

MAI4CAREU

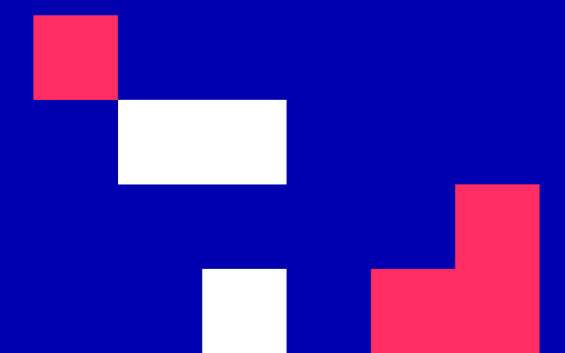
Master programmes in Artificial
Intelligence 4 Careers in Europe

University of Cyprus

HUMAN-CENTERED INTELLIGENT USER INTERFACES - MAI648

Marios Belk

2022



CONTENT 12

Evaluating Intelligent User Interfaces

CONTENTS

- Overview
 - Importance of Evaluation
 - Evaluation Types
 - Analytical Evaluation Techniques
 - Experimental Evaluation Techniques
- Exploratory Evaluation Techniques
 - Advantages and Disadvantages

CONTENT 12

Learning Outcomes

- Describe the main dimensions of user evaluation
- List the core characteristics of user evaluation techniques
- Explain the main methods for evaluating intelligent user interfaces
- Assess the relevance of certain evaluation techniques given a certain application domain

CONTENT 12

Overview

- User evaluation is the process of testing the design of a system and the final product
- Main purpose is to ensure the behavior expected by the designer and the satisfaction of user requirements.
- It should not be considered as a phase independent of the design.
- It should not be regarded as an activity that takes place only at the end of the process.

CONTENT 12

Overview

- System evaluation is a key process during the various phases of design and development
- Depending on the phase in which the evaluation is performed, we can distinguish two types:
 - **Formative evaluation:** Which aims to improve the system during the design-development phase
 - **Summative evaluation:** The quality of the final system / product is evaluated
- In the design of interactive systems and user interfaces in particular greater emphasis is given to formative evaluation

CONTENT 12

User Evaluation Goals

- Determining the scope of the system's functionality
 - User requirements must be met
 - The functionality must not only exist, but also be easily accessible to the user.
- Determine the impact of the interface on the user
 - How easy is the system to learn?
 - How is it in terms of usability?
 - What is the user's attitude towards the system?
- Identification of problems in the system

CONTENT 12

What do we Evaluate and When?

- The evaluation in iterative design examines:
 - The initial concept of the system and the corresponding mental model that users must form
 - Initial prototypes of the system (possibly of low fidelity)
 - More complete prototypes (this stage is repeated several times)

CONTENT 12

What do we Evaluate and When?

- From the system designers' point of view, the evaluation is aimed at testing:
 - Whether the user requirements have been correctly captured
 - Whether the design guidelines have been followed
 - Whether the user experience has been properly documented
 - Whether the designers' requirements have been fulfilled
 - Whether the user experience has been met
 - Alternative prototypes
 - The subjective user satisfaction

CONTENT 12

Why is User Evaluation Important?

- *“Iterative design, with its repeating cycle of design and testing, is the only validated methodology in existence that will consistently produce successful results. If you don’t have user-testing as an integral part of your design process you are going to throw buckets of money down the drain.” Bruce Tognazzini*

CONTENT 12

Evaluation Types

- Evaluation in controlled/in-lab settings
- Evaluation in real-life settings

CONTENT 12

Evaluation in Controlled/in-lab Settings

- It typically does not include end-users
- It is carried out by designers

CONTENT 12

Evaluation in Controlled/in-lab Settings

- **Advantages**
 - Includes special equipment
 - Special video/audio recording equipment
 - Specially configured computers
 - The environment is free from noise/interruptions
- **Disadvantages**
 - Abnormal situations
 - It is the only solution when the installation is dangerous to humans or in a remote location

CONTENT 12

Evaluation in Real-life Settings

- The designer visits the real user environment and monitors the system in practice.

