hcaim human centred artificial intelligence masters

MŰEGYETEM 1782

Posters

presented at the Conference on Human-Centred Regulation of AI (HCRAI)
Faculty of Electrical Engineering and Informatics,
Budapest University of Technology and Economics

Wednesday, June 28, 2023

Conference website: https://hcaim.bme.hu/hcrai

Sponsors:

Continental The Future in Motion

E-GROUP SOFTWARE & BEYOND

QTICS GROUP

BME HUMÁN

Supporters:

ARTIFICIAL INTELLIGENCE National Laboratory

Mesterséges Intelligencia Koalíció

eit Digital

Co-financed by the Connecting Europe Facility of the European Union

These posters – with a single exception – were prepared by students participating in the Human-centred Artificial Intelligence Master's (HCAIM) Programme at one of the four universities belonging to the HCAIM consortium:

- Budapest University of Technology and Economics, Hungary
- HU University of Applied Sciences, Utrecht, The Netherlands
- Technological University Dublin, Ireland
- University of Naples Federico II, Italy

Part of the posters are based on the Master's theses of the students while others were written during the HCAI Blended Intensive Programme in Utrecht between 30 January – 3 February 2023 (https://hcaim.bme.hu/en/bip/).

The last poster is an overview of the HCAIM programme implementation at the Faculty of Electrical Engineering and Informatics of the Budapest University of Technology and Economics (https://hcaim.bme.hu/en/msc-bme/).

Budapest, July 6, 2023

https://humancentered-ai.eu
https://hcaim.bme.hu

Co-Financed by the Connecting Europe Facility of the European Union
Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068

List of posters

Health and Bio

- Human-centered reinforcement learning in de novo molecular design by Mátyás Antal[1], Supervisors: Bence Bolgár[1], Péter Antal[1]
- Framework for Trustworthy AI in the Health Sector by Mykhailo Danilevskyi[9], Supervisors: Fernando Perez Tellez[9], Davide Buscaldi[9]
- Noisy Datasets of Unstructured Text from Cardiology Records with NER Classifier by Mario Minocchi[10]
- GPT-3-based features extraction from EHR of cancer patients by Chiara Salzano[10], Supervisors: Matteo Pallocca[5],·Stefano Marrone[10]

Language

- Detecting Fake News using Machine Learning by Roberto Buono[10]
- A HC approach to abusive language detection: Efficacy of LLMs by Zaur Gouliev[9], Supervisor: Rajesh Jaiswal[9]
- Multi-modal Fake News Detection on Twitter through User's Social Network by Wan Yit Yon[9]

Image and video

- Data Augmentation Approaches for Lip Movements Detection by Daniele Iuliano[3,10]
- MLOps: Using machine learning in the industry by Balázs Tibor Morvay[1]
- Application of Semantic Segmentation by Mátyás Pelle[1], Gábor Sörös[7], Luca Szegletes[1]

Navigation

- Mapping and navigation based on Neural Radiance Fields by Ágoston Csehi[1], Supervisors: György Józsa[7], László Lengyel[1]

Ethical behaviour

- Balancing Innovation and Ethics: AI Digital Twins in Remote Working Hubs by Yilin Li[4], Supervisor: Alireza Dehghani[4]
- A Deep Learning approach for identifying sexual harassment by Beatrix Tugyi[1], Consultants: Tarry Singh[8], Dr. Tannistha Maiti[8]

Explainability

- Revalidation and Explainability of PreSS by Cloë van Geest[9]
- Explainable AI to understand machine learning models by Szabolcs Weyde[1]

Other

- Drone and Digital Twins for Crowded event management – A FIWARE-based approach by Dario De Dominicis[6,10]
- Reinforcement learning in mechatronic systems by Pim Jansen[4]
- Object Detection Model for Defect Inspection in Aerospace: Predictive Maintenance and Ethical Consideration in Industry 4.0 by Flavia Napoletano[10]

HCAIM @ BME at a glance

- Human-Centred Artificial Intelligence Master's Supplementary Programme at the Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics

