVELOPMENT OF A TOOL FOR MEASURING AND ANALYZING COMPUTER USER SATISFACTION.* Management Science, 29(5):530 – 545, 1983. 14

[46] McNamara, Niamh: *The psychometric approach to user satisfaction measurement*, 2013. ISSN 09535438. 14

[47] Picard, R.W., E. Vyzas, and J. Healey: *Toward machine emotional intelligence: Analysis of affective physiological state.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(10):1175–1191, 2001. 14, 15

[48] Zeng, Zhihong, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang: *A survey of affect recognition methods: Audio, visual, and spontaneous expressions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1):39–58, 2009, ISSN 01628828. 15, 16, 17

[49] Fasel, B. and J. Luettin: *Automatic facial expression analysis: A survey.* Pattern Recognition, 36(1):259–275, 2003. 16, 17

[50] Alben, Lauralee: *Quality of experience: Defining the criteria for effective interaction design.* Interactions, 3(3):11–15, may 1996, ISSN 1072-5520. https://doi.org/10.1145/235008.235010. 20

[51] Kolakowska, A., A. Landowska, M. Szwoch, W. Szwoch, and M.R. Wrobel: *Emotion recognition and its application in software engineering.* In *2013 6th International Conference on Human System Interactions, HSI 2013*, pages 532–539, 2013, ISBN 9781467356374. 25

[52] Ekman, P.: *Are There Basic Emotions?* Psychological Review, 99(3):550–553, 1992. 25, 41, 44

[53] Petersen, Kai, Sairam Vakkalanka, and Ludwik Kuzniarz: *Guidelines for conducting systematic mapping studies in software engineering: An update.* Information and Software Technology, 64:1–18, August 2015, ISSN 09505849. 30, 34, 49

[54] Parsifal - Perform Systematic Literature Reviews: *Parsifal*, 2023. https://parsif.al/. 30, 49

[55] Petticrew, M. and H. Roberts: *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishing Ltd, 2008, ISBN 9781405121101. 32, 35

[56] Dybå, T., T. Dingsøyr, and G.K. Hanssen: *Applying systematic reviews to diverse study types: An experience report*. In *Proceedings - 1st International Symposium on Empirical Software Engineering and Measurement, ESEM 2007*, pages 225–234, 2007, ISBN 9780769528861. 34

[57] Ceccacci, Silvia, Andrea Generosi, Luca Giraldi, and Maura Mengoni: *Emotional Valence from Facial Expression as an Experience Audit Tool: An Empirical Study in the Context of Opera Performance*. Sensors, 23(5), March 2023, ISSN 14248220. 39, 41, 44, 124

[58] Agrawal, Sanket, Rucha Rangnekar, Aditya Das, Shantanu Gawde, and Sudhir Dhage: *Gauging Customer Interest Using Skeletal Tracking and Convolutional Neural Network*. IEEE, 2019, ISBN 9781538681589. 39, 41, 42, 44, 47

[59] Rai Jain, Priti, S. M.K. Quadri, and Muskan Lalit: *Recent trends in artificial intelligence for emotion detection using facial image analysis*. In *ACM International Conference Proceeding Series*, pages 18–36. Association for Computing Machinery, August 2021, ISBN 9781450389204. 39, 42, 44, 46

[60] Pantano, Eleonora, Charles Dennis, and Eleftherios Alamanos: *Retail Managers' Preparedness to Capture Customers' Emotions: A New Synergistic Framework to Exploit Unstructured Data with New Analytics*. British Journal of Management, 33(3):1179–1199, July 2022, ISSN 14678551. 39, 45, 47, 48

[61] Dampage, Udaya, D. A. Egodagamage, A. U. Waidyaratne, D. A.W. DIssanayaka, and A. G.N.M. Senarathne: *Spatial Augmented Reality Based Customer Satisfaction Enhancement and Monitoring System*. IEEE Access, 9:97990–98004, 2021, ISSN 21693536. 39, 42, 45, 47, 48, 71

[62] Lee, J. S. and D. H. Shin: *The relationship between human and smart TVs based on emotion recognition in HCI*, volume 8582 LNCS. Springer Verlag, 2014, ISBN 9783319091464. 39, 42, 45

[63] Martins, Ana Isabel, Ana Filipa Rosa, Alexandra Queirós, Anabela Silva, and Nelson Pacheco Rocha: *European Portuguese Validation of the System Usability Scale (SUS)*. In *Procedia Computer Science*, volume 67, pages 293–300. Elsevier B.V., 2015. 52, 53

[64] Borsci, Simone, Robert D Macredie, Julie Barnett, Jennifer Martin, Jasna Kuljis, and Terry Young: *Reviewing and Extending the Five-User Assumption: A Grounded Procedure for Interaction Evaluation*. ACM Trans. Comput.-Hum. Interact., 20(5), November 2013, ISSN 1073-0516. https://doi-org.ez54.periodicos.capes.gov.br/10.1145/2506210. 66

[65] Choi, Wansuk, Taeseok Choi, and Seoyoon Heo: *A Comparative Study of Automated Machine Learning Platforms for Exercise Anthropometry-Based Typology Analysis: Performance Evaluation of AWS SageMaker, GCP VertexAI, and MS Azure.* Bioengineering, 10(8), August 2023, ISSN 23065354. 69