Univerza v Mariboru

FRAMEWORK PROGRAMME OF EARLY STAGE RESEARCHER TRAINING[1]

1. BASIC DATA

Mentor's name and surname	**Matej Rojc**	Mentor's register number at ARIS (SICRIS)**:**	**18876**
Mentor's e-mail:	matej.rojc@um.si	Mentor's tel. no.:	02 220 7223
Research programme (RP) leader's name and surname:	Zdravko Kačič	RP leader's register number at ARIS (SICRIS):	06821
Title of research programme:	ADVANCED METHODS OF INTERACTION IN TELECOMMUNICATION	RP's Register number at ARIS (SICRIS):	P2-0069 (B)
Research organisation (RO) of University of Maribor, where training shall be conducted:	Faculty of Electrical Engineering and Computer Science	RO Register number at ARIS (SICRIS):	0796
Research field according to ARIS classification:	2.08 Telecommunications	Research field according to Ortelius classification (EURAXESS)	15.7 Communication engineering

2. DEFINITION OF RESEARCH PROBLEM AND GOALS OF DOCTORAL RESEARCH[2]

Starting point of research task of the early stage researcher and its position in the research programme, where the mentor is included, work hypothesis, research goals and foreseen result with emphasis on an original contribution to science:

The mentor is included in the research program "Advanced interaction methods in telecommunications" P2-0069, which involves the development of advanced multimodal interfaces and covers theoretical, development and applied research. In the field of AI-based multimodal interfaces, research is focused on developing new ways of verbal and non-verbal communication, taking into account the target applied domains of the Internet of Things. The goal is to develop multimodal interfaces based on AI technologies that will enable more natural and user-friendly communication with built-in virtual agents, capable verbal and non-verbal communications. The

[1] Term early stage researcher (ESR) is written in male form and used as neutral for women and men.

[2] Research and study programme of training have to harmonise with contents of the research programme, where the mentor is a member.

mentor was involved in a national basic project (Humanipa J2-1737 in which a new concept of understanding of conversational Intelligence (CLU - Conversational Language Understanding) is developing, as a new approach that further develops the idea that verbal and non-verbal conversational signals are complementary and equal important in conversational terms. As part of this project, we have developed new models based on data fusion and algorithm on the basis of artificial intelligence AI, which capable of creating an in-depth understanding in cognitive mutual mutual with communicative purpose, as the central core in human interaction. The development of algorithms is based on techniques Deep learning and uses comprehensive talk signals contained in EVA-CORPUS. The mentor has been a national coordinator of the COST 18231 campaign, where project members have been dealing with the problem of more advanced generation of text, which represents a key technology for developing advanced human-machine interactions using natural language.

The starting point of a young researcher's research work:
The starting point of the work of the young researcher is directed to an area where a mentor is also actively involved. The starting point of the young researcher's research work is focused on the problems we face and know in these research areas, and are difficult to address isolated, since it is necessary to connect the knowledge of one field, with the knowledge of another field. Conversational Language Generation CLG, the problem of advanced conversation responses, which would be more similar to the natural response of people in the AI-based talk situations by engaging conversational agents, is a very topical area of research, as research is often concerned with more advanced generation of natural language on Natural Language Generation NLG, which is otherwise a sub-processing of natural language processing (NLP), than with a comprehensive treatment of the problem in terms of multimodal communication, which is present in talking communication between people. The starting point of research work is thus not only to integrate the basic aspects of artificial intelligence, but also to understand verbal and non-verbal information, and to connect them for more advanced automatic creation of conversational responses, but also touches the fields of cognitive science. Today, this is one of the biggest challenges in the development of general artificial intelligence (Artificial General Intelligence AGI). Under the auspices of the work results, we can place applications for dialogue systems-targeted systems and open domain systems (Goal Oriented Systems, Open-Domain Systems), which are also the fields of work of projects in which the mentor is involved. In recent years, the processing of natural language with neural network models has made tremendous progress regarding the acquisition of text information, understanding texts, and also at the level of automatic generation of texts for better naturalness and conformity. This was provided by more efficient algorithms for machine learning and neural network applications that are today designed to learn presentation at higher, more abstract levels where presentations or presentations. Knowledge can be more "thickened" and consumes less dimensions, and consequently, deep learning architectures are able to better capture grammar and semantic generalization in texts. The use of neural network architectures also allows for a longer history, with it very important that they reduce dispersion, as well as an explosion of the number of parameters through the projection of history into space with less dimensions so that similar history "knows" to share related presentations. In order to produce more complex responses for advanced interaction systems, we certainly need effective machine-learning algorithms based on deep learning techniques, where in the case of a young researcher's research work, it is a challenge for how such models and techniques develop non -verbal signals for inclusion, with appropriate The models are not yet known, nor is the success of the automatic formation of advanced natural multimodal responses. It is known that neural networks from end to end can be very effective in various teachings, and the problem is the interpretation of models, which can be very important in the creation of such complex conversational responses and understanding the decisions of the model in specific conversational contexts and situations. The exploitation of the attention mechanism in the field of deep learning techniques in the basic description of the model for different modality is an area that can help develop more interpretable models as well as visualizing learned presentations and can also be the subject of research work. By increasingly understanding the deep learning of neural networks in the field of forming multimodal conversational responses, it will certainly be more possible to integrate them into a wider range of tasks and goals that develop in the fields of projects that also mentor is involved. Dialogue and conversational agents (CA), which are also the subject of research, projects P2-0069 and J2-1737 from different angles, today include mostly generating semantically meaningful and