CONTENT 12

Evaluation in Real-life Settings

- **Advantages**
 - Ability to observe interaction phenomena between systems and between users
 - Ability to monitor tasks that take a long time (e.g. months) to complete
 - Behaviors not seen in laboratory conditions are revealed
- **Disadvantages**
 - High level of noise, mobility, interruptions in the environment

CONTENT 12

Evaluation Techniques

- **Analytical techniques:** in the lab without user participation - evaluation by experienced scientists in the field (expert reviews)
 - Cognitive Walkthrough
 - Keystroke Level Model analysis
 - Heuristic Evaluation

CONTENT 12

Evaluation Techniques

- **Experimental techniques:** in the lab with users
 - Usability testing
 - Think aloud protocol
 - User logging

CONTENT 12

Evaluation Techniques

- **Exploratory techniques:** in the field with users
 - User observation
 - Field studies (multiple users)
 - Semi-structured interviews
 - Completion of questionnaires

CONTENT 12

Analytical Techniques

CONTENT 12

Analytical Techniques

- They are also known as predictive models
- They are mainly performed by experienced scientists in the field (experts in usability engineering)
- They provide an effective way for quantitative and qualitative evaluation of interactive systems without direct user involvement
- Various psychological, cognitive and kinetic models of users are used to evaluate the design

CONTENT 12

Analytical Techniques

- They are less costly than evaluation with user participation
- Their usefulness mainly concerns the evaluation of alternative designs (prototypes) or systems whose functions and the way they will be performed by users are predictable

CONTENT 12

Cognitive Walkthrough

- It examines the ease of learning the system by new users
 - It can be applied either in the system design phase or in the product integration phase
 - It requires no special equipment
 - It is easy to use
 - It is efficient
- The evaluator walks through the system step-by-step and investigates whether it supports the user effectively

CONTENT 12

Cognitive Walkthrough Procedure

- Typical scenarios of system use are defined
- They are then broken down into action sequences for each user task
- Examine whether the inexperienced user can successfully complete the task based on the following criteria:
 - Is the next action clear to the user?
 - Can the user link the next action provided by the system to the next goal?
 - After performing the action, is the user able to understand the system's response (whether he/she made the right or wrong choice)?

CONTENT 12

Heuristic Evaluation

- It is a subjective method of examination of the system by usability experts
- It examines whether the various rules and principles of design of interactive systems in general and in particular regarding the specific application are respected

CONTENT 12

Heuristic Evaluation

- It can be applied to:
 - Paper prototypes
 - Intermediate fidelity prototypes
 - Final systems
- It is also known as usability inspection
- When up to 5 evaluators are used it is also called "discount evaluation". Experience from using the method suggests that 5 evaluators can identify on average 75-80% of usability problems in the system

CONTENT 12

Heuristic Evaluation

- The method was developed by Jacob Nielsen in the early 90s.
- Nielsen also proposed ten general rules (heuristics) for the evaluation of interactive systems
- These rules were derived from the analysis of 249 usability problems

CONTENT 12

10 Usability Heuristics for User Interface Design

- Jakob Nielsen's 10 general principles for interaction design. They are called "heuristics" because they are broad rules of thumb and not specific usability guidelines.

CONTENT 12

10 Usability Heuristics for User Interface Design

Heuristics

- #1: Visibility of system status
- #2: Match between system and the real world
- #3: User control and freedom
- #4: Consistency and standards
- #5: Error prevention
- #6: Recognition rather than recall
- #7: Flexibility and efficiency of use
- #8: Aesthetic and minimalist design
- #9: Help users recognize, diagnose, and recover from errors
- #10: Help and documentation

<https://www.nngroup.com/articles/ten-usability-heuristics/>

CONTENT 12

Heuristic Evaluation

- The heuristic rules (criteria) proposed by Nielsen are not a cure-all solution
- For effective evaluation the rules need to be modified to reflect the logic of the application being evaluated
 - Different rules have been developed for website evaluation
 - For the evaluation of collaborative environments different heuristics are needed
 - Heuristics have not yet been developed for the evaluation of other application domains
- The design principles and guidelines form the basis for the development of heuristics for evaluation

CONTENT 12

Methodology

- The evaluation using this method focuses on two main points:
 - The overall design of the system screens.
 - The flow of dialogs, messages and actions required to perform a particular process.
- A small set of evaluators (5-10) is involved who examine the interface and the identity of the problems
- In some cases, there is an observer who records the evaluator's comments

CONTENT 12

Methodology

- Evaluators should be informed about the scope of the system
 - Who are the users, what are the main objectives that users are seeking to achieve by using the system and what usability objectives have been set
- Evaluators may be asked to focus on a specific category of users (e.g., occasional users, students, teachers, etc.)

