Jan Daciuk, Jerzy Dembski, Magdalena Godlewska,
Adam Kaczmarek, Agata Kołakowska, Jacek Lebiedź,
Maciej Smiatacz, Mariusz Szwoch, Wioleta Szwoch

INTELLIGENT INTERACTIVE SYSTEMS TECHNOLOGIES

edited by Bogdan Wiszniewski

Gdańsk 2022

GDAŃSK UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE
CHAIRMAN OF EDITORIAL BOARD
Dariusz Mikielewicz

EDITOR OF SCIENTIFIC PUBLICATIONS
Michał Szydłowski

REVIEWERS
Wojciech Jędruch
Marek Moszyński

COVER DESIGN
Wioleta Lipska-Kamińska

Published under the permission
of the Rector of Gdańsk University of Technology

Gdańsk University of Technology publications may be purchased at
https://www.sklep.pg.edu.pl

ISBN 978-83-7348-873-1

Preface

I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9*, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning.*

Alan M. Turing. Computing Machinery and Intelligence. *Mind*, 59(236): 433–460, 1950.

In the famous imitation game proposed by Turing the human interrogator has to determine which of the other two players is the person, and which is the machine. The machine is responding to convince the interrogator to mistakenly conclude that the machine is the other person playing the game, whereas the other person is trying to help the interrogator to correctly identify the first player as a machine [278]. From the publication of this concept by Allan Turing more than half a century ago to the present day, there has been a lively discussion in the world literature how to interpret this concept when testing for machine intelligence [409]. The term 'Turing test' was coined to refer to various kinds of behavioral tests performed by humans on a machine for the presence of mind, or thought, or intelligence in it. An interesting fact is that Turing himself stated in his earlier work [387] that intelligence should be viewed as an *emotional* concept rather than a mathematical one. He drew a clear distinction between these two concepts by indicating that mathematical thinking drives the subject to consider an object to behave in an intelligent manner by perceiving only its objective features, whereas the emotional concept indicates the need for both, perceiving the objective features of the object and the state of mind and training of the subject. That probably explains, why even at the beginning of the third decade of the twenty first century, it is still unclear how much progress has been made towards the goal of computer systems to interact with humans as their peers.

In 1991, American inventor and philanthropist Hugh Loebner launched a competition aimed at recreating the conditions of the Turing test to assess the success of conversational programs in passing as human. Unfortunately, computer programs submitted to the Loebner Prize competition conducted every year since then still has not passed the 'imitation game' defined by Turing. Nevertheless the competition staged the ground for proving AI's ability to deceive humans with a natural language capability, both in speech recognition and text generation. Contemporary AI voice assistants, like Amazon's Alexa or Apple's Siri are becoming more and more widespread in everyday use of people worldwide. On the one hand, developments in language generators are also quite impressive. For example, the language AI model GPT-3 [51] can generate fiction, poetry, press releases, even music or jokes. But of course

neither intelligent chatbots nor text generators have been yet close to passing the Turing test by surviving interrogation with a high degree of success over a repeated number of trials by a human subject who knows that one of the other two participants in the conversation is a machine. But if computer programs, although unable to act skillfully in the diverse range of situations that a person with common sense can, could somehow outperform human beings in limited tasks in specific environments, the human perception of the machine intelligence may change dramatically. And this is the idea behind this book.

Natural User Interfaces (NUIs) are an emerging field of human-machine interaction that uses natural human abilities to integrate technology with human-learned modes of interaction. Most computer interfaces are not natural, as they use artificial control devices that must be learned to operate. Instead, NUIs offer users the ability to perform natural human activities as when communicating with other people, involving not only vision, touch, speech, and textual utterances, but also cognition, creation, and exploration of the surrounding world environment [208]. Owing to the AI solutions embedded in a NUI enabled application, these activities can be quickly detected by the computer end devices to control all of its operations or just to manipulate the contents of its screen. In this sense, the Turing prophecy we quoted at the beginning can be considered fulfilled. Maybe today machines are not able to think yet (whatever we shall mean by that), but even if they would never be able to do so in the future, they can certainly react ”humanly” to the actions of their users.

The key to achieving that is AI, therefore we start the book from a comprehensive snapshot of the current state of the art and trends in machine learning in Chapter 1. The issues of feature extraction crucial for the detection and recognition of graphic objects, from printed character recognition to face recognition systems are discussed, as feature extraction is known to produce particularly good results in situations when the training sets are small. We focus on the extractor training methods when the extractor training process is not a part of the classifier training process and labeled data from similar classification tasks are used for extractor training. Particular attention is paid to feature extraction methods related to deep learning, such as multitask or transfer learning techniques.

Next in Chapter 2 we concentrate on conversation based on text, the basic mean used by people not only to convey information, but also to gather it for further use in document archives. The ultimate goal is to make computers understand text, both written and spoken. Given the theoretical limitations set up by Chomsky’s mathematical theory of syntax [67], NLP algorithms should be able to deal with a substantial ambiguity in the structure of human languages, whose usage is to convey meaning amongst a huge range of possibilities, making the memory of each such a language practically unlimited. Owing to AI much progress was made to cope with this problem. So the focus of this chapter is on the key component of any modern speech recognition system which is a language model. Syntactic, semantic and the relatively new neural models are discussed, along with their possible combinations to further improve the scores.