consistent responses on the basis of text only to which people due to greater naturalness They respond simpler and are also more informative for them when the system interacts with people. In more advanced models of dialogue, which also include the synthesis of speech and conversational agents on the responses, it is impossible to imagine the appropriate level of naturalness, if it is based on the formation of responses only at the level of the text, as conversational agents have no knowledge of how to generate a response similar to those , who perform in everyday communication between people. On the other hand, conversational agents are already very important for many different applications, both in the case of targeted dialogue with a closed domain that help users achieve a specific goal, as well as in the case of an open domain conversational agents involved in talking with people - called Also chatting models (chit-chat models). IBM Watson, Apple Siri, Microsoft Cortana, Amazon Alexa and Ipsoft Amelia are examples of such systems that indicate great needs for new human -computer interactions (HCI) in the field of multimodalism and connecting verbal and non -verbal information on the page Creating responses, not only natural in the context of multimodalism, but also in the direction of more comprehensive conversational responses, which can also take into account the conversational context. The development of conversational agents and dialogue systems are now addressing a lot of problems, which is also focused on the candidate's research work, with the results of the research being an important upgrade of existing knowledge and solutions in this field of work that will go beyond the restrictions represented by the formation of responses only to the level of the text. Research will thus be directed to the problem of encoding context or. Coding context information, for example, using knowledge or previous conversation databases that can represent important information and to ensure that a conversational agent has more information on how to create a more consistent, informative, and original conversational response that will know to follow the context of the conversation. It will also need to be involved in the field of understanding of conversational behavior and language. Research work will also be focused on understanding text and personalization, as it is crucial to follow the goal of developing a talk agent who would have such a coherent personality (Coherent Persona). Although the concept of personalization itself in psychology is already quite well researched (research on the use of personalization properties is based on Big Five), it is still difficult to define the characteristics of personalization, and data on this is difficult to obtain. Alternative approaches that take advantage of psycholinguistics are still in its infancy. Some approaches are based on explicit or implicit modeling of personality. Research work in this respect will also address the treatment of problems such as: how to present personalization, how to present non -verbal behavior, and how to bring it through deep learning techniques into the model of conversational responses that would be as natural as possible. or. Understanding the entrance or. the user. Not only in the form of text, but above all multimodal. Today, the most commonly used models of transformers are used to create responses at the text level, so research work will also touch on questions whether such models are also suitable for integrating verbal and non -verbal information on the response page.

The use of large bases of knowledge and language sources always represents a key element of the development of NLP, NLU, CLG, etc., as they can ensure the high performance of ML techniques in the deep learning field. Still, the key problem of research is the question of how to include various language resources and knowledge bases in models - e.g. How to process appropriate relevant content, taking into account different sources and data formats. Looking only at the text level, the performance of the models is improved by using sources based on semantic structured knowledge, e.g. Wordnet, Babelnet and Imagenet. The most commonly used corps are Open subtitles, Twitter Conversation Dialogues, Movie Triples, Cornell Movie Dialogues. Recently, new data sets such as Persona Chat Dataset, Reddit Dataset, have also become available. One of the important problems today is boring and general responses from conversational systems, which in turn also affect the course of communication with conversational agents, which therefore do not mostly allow longer interactions with users. In these cases, conversational agents are often used from end to end based on sequential models, with generated boring and general responses, e.g. I don't know, I'm not sure. etc. The area of research work will therefore be aimed at finding more complex modeling of multimodal conversations, even at the expense of integrating more contextual information, understanding of the word and with the help of attention mechanisms.

