Funded by the European Union

www.gain-twinning.eu

Project Acronym: **GAIN**

Project Full Title: **Georgian Artificial Intelligence Networking and Twinning Initiative**

Grant Agreement Number: 101078950

HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning)

WEBSITE: www.gain-twinning.eu

Project Start Date: 1st October 2022

Funded by the European Union

www.gain-twinning.eu

Deliverable D3.5 GAIN highly specialized trainings report

Dissemination level:	PU
Due Date:	September 30, 2024
Extended date:	October 31, 2024
WP leader:	DFKI
Author:	G. Giorgobiani
Reviewers:	F. Bremond, B.E. Wirth, P. Müller,
Type:	Report
Task responsible:	MICM

Contents

1. Executive Summary

This document describes the **GAIN highly specialized trainings (D3.5, Task 3.5),** conducted at MICM. The document outlines the planning for the forthcoming trainings set out by the DoW of the project.

Document History			
Version	**Date**	**Contributors**	**Description**
1	15/10/2024	G. Giorgobiani, N. Kukhilava, B. Mikaberidze, R. Kalandadze, T. Saghinadze – MICM; P. Müller, B. Wirth – DFKI; M. Balazia – INRIA	Initial draft of the report
2	29/10/2024	G. Giorgobiani, N. Kukhilava, B. Mikaberidze, R. Kalandadze, T. Saghinadze – MICM; P. Müller, B. Wirth – DFKI; M. Balazia – INRIA	Final version

2. Introduction

Project GAIN Highly Specialized Trainings are organized in Georgia by the young researchers of GAIN-MICM team, re-trained by the scientists from DFKI and INRIA. The concept and content are developed jointly. French and German colleagues participate in the trainings via teleconferencing.

The trainings are prepared and conducted by the research groups leaders (as set out in D 2.1 TRIP[1] framework parameters). Each training represents short course for the young researchers and students from MICM, GTU and several other universities.

The overall aim of these activities is to cover specific gaps in knowledge in the topics of AI defined by the TRIP, like AI methods and technologies for Deep Speech Analysis, Human Behavior Understanding, etc. To prepare a foreground for the advanced topics, it was decided to start the training sessions from the basics of Machine Learning, Deep Learning and Natural Language Processing.

The trainings were organized as an interactive event, where the participants posed questions and were involved in discussion and brainstorming. Participation was made available online as well by use of Microsoft Teams. Each training session has been recorded.

The training materials and recordings of the training sessions are available on the project website[2]. The photos and the screenshots of the training sessions are given in the Appendix of this document.

[1] Twinning Research and Innovation Programme

[2] https://www.gain-twinning.eu/?page_id=988

3. Venue & Audience

GAIN Highly Specialized Trainings were held in the meeting room of MICM in Tbilisi, Georgia.
The workshop's audience mainly included GAIN-MICM team members. The invitations have been sent to 20 most active students from 5 universities (TSU, IBSU, GTU, SDSU, TSMU), who took part in the GAIN 2nd Summer School.

4. 1st GAIN Highly Specialized Training: *AI: From Its Beginnings to the Present*

The first GAIN Highly Specialized Training was held on 26 September 2024. The title of the Training was "AI: From Its Beginnings to the Present". It was delivered by Natia Kukhilava, GAIN-MICM team leader, assistant researcher of MICM.
Participation was available online as well. The training session was attended by 17 participants.

Agenda

GAIN Georgian Artificial Intelligence Networking and Twinning Initiative HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning) Project #101078950			
1st GAIN Highly Specialized Training N. Kukhilava. "AI: From Its Beginnings to the Present"			
Date:	16. 09.2024		
Start Time:	14:00 PM		
End Time:	17:00 PM		
Location:	Grigol Peradze 4, Tbilisi, Georgia, MICM, Meeting Room, 2nd floor		
Time	**Minutes**	**Topic / Discussion**	**Comments**
14:00 PM	90	AI: History and Modern State of the Arts	Questions/Answers
15:30 PM	15	**Coffee Break**	
15:45 PM	75	Practical work/quizzes	Problem solving
17:00 PM		Closing	All
17:00 PM	**180 min**	**End of Meeting**	

4.1. Content of training

The training content was structured as follows:

- **Introduction.** Here the participants gained insight into the origins and the development history of AI and its subfields
- **Machine Learning**. In this part Supervised and Unsupervised Learning, Model Training and relevant examples were discussed.
- **Deep Learning**. The principle of work of a Perceptron and Neural Network have been considered.
- **Generative AI**. The topics on Text Generation and Image Generation have been discussed.
- **Problems Solving Session**: House Pricing[3], Image Classification[4].
- **Quiz**. Multiple choice questions[5] have been delivered to the participants via the QR code.
- **Q/A**.