Besides exchanging textual content, seeing is another important way of communicating for humans. From an almost endless list of possible applications of computer vision we address in Chapter 3 the issue of biometric identity verification, a very convenient and widely desired way to protect devices and their resources against unauthorized access for its ease of use. The computer vision techniques matured over the years to the point that now showing up in front of a computer can be like looking at someone just to be recognized and to hear words of greeting, a very attractive alternative to typing in hard to remember complex tex-

tual passwords. In that regard the chapter discusses methods which are capable of protecting automatic speaker verification systems from playback attacks and proposes a new method to detect attacks performed in an environment entirely different from the training one and with the use of the equipment that differs considerably from the devices that captured the training samples.

After recognizing a user, computer systems may show more interest in the person, trying to understand his/her state of mind – to become more friendly, understanding, or simply more effective during the conversation. This is the issue of *affective computing*, a relatively new area of computer science. We focus on these issues in the next three chapters of the book.

In Chapter 4 we review the current state of the art of collecting and processing user data, which as biosignals emitted from various places of the human body can help the system to identify the actual emotions of the user and adjust accordingly its operations. Models of emotions, various ways of evoking emotions, as well as their theoretical foundations are discussed. In particular physiological signals and their relationship to emotions are examined, along with the relevant methods and algorithms, including possible further research directions.

In Chapter 5 yet more aspects related to understanding human users by computer systems are addressed, namely their behaviors exhibited during interaction. Formally they shall be considered as a component of a wider issue of biometrics addressed in part in Chapter 3, however a properly designed biometric system may infer more about users than just verifying their identity. AI enabled analysis of users' characteristics may also tell much the system about their skills, preferences or mood. Several application areas for that are reviewed, including user authentication, emotion recognition, diagnosis and therapy of disorders based on data from standard input devices, without resorting to solutions requiring sensors like previously in Chapter 4.

Finally, Chapter 6 goes beyond viewing the intelligent computer system as just a reactive entity, where the only source of change are the users and their actions. Here we take a closer look on the situation when the system may have some "conflicting interests" with the user. This is the case of computer games, where the user is confronted with a problem to be solved, but the system is not helping so much in that. This mechanism is explored in serious games supporting the concept of active "experiential" learning, a very promising research direction in modern computer-assisted education [407]. We focus on the emotional aspects of such games and explore AI capability to recognize emotions of the user (player) to make the gameplay (the education process) more effective by avoiding frustration of the player if the problem to be solved becomes too difficult for him/her, as well as boredom if it becomes too simple.

Yet a higher level of active systems may be achieved when users are fully immersed in an interface which is a 3D computer generated virtual world and can interact with surrounding objects of that world as they were in a real one. This is the issue covered by Chapter 7. Interaction in such a world is both multidimensional and multimodal, with the possibility of free movement of the user in any direction and the simultaneous stimulation of several of his/her senses, most often sight, hearing, touch and sense of balance. Moreover, all the techniques discussed previously in Chapters 4-6 may be integrated in the virtual worlds to do both, inducing a specific behavior of the immersed user and providing feedback from the user to the 3D scene generation system. Numerous examples of such applications are presented and discussed in the chapter.

However, artificial intelligence is not always required for the computer application to be active or even take the initiative in a conversation with a user. This is the case of reactive or

proactive electronic documents which are capable of simultaneously playing two roles in a computer system: a static information unit and a dynamic interface component [128]. These issues are addressed in the last two chapters of the book.

In Chapter 8 a case study of implementing the *executable paper* concept is presented. This concept involves augmenting the original document informative (data) content with functionality embedded in it to enable implementation of various interactive scenarios, like repeating experiments reported in a scientific paper in reproducible research [114], initiating and tracking of reader's actions in interactive forms [151], interpreting document data in alternative ways to extract useful information [288], or investigating their provenance and meaning, as discussed in the chapter. The IODA model of document architecture used in this case study enables implementation of reactive document behaviors, where users initiate and lead the conversation with a system.

In Chapter 9 we present a proactive document concept. It provides a content for the user to work on and at the same time some embedded functionality enabling the document to migrate in a network organization as a standard email MIME attachment. Upon delivery, its carrier message is retrieved from a mailbox by a specially developed email client and the attached document is activated to interact on its own with a local execution device and its user as a proactive software agent. With that arbitrary complex distributed collaborative processes may be implemented. Although no specific AI solutions are needed for that, the system of such mobile document agents can introduce self-organization in the dynamically (ad hoc) set virtual organization owing to an individual migration path embedded in each document.

All research areas discussed in this book are the subject of intensive research conducted for many years at the Department of Intelligent Interactive Systems at the Faculty of Electronics, Telecommunications and Informatics at Gdansk University of Technology. The common denominator of this activity is allowing people to interact with the computer in the most natural and obvious way, regardless of their age, training, education or degree of disability. Arguing somewhat with Turing's expectations as expressed in the concept of his famous 'imitation game', there is hardly no need for computers to deceptively impersonate human beings. Nobody expects them to today. Instead, users would rather expect computer applications and systems to play all day long a role of obedient servants – skillful executors of commands, wise advisers, understanding teachers, good entertainers, caring helpers, and so on – depending on the time of day, need of a current situation or user's mood. The examples in this book show that today's AI and visualization solutions, supported by the vast resources of data available on the Internet, are dealing with these challenges quite well. And, as may be seen, each day they can do more.

Enjoy the book!

Bogdan Wiszniewski
Gdańsk, June 2022

Contents