Evaluation is one of the most important aspects of finding new knowledge and new methods in the field of a candidate's research work. In this area, this is still an open research problem, as there are no suitable or standardized performance metrics. Researchers are mainly used to adjust automatic metrics, such as Bleu and meteor, to certify performance. However, research has shown that these

metrics show less or no correlation by evaluating people. Rating with people is now the second most common metric of assessment that exists in this area, and researchers use different criteria such as: semantic importance, appropriateness, interesting, fluidity and grammatical appropriateness. In the context of the research work, both objective and subjective valuation methods will be used, addressing the question whether objective metrics allow comparable assessment performance compared to the assessment metric, which includes people.

Working Hypothes of Research:
Based on understanding verbal and non -verbal signals, by incorporating elements of cognitive architectures, deep learning algorithms, neural network architectures, generative AI, and on the other hand, advanced modeling (by integrating verbal and non -verbal, eg emotional signals), we can create more natural and more complex talk responses , which will be more appropriate for their use in communication by conversational agents. Based on the results of the research, which will be obtained on the basis of the adaptation and upgrading of the use of deep learning and with better use of memory mechanisms, more advanced dialogue systems will also be able to use the knowledge of longer conversational contexts as well as domain knowledge from databases. With understanding and using non -verbal content, it will also be possible to actually face the very important challenge of modeling the personality of the conversation agent.

Research Objectives:
(1) How to effectively use multimodal information from different sources (signals),
(2) How modern machine learning methods and generative AI can be used in the formation of more natural and contextually demanding responses,
(3) or new AI models may ensure the development of advanced dialogue systems, also due to the limitation of the necessary data,
(4) Can multimodal information in corpora provide a more appropriate and more natural response from the system than if the response is generated only on the basis of the text?
(5) What is the relevance of individual conversational signals to form an expected response from the multimodal system,
(6) How much are different talk signals important in creating responses from the personality of the conversational agent as well as naturalness?
(7) How to understand multimodal information at the entrance (e.g. communicative purpose, emotions, gestures, facial terms, etc.) helps generate a proper response with gestures that would include appropriate context, communicative purpose, emotions, gestures, facial expressions, also choosing Speaker, age, speaker's gender, speaker profile?

Expected Results:
Research work will be directed to the areas of NLP, NLU, CLG and systems with multimodal interaction. In doing so, finding solutions for problems that will upgrade and improve automatic dialogue responses. In this context, the work addresses several core challenges that represent the basis for original contributions to science:
(1) Data processing, with various sources involved (text, images/video, sensors),
(2) Modern approaches to machine learning to develop more complex AI models,
(3) More advanced more natural and contextual interaction within multimodal AI systems.

Aim of research:
(1) A new model of formation of a colloquial response in multimodal systems, which will be coherent and understandable to people even at the level of non-verbal signals, and will be more advanced in terms of inclusion of verbal and non-verbal information, and as such represents an original contribution to science.
(2) Personality modeling and personalization through fusion of verbal and non-verbal information, which represent the appropriate or appropriate. The natural "answer" to the user's input data is an original contribution in science as a solution,
(3) Research in machine learning methods for the elimination and adequate weight of relevant information may be an original contribution of science,
(4) New methods that make the learning data more effectively are the original contribution. Neural networks usually need large amounts of data to ensure good performance on a particular task.

Using knowledge transfer is an example of good practice on how to use the existing data for a new problem.
(5) The role of cognitive architectures has not yet been investigated for deep learning architectures and needs to be better explored, as they are very important for the construction of intelligent conversational agents by modeling conversational behavior. Cognitive approaches are conceptually and practically embarking on such long-term memory (short-term memory), along with the mechanism of selecting an action that represents a bridge between them. Modeling such architectures in deep learning in the field of advanced responses is an original contribution to science.

3. STUDY PROGRAMME

Foreseen study programme, to which early stage researcher shall be enrolled in academic year 2024/2025:

Framework plan (year 1):
1st semester:
- Examination
- Examination
- Optional subject
- Optional subject
- Optional subject
- IRD with a seminar
- at least one conference post
- analysis of verbal and non -verbal signals
- analysis of existing knowledge in the field of responses in advanced HCI systems,
- analysis of the use of modern approaches to deep learning relating to the goals of research work,
- Using suitable databases and processing data for DL learning
- Functional design of an advanced module to form intelligent and natural responses

2nd semester:
- Optional subject
- Ird
- at least one conference post
- learning advanced models, evaluating models,
- Defining/Setting up the development environment and the frames used, visualizing for development
- Advanced Model Model Modeling Interviews and Platforms for Objective Testing
- In-depth study of working hypothesis and research issues
- Modeling context in human-human conversations, personality modeling and personalization to the theoretical level
- the role of cognitive architectures and the possibilities of use in the design of the responses formation