4.2. Training materials

The following training materials have been provided:

- The PP presentation, including the clickable links to the Google Colab to illustrate the AI practical problems, and a QR code for the quiz.
- Online links for practical session:
  - ✓ https://colab.research.google.com/drive/1-rXK-M3IY3YzYMy_X6qWtGWeiBBPqI4Z
  - ✓ https://colab.research.google.com/drive/16BfuQMa8PO3dMcZs4lSa0EkX6GDTDuxv
  - ✓ https://freeonlinesurveys.com/s/aoACvGtE
- Quiz questionnaire.

The Guests list and PP presentation are available on demand at MICM. The presentation and the recording are downloadable at the project website.

[3] https://colab.research.google.com/drive/1-rXK-M3IY3YzYMy_X6qWtGWeiBBPqI4Z
[4] https://colab.research.google.com/drive/16BfuQMa8PO3dMcZs4lSa0EkX6GDTDuxv
[5] https://freeonlinesurveys.com/s/aoACvGtE

5. 2nd GAIN Highly Specialized Training: *Quick Start with NLP and Hugging Face for Georgian Language*

The 2nd GAIN Highly Specialized Training was held on 10 October 2024. The title of the Training was "Quick Start with NLP and Hugging Face for Georgian Language". It was delivered by Beso Mikaberidze, GAIN-MICM team leader, assistant researcher of MICM.
Participation was available online as well. The training session was attended by 16 participants.

Agenda

GAIN Georgian Artificial Intelligence Networking and Twinning Initiative HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning) Project #101078950			
2nd GAIN Highly Specialized Training B. Mikaberidze. "Quick Start with NLP and Hugging Face for Georgian Language"			
Date:	10. 10.2024		
Start Time:	14:00 PM		
End Time:	17:00 PM		
Location:	Grigol Peradze 4, Tbilisi, Georgia, MICM, Meeting Room, 2nd floor		
Time	**Minutes**	**Topic / Discussion**	**Comments**
14:00 PM	90	NLP fundamentals, applications to the Georgian Language	Questions/Answers
15:30 PM	15	**Coffee Break**	
15:45 PM	75	Practical work/code writing	Problem solving
17:00 PM		Closing	All
17:00 PM	**180 min**	**End of Meeting**	

5.1. Content of training

The training content was structured as follows:

1. **Introduction.** NLP as the field at the intersection of Computer Science, AI and Linguistics.
2. **Key tasks**. Sentiment Analysis, Named Entity Recognition, Part-of-Speech Tagging, Machine Translation, language understanding (also generation)
3. **Real-Life Applications**. Voice Assistants, Chatbots, Translation, Spam filtering, etc.

4. **Impact on Everyday Life**. Communication (breaking language barriers), Accessibility (Assisting disabled people), Business (Analyzing customer feedback), Education (Enhancing language learning tools).
5. **NLP and Georgian Language**. Overcoming the Challenges of Low-Resource Languages, Vital Role of NLP Tools in Preserving Georgian, Unlocking New Opportunities for Technological Growth.
6. **Hugging Face**. Extensive model repository, User-friendly APIs, Strong community support, supporting multiple languages, including Georgian.
7. **Hands-On Workshop**. Load and process datasets, Fine-tune, evaluate and test a pre-trained model on your data, Push/pull a model to/from the Hugging Face Hub, infer a model using the Hugging Face pipeline.

5.2. Training materials

- The PP presentation, including the clickable links.
- Online links:
  - ✓ Workshops materials [6,7]: Georgian Sentiment Analysis (Original source: Hugging Face 02 text classification), Hugging Face workshops.
  - ✓ Datasets [8]: Georgian Sentiment Analysis (Source Paper), IMDB Movie reviews.
  - ✓ Hugging Face docs [9,10]: NLP Course, Transformers, Hugging Face videos.