[1] Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics, Hungary
[2] CeADAR, Ireland's Centre for Applied AI, Dublin, Ireland
[3] CNR, National Research Council, Italy
[4] HU University of Applied Sciences, Utrecht, The Netherlands
[5] IRCCS Istituto Nazionale Tumori Regina Elena, Rome, Italy
[6] Meditech, Naples, Italy
[7] Nokia Bell Labs, Budapest, Hungary
[8] Real AI B.V., The Netherlands
[9] Technological University Dublin, Ireland
[10] University of Naples Federico II, Italy

Human-centered reinforcement learning in de novo molecular design

Mátyás **Antal**[1], Bence **Bolgár**[1], Péter **Antal**[1]

1. Budapest University of Technology and Economics, Budapest, Hungary

ABSTRACT

The proposed research focuses on the application of reinforcement learning in de novo molecular design. By incorporating generative models and human feedback, this study aims to fine-tune the structure of a molecule while satisfying certain constraints. The significance of this research lies in its ability to upgrade current molecular design models and leverage human expertise. The expected outcome of this study is to reinforce the success of reinforcement learning in the field of molecular design and contribute to the advancement of this area of research.

Fig. 1. Imaginary depiction of expert guided molecule design. *Source: DALL-E*

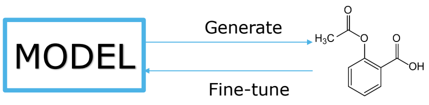
INTRODUCTION

Recent developments in human-feedback reinforcement learning (HFRL) have demonstrated the potential to significantly improve the performance of agents by incorporating human expertise with machine learning algorithms. Building upon the success of machine learning in the field of de novo molecular design, the integration of HFRL has the potential to further enhance molecule design process.

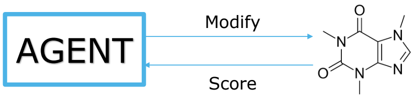
LITERATURE REVIEW

RL-based methods in molecular design can be broadly categorized into two main approaches[1].

1. Model fine-tuning: Pre-trained generative models guided by RL to generate optimized molecules[2].



2. Direct reinforcement: RL directly modifies molecular structure[3]."



Despite the absence of prior research on **Human-feedback RL** in the domain of molecule design, it is currently a rapidly growing area of study, with most methods focusing on guiding the training process by utilizing human feedback[4].

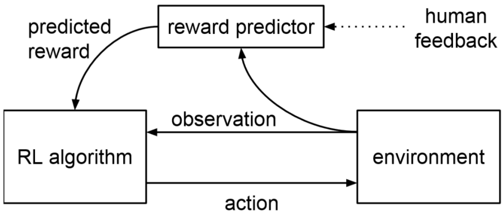


Fig. 2. Schematic illustration of a human-feedback reinforcement learning method developed by Deepmind. *Source: [5]*

RESEARCH METHODOLOGY

1.Review of generative models: Review existing literature on generative models for molecular design, considering factors such as performance and compatibility with reinforcement learning.

2.Model selection: Select the most appropriate generative model to serve as a base for the proposed framework.

3.Framework: Develop a framework for training reinforcement learning algorithms using human feedback, taking into account desired molecule properties.

4.Flexible integration: Design framework to allow easy integration of new generative models and reinforcement learning algorithms.

PRELIMINARY CONSIDERATIONS

Ensuring the success of the design process requires considering human feedback accuracy and consistency. This includes evaluating expert qualifications and feedback collection methods. The limitations of generative models and reinforcement learning must be addressed to achieve optimal results. Responsibility and credit for the final output must also be determined

CONCLUSIONS

The proposed research combines human expertise and machine learning in de novo molecular design through human-centered reinforcement learning. This study aims to provide a framework for more efficient, effective, and ethical molecular design processes. The expected outcome is improved quality of molecular design and advancement of the field.

[1] Du, Y., Fu, T., Sun, J., & Liu, S. (2022). Molgensurvey: A systematic survey in machine learning models for molecule design. arXiv preprint arXiv:2203.14500.

[2] Blaschke, T., Arús-Pous, J., Chen, H., Margreitter, C., Tyrchan, C., Engkvist, O., ... & Patronov, A. (2020). REINVENT 2.0: an AI tool for de novo drug design. Journal of chemical information and modeling, 60(12), 5918-5922..