CONTENT 12

Methodology

- Each evaluator is free to choose the evaluation method (e.g. criteria) he/she wishes to use
 - The evaluators evaluate the system independently
 - They need to test the system at least twice
- **First iteration**
 - Problems that novice users may encounter are sought
 - Look for problems related to the mental model of the system that the designers are trying to create for users

CONTENT 12

Methodology

- **Second iteration**
 - Primary user goals, usability goals set, design rules and guidelines, and environmental parameters are checked
 - The above checks are usually in the form of checklists

CONTENT 12

Methodology

- **Third iteration** - *performed if there is time*
 - Any problems not diagnosed in the previous passes are sought
- At each stage the evaluators may ask for clarification on the use of the system either:
 - In advance
 - If they identify a problem
- The process is concluded with a meeting of the evaluators and the creation of a common list of problems identified and their overall assessment

CONTENT 12

Assessment of the severity of problems

- The assessment of the usability problems identified is based on four factors:
 - Frequency of occurrence of the problem
 - Ease with which the problem is overcome
 - Whether the problem is on-off (occurs only once) or persistent, i.e. it requires resolution before letting the user proceed to the next step
 - What impact the occurrence of the problem has on the user and the system

CONTENT 12

Assessment of the severity of problems

- The severity of the problems is rated on the following scale:
 - **0** => I do not agree that this is a usability problem
 - **1** => *Aesthetic problem*: not required to be fixed unless time is available
 - **2** => *Minor usability problem*: solving the problem has low priority
 - **3** => *Major usability problem*: it is important to fix, solving the problem has high priority
 - **4** => *Usability catastrophic*: it is absolutely necessary to fix the problem before the system is completed

CONTENT 12

Evaluation Results

- The result of the evaluation is a report which includes:
- A brief description of the system, the key users and the key objectives that the users are pursuing through the systema
- Summary of the evaluation methodology followed

CONTENT 12

Evaluation Results

- A description of the usability problems identified:
 - Description of the problem
 - Who (category of users) is likely to encounter the problem and when
 - Path (actions) to be taken to encounter the problem
 - Number of evaluators who identified the problem and number of evaluators who agreed that it is indeed a problem
 - Severity of the problem

CONTENT 12

Evaluation Results

- Design ideas to address the problem
- Good usability points in the system

CONTENT 12

Some Statistics on the Method

- In total, each evaluation session lasts about 2-3 hours
 - 30 minutes are left for briefing the evaluators on the system and users
- 5 evaluators identify about 75% of the problems
- 10 evaluators identify about 90% of the problems

CONTENT 12

Some Statistics on the Method

- It is helpful if the evaluation team is composed of a combination of usability experts and domain experts

- It is helpful if the evaluation team is composed of a combination of new and old evaluators (who have participated in other evaluations of the system in the past) in each design phase

CONTENT 12

Advantages

- There are few ethical and practical issues to resolve given that no users are involved
- It is a minimalist approach
 - Some design guidelines can identify many common usability problems
 - Can be implemented with little effort and is easy to remember at a later stage

CONTENT 12

Advantages

- It is a discounted usability approach
 - No end users of the system are required
 - Cheap and quick way of inspecting the system
 - Can be performed not only by usability experts but also by end users or domain experts

CONTENT 12

Disadvantages

- Nielsen's general heuristic rules cannot be applied in all cases:
 - In many fields empirical rules either do not exist or are in their infancy
 - Developing empirical rules with practical value is a relatively painful process
- Even when Nielsen rules can be applied:
 - It is not feasible to transform them into a checklist format
 - Their application (use) requires some skill and experience (sometimes evaluators need to be trained for up to a week)

CONTENT 12

Disadvantages

- In many cases significant problems are not detected using this method
 - It is advisable to combine it with other forms of evaluation
- In many cases it is the trivial (non-significant) problems that are identified
- It may be difficult to find suitable experts for the evaluation
 - The best evaluators are those who, in addition to being usability experts, are knowledgeable about the application domain and the end users

Experimental Techniques

CONTENT 12

Experimental techniques

- The sample size should be sufficiently large to be representative of the user population.
- Experimental evaluation variables:
 - Independent (number of menu options, interface style, icon design, etc.)
 - Dependent (speed of selecting a command, number of errors, etc.)
- Hypothesis: a change in the value of the independent variable will cause a specific change in the dependent variable.