The Guests list and PP presentation are available on demand at MICM. The presentation and the recording are downloadable at the project website.

6. 3rd GAIN Highly Specialized Training: *Georgian first Large Language Model: Techniques and Approaches for Low-Resource Languages*

The 3rd GAIN Highly Specialized Training was held on 15 October 2024. The title of the Training was "Georgian first Large Language Model: Techniques and Approaches for Low-Resource Languages". It was delivered by Raphael Kalandadze, GAIN-MICM team leader, assistant researcher of MICM. Participation was available online as well. The training session was attended by 18 participants.

[6] Georgian Sentiment Analysis
[7] Hugging Face workshops
[8] IMDB Movie reviews
[9] NLP Course
[10] Transformers

Agenda

GAIN Georgian Artificial Intelligence Networking and Twinning Initiative HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning) Project #101078950			
3rd GAIN Highly Specialized Training R. Kalandadze. "Georgian first Large Language Model: Techniques and Approaches for Low-Resource Languages"			
Date:	15. 10.2024		
Start Time:	14:00 PM		
End Time:	17:00 PM		
Location:	Grigol Peradze 4, Tbilisi, Georgia, MICM, Meeting Room, 2nd floor		
Time	**Minutes**	**Topic / Discussion**	**Comments**
14:00 PM	90	Data collection and augmentation approaches, language adaptation methods	Questions/Answers
15:30 PM	15	**Coffee Break**	
15:45 PM	75	Practical work / developing an LLM for Georgian	Case study
17:00 PM		Closing	All
17:00 PM	**180 min**	**End of Meeting**	

6.1. Content of training

- **Large language models:** this part deals with LLMs, which have transformed the field of natural language processing, driving notable progress across various applications.
- **Development of the first Georgian Large language model**: building LLMs for low-resource languages presents unique challenges due to limited data and linguistic resources. This part focuses on the techniques and strategies used to overcome the existing obstacles.
- **Data collection and augmentation**: the approaches specifically designed for low-resource environments, enabling to build a robust dataset that serves as the foundation for model training are considered. Additionally, language adaptation methods aimed at enhancing the model's performance while maintaining its focus on the Georgian language is discussed.

- **Case study**: here the insights into fine-tuning processes allowing to tailor general-purpose models for instruction-following tasks are shared. By presenting this case study on developing an LLM for Georgian, attendees have gained practical insights applicable to other underrepresented languages.

6.2. Training materials

- The PP presentation.
- Online link for demonstration and practical work: https://chat.ailab.ge/

The Guests list and PP presentation are available on demand at MICM. The presentation and the recording are downloadable at the project website.

7. 4th GAIN Highly Specialized Training: *Parameter Efficient Transfer Learning for Visual Transformers*

The 4th GAIN Highly Specialized Training was held on 29 October 2024. The title of the Training was "Parameter Efficient Transfer Learning for Visual Transformers". It was delivered by Teimuraz Saghinadze, GAIN-MICM team leader, researcher of MICM.

Participation was available online as well. The training session was attended by 16 participants.

Agenda

GAIN Georgian Artificial Intelligence Networking and Twinning Initiative HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning) Project #101078950			
4th GAIN Highly Specialized Training **T. Saghinadze. "Parameter Efficient Transfer Learning for Visual Transformers"**			
Date:	29. 10.2024		
Start Time:	14:00 PM		
End Time:	17:00 PM		
Location:	Grigol Peradze 4, Tbilisi, Georgia, MICM, Meeting Room, 2nd floor		
Time	**Minutes**	**Topic / Discussion**	**Comments**
14:00 PM	90	Methods for transfer learning; transformer-based vision models; injecting adapters into	Questions/Answers

		the transformer blocks and embedding layer	
15:30 PM	15	**Coffee Break**	
15:45 PM	75	Practical work/ exploring the subject step by step using Google Colab	Problem solving
17:00 PM		Closing	All
17:00 PM	**180 min**	**End of Meeting**	

7.1. Content of training

- **Introduction**. Transfer learning methods and Large Language Models.
- **Exploring a new domain**. Creation of transformer-based vision models.
- **Using PyTorch, implementing ViT**. Injecting adapters into the transformer blocks and embedding layer.
- **Step by step tutorial using Google Colab**. Build the whole system from the bottom up to deepen our understanding of the process.