[3] Zhou, Z., Kearnes, S., Li, L., Zare, R. N., & Riley, P. (2019). Optimization of molecules via deep reinforcement learning. Scientific reports, 9(1), 1-10

[4] Li, Guangliang, et al. "Human-centered reinforcement learning: A survey." IEEE Transactions on Human-Machine Systems 49.4 (2019): 337-349.

5] Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. Advances in neural information processing systems, 30

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This poster was created as part of the Blended Intensive Programme organized under the Erasmus + Programme of the European Union

Framework for Trustworthy AI in the Health Sector

Mykhailo Danilevskyi - D22126578@myTUDublin.ie

School of Enterprise Computing and Digital Transformation,
Faculty of Computing, Digital and Data, TU Dublin, Ireland

Supervisors: Dr. Fernando Perez Tellez, Dr. Davide Buscaldi

human centred
artificial intelligence
masters

OLLSCOIL TEICNEOLAÍOCHTA
BHAILE ÁTHA CLIATH
DUBLIN
TECHNOLOGICAL
UNIVERSITY DUBLIN

Introduction

The European Commission defines that Trustworthy AI should be lawful, ethical and robust. The ethical component and its technical methods are the main focus of the research. According to this, the initial research goal is to create a methodology for evaluating datasets for ML modeling using ethical principles in the healthcare domain. Ethical risk assessment will help to ensure compliance with principles such as privacy, fairness, safety and transparency which are especially important for the Health Care sector. At the same time, risks must be evaluated with respect to AI model performance and possible scenarios of risk mitigation. Ethical risk mitigation techniques involve data modification, elimination of private information from datasets that directly influence AI modelling. Therefore ethical risk mitigation techniques should be carefully selected depending on domain and context. In this research work, we present an analysis of these techniques.

Privacy

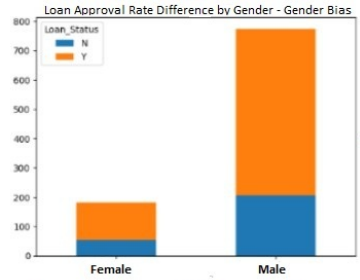
Privacy principle declares that data is protected and used with owner consent. The aim of the section is to research methods for private data detection and de-identification with minimum information loss.
The following methods have been researched for the best suitability for healthcare data: Regex, Conditional Random Fields, Machine Learning (LR, SVM, Decision Trees), Deep learning (CNN, RNN, LSTM, BERT)

Fairness

Compliance with fairness principle ensures that model decisions are not discriminating and equally correct for any group of AI users by gender, age, race, etc. For healthcare AI powered solutions, fair and unbiased decisions are critical for people's health. For example, younger people have higher underdiagnosis AI decision rate than older people, because they are naturally healthier and this is reflected in AI train data. In order to avoid such cases, it is required to measure and mitigate bias in train data. The following detection methods and techniques are being used in experiments with healthcare data: Data pre-processing methods - Re-weighing [1], Optimized pre-processing [2], Learning fair representations [3], Disparate impact remover [4]; Data in-processing methods - Adversarial debiasing [5], Prejudice remover [6]; Data post-processing - Equalized odds post-processing [7], Reject option classification [8]. Tools: AIF360, Google What-if, Fairkit, Fairlearn.

Case Study: Improve Fairness using Disparate Impact Remover Algorithm

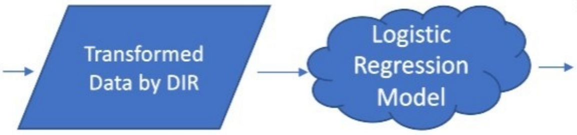
The initial experiment was conducted on the dataset that contains information about the loan applicants*. Gender was selected as applicants protected attribute which means that in the ideal world loan approval rate is equal for applicant regardless of gender. Fairness was estimated with Disparate Impact metric. If the value of the metric is greater or equal to 0.80, then it is considered that there is no discrimination in loan approvals between genders [9]. During the experiment, Disparate Impact Remover algorithm was applied to improve fairness of the loan approval logistic regression model. Disparate Impact Remover algorithm transforms numeric data to minimize differences between genders in loan amount, terms etc. The algorithm was selected randomly for initial exploration of the fairness problem. The algorithm improved Disparate Impact metric from 0.66 to 0.71. The improved result is still lower than normal 0.80, which means that further analysis of other methods is required.
* - experiments with healthcare related data will be conducted in the next stage.