Framework plan (year 2)
3rd semester:
- Seminar
- Ird
- Creating a new Model Model in Multimodal Systems
- Research Methods ML for extraction and weighting relevance of several information
- Research to model relational addictions of verbal and non -verbal signals

approach resolution of ambiguity in creating responses, methods for optimal use of teaching data
the derivation of all formal procedures to apply for doctoral thesis

4th semester:

Seminar
Ird
Development and first testing of an advanced response model in multimodal systems
Experiments using various approaches and solutions, checking the work hypothesis,
Defining the limits within the model
Publication of dissertation findings in JCR index

Framework plan (year 3)

5. Semester:

Ird
Developing a model and finding solutions to problems and restrictions
Evaluating a model using objective and subjective tests
Publication of dissertation findings in JCR index
Preparation of doctoral dissertation (theoretical and experimental part of research)

6. Semester:

Production and defense of doctoral dissertation

4. DESCRIPTION OF WORK AND TASKS

Working Hypothes of Research:
Based on the understanding of verbal and non -verbal signals, by incorporating elements of cognitive architectures, deep learning algorithms, neural network architectures, and on the other hand, more advanced modeling (by integrating verbal and non-verbal, such as emotional signals), create more natural and more complex conversational responses that will More suitable for their use in communication by conversational agents. Based on the results of the research, which will be obtained on the basis of the adaptation and upgrading of the use of deep learning and with better use of memory mechanisms, more advanced dialogue systems will also be able to use the knowledge of longer conversational contexts as well as domain knowledge from databases. With understanding and using non -verbal content, it will also be possible to actually face the very important challenge of modeling the personality of the conversation agent.

Research work:
Research work will be directed towards NLP, NLU, CLG and multimodal dialogues. Not only classic approaches will be included, but will find solutions to known problems that can upgrade and improve the automatic formation of dialogue systems. In this context, the work addresses several core challenges:
(1) Data processing, with various sources involved (text, images/video, sensors),
(2) The use of modern approaches to machine learning to develop more advanced more complex models,
(3) A more advanced more natural and contextual interaction within multimodal systems.

Work methods:
(1) Analysis of existing deep learning techniques used in the NLP, NLU, CLG,
(2) Use of appropriate databases, processing data that will allow new models of automatic responses,
(3) Learning advanced DL models,
(4) Evaluating the naturalness of the new system,
(5) Analysis, review and comparison of methods of use of structured bases, improvement of awareness of available resources and possible use of them for the challenges of research work.

Questions will be addressed as part of the survey:
(1) How to effectively use multimodal information from different sources (signals),

(2) How Modern Mechanical Learning Methods can benefit from the formation of more natural and contextually demanding responses,
(3) or new models may ensure the development of advanced dialogue systems, also due to the limitation of the necessary data,
(4) Can multimodal information in corpora provide a more appropriate and more natural response from the system than if the response is generated only on the basis of the text?
(5) What is the relevance of individual conversational signals to form an expected response from the multimodal system,
(6) How much are different talk signals important in creating responses from the personality of the conversational agent as well as naturalness?
(7) How to understand multimodal information at the entrance (eg communicative purpose, emotions, gestures, facial terms, etc.) helps generate an appropriate response that would include appropriate context, communicative purpose, emotions, gestures, facial expressions, including the choice Age, speaker's gender, speaker profile?

5. REQUESTED LEVEL OF EDUCATION

Master of science

6. REQUESTED FIELD OF EDUCATION

elektrotechnic, electronics, computer and information sciences

7. KLASIUS SRV

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

8. KLASIUS P

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

9. REQUESTED KNOWLEDGE

Knowledge of programming languages, machine learning, basic artificial intelligence

10. REQUESTED SPECIAL REQUIREMENTS

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

11. REQUESTED LANGUAGES

english language

12. REQUESTED WORK EXPERIENCE

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

13. FORESEEN POSTDOCTORAL TRAINING

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

Mentor's signature:

Matej Rojc

Digitalno podpisal Matej Rojc
Datum: 2025.02.12 08:10:39 +01'00'

Research programme leader's signature:

Digitalno podpisal ZDRAVKO KAČIČ
Datum: 2025.02.12 11:14:24 +01'00'

Name and surname of Dean or authorised person[3]:

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

Signature of dean or authorised person:

Gorazd Štumberger

Digitalno podpisal Gorazd Štumberger
Datum: 2025.02.12 11:44:40 +01'00'

Barbara Ojsteršek

Digitalno podpisal Barbara Ojsteršek
Datum: 2025.02.12 11:23:55 +01'00'

Place and date:

Kliknite ali tapnite tukaj, če želite vnesti besedilo.

Kliknite ali tapnite tukaj, če želite vnesti datum.

Stamp:

[3] The training program is signed by the dean of the member where the ESR's employment and training will take place.