7.2. Training materials

- The PP presentation.
- Online links for demonstration and practical work:
  - ✓ https://colab.research.google.com/drive/16X4E6dOj_-D6aMFYYNFrXSj_saHLPZ3a?usp=sharing
  - ✓ https://colab.research.google.com/drive/1WdTMeVxpXdX0i4W1wAscrSelc8gKC8AR?usp=sharing

The Guests list and PP presentation are available on demand at MICM. The presentation and the recording are downloadable at the project website.

8. Plans for the remaining trainings

Within Task 3.5 MICM will organize the following trainings:

- **In November 2024. The tentative topic: Detecting and drawing bounding boxes in videos using ByteTrack.** *Research on emotion recognition from human body movements is the part of a relevant subproject of the GAIN project. Detecting and cropping the people out of videos is one of the steps in this task. The training includes the step-by-step introduction to the above-mentioned technologies and the problem-solving process.*

- **In March 2025. The tentative topic: Developing an Annotation Scheme for the Georgian Psychiatric Corpus.** *This training session will guide participants in creating an effective AI annotation scheme for the Georgian Psychiatric Corpus. The goals of the GAIN-MEPHESTO project will be discussed.*

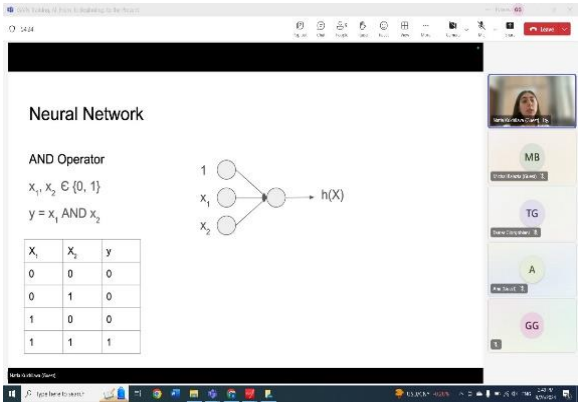
9. Conclusions

- The GAIN Highly Specialized Trainings, conducted by the GAIN research groups leaders at MICM trained by the project partners from DFKI and INRIA, demonstrated high interest in the AI topics of the GAIN team members as well as the external participants. It raised the motivation of the Georgian students to further enhance their knowledge in the field.
- The chosen topics were assessed by the trainees as relevant and interesting.
- The foreground to move to the more advanced topics, which covered the basics of Machine Learning, Deep Learning and Natural Language Processing, was prepared during the firsts 2 training sessions.
- The trainings on Visual Transformers and Georgian LLM were devoted to the advanced topics of AI and contained the achievements of Georgian students as well.
- Discussion of the relevant real-life problems stimulated the interest and strengthened the impact of training. Hands-on training gave the trainees ability to implement AI solutions, from data preprocessing to model deployment.
- Training fostered a systematic approach to problem-solving, encouraging participants to think critically about how to leverage AI in various contexts.
- As the field of AI is rapidly evolving, the training events emphasized the importance of ongoing education to stay current with advancements and best practices.
- The feedback received from the audience at the end of each training after the Q&A session confirmed a high level of satisfaction among participants. They expressed interest in participating in the forthcoming training events.
- Finally, the Georgian students, who had not conducted the trainings before, gained one more valuable experience in this regard.

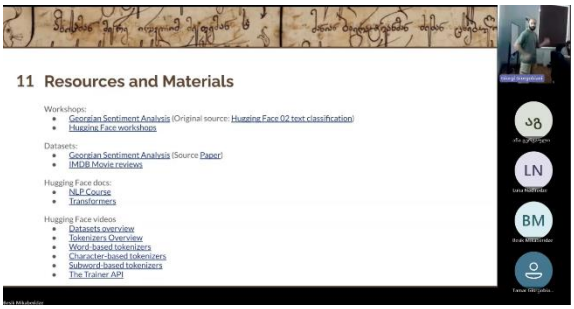
10. APPENDIX

Training venue

Training 1



Training 2



Training 3

Training 4