CONTENT 12

Methods of Experimental Evaluation

- Experimental controlled use of the system
 - A classical software evaluation technique that provides quantitative measurements of system performance when users perform predefined tasks
- Protocols for recording user feedback and behaviour
 - Think aloud protocol

CONTENT 12

Experimental controlled use of the system

- A classic technique that provides quantitative measurements of system performance when users perform predefined tasks.
- Measurements are made in laboratories under conditions that simulate real-life operating conditions.

CONTENT 12

Experimental controlled use of the system

- Quantitative data is useful when testing system performance against some other performance target.
- Experiments may be combined with qualitative measurements and observations interviews or questionnaires.

CONTENT 12

Think Aloud Protocol

- The emphasis is on system efficiency and user satisfaction.
- Users express their thoughts and feelings loudly as they interact with the system.
- The user's actions and thoughts are recorded and then analyzed.
- The method requires few resources and has proven to be very effective.

CONTENT 12

Advantages

- The evaluator draws conclusions about the user's mental model.
- If the user's sequence of actions is different from what is expected for the task, it is concluded that the system is not clear enough.
- A record is made of the terminology used by the user to check whether it is in line with that used in the manuals and in the system interface.

CONTENT 12

Disadvantages

- Expressing thoughts in a loud voice may disturb the user's concentration, e.g. a young student trying to solve a difficult problem.
- It is difficult for experienced users to express all their thoughts after they have automated many of their actions

CONTENT 12

Methods of recording actions

- Evaluator's notes
 - Less expensive method
- Recording of subjects
 - Useful in protocols of the 'talking subjects' category.
- Video recording of subjects
 - Loss of detail, such as facial expressions, etc., which can only be captured in closer shots.
 - Need to synchronize with screen image

CONTENT 12

Methods of recording actions

- Computer logging of events
 - Logging at keystroke level
 - High-volume material, so analysis is a very laborious process
- User logging (user logging)
 - of a subjective nature

CONTENT 12

Exploratory Techniques

CONTENT 12

Exploratory techniques

- They aim to investigate the usability characteristics of the system by recording the user's opinions.
- They are carried out in the physical space.
- They do not require special equipment.
- **Main Methods**
 - User interviews
 - Completion of questionnaires
 - Field observations

CONTENT 12

Interviews

- Direct contact between the user and the evaluator.
- **Advantages**
 - Flexibility (e.g. the level of questions can be changed according to the situation)
 - Possibility to record in detail users' views on the system
- **Disadvantages**
 - Inherent difficulty in analyzing and comparing the material and drawing general conclusions
 - Inability to draw quantitative conclusions.

CONTENT 12

Questionnaires

- Clarity in the wording of questions is a key feature of questionnaires.
- **Advantages**
 - Administering questionnaires and drawing conclusions is much easier than in interviews.
- **Disadvantages**
 - Unlike interviews, less flexible, since the questions are written down and cannot be changed on occasion.

CONTENT 12

Questionnaire Types

- Open ended questions
- Closed questions
- Semantic Differentials
- Likert Scale
- Classification of objects
- General questions

CONTENT 12

User Observation

- Observation of users in the physical space in which they carry out their work.
- **Advantages:**
 - greater naturalness of the user's actions, stronger team spirit in their action, since they are in a known place with known team members.
- **Disadvantages:**
 - less quantitative recording of actions, possibly the recording conditions may not be completely satisfactory (noise).

CONTENT 3

Sources and further readings

- HCI course, Department of Computer Science, University of Cyprus

MAI4CAREU

Master programmes in Artificial
Intelligence 4 Careers in Europe

Thank you.