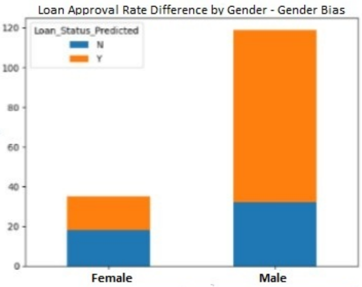
Gender	N	Approval Rate
Female	624	68%
Male	145	74%

DI = 0,92
(0,68 / 0,74)

Disparate Impact (DI) >= 0,80 considered as normal



Accuracy: 0.82
Precision: 0.88
Recall: 0.86



Gender	N (testset)	Approval Rate
Female	35	51%
Male	119	72%

DI = 0,71
(0,51 / 0,72)

(without DIR Transformation)

Gender	N (testset)	Approval Rate
Female	35	48%
Male	119	73%

DI = 0,66
(0,48 / 0,73)

Conclusions and Future Work

Creation of a framework of trustworthy AI in the healthcare sector is critically important for ensuring privacy, bias-free and fair decision by AI-powered healthcare systems. As a future work, the following subgoals are planned to achieve: 1) Identify peculiarities of health datasets; 2) Apply methods and techniques for detecting and de-identification of private information. Define de-identification methods that best work with healthcare data taking into consideration possible loss of meaningful data; 3) Apply methods and techniques for detection and mitigation of bias in datasets. Define the best methods for healthcare.

** The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

References

[1] Kamiran, Faisal & Calders, Toon. (2011). Data Pre-Processing Techniques for Classification without Discrimination. Knowledge and Information Systems. 33.10.1007/s10115-011-0463-8. [2] Calmon, Flavio & Wei, Dennis & Natesan Ramamurthy, Karthikeyan & Varshney, Kush. (2017). Optimized Data Pre-Processing for Discrimination Prevention. [3] Zemel, Richard & Wu, Y. & Swersky, K. & Pitassi, T. & Dwork, C.. (2013). Learning fair representations. 30th International Conference on Machine Learning, ICML2013.1362-1370. [4] Friedler, Sorelle & Scheidegger, Carlos & Venkatasubramanian, Suresh. (2014). Certifying and Removing Disparate Impact. 10.1145/2783258.2783311. [5] Zhang, Brian & Lemoine, Blake & Mitchell, Margaret. (2018). Mitigating Unwanted Biases with Adversarial Learning. 335-340.10.1145/3278721.3278779. [6] Kamishima, Toshihiro & Akaho, Shotaro & Asoh, Hideki & Sakuma, Jun. (2012). Fairness-Aware Classifier with Prejudice Remover Regularizer. [7] Hardt, Moritz and Price, Eric & Srebro, Nathan. (2016). Equality of Opportunity in Supervised Learning.[8] Kamiran, Faisal & Karim, Asim & Zhang, Xiangliang. (2012). Decision Theory for Discrimination-Aware Classification. Proceedings - IEEE International Conference on Data Mining, ICDM.924-929.10.1109/ICDM.2012.45. [9] Dan Biddle (2006). Adverse Impact And Test Validation: A Practitioner's Guide to Valid And Defensible Employment Testing. Aldershot, Hants, England: Gower Technical Press. pp.2–5.

NOISY DATASETS OF UNSTRUCTURED TEXT FROM CARDIOLOGY RECORDS WITH NER CLASSIFIER

Mario **Minocchi**
University of Napoli Federico II

ABSTRACT

With the health care system, the data labels associated with them are **intrinsically noisy**, and training models using noisy labels severely degrades their generalization performance, even if they are essential in modern deep learning applications.
Specifically, the research will be based on noisy dataset of unstructured Italian texts, from **cardiology medical records** using active learning to improve the performance of a medical **NER** classifier based on deep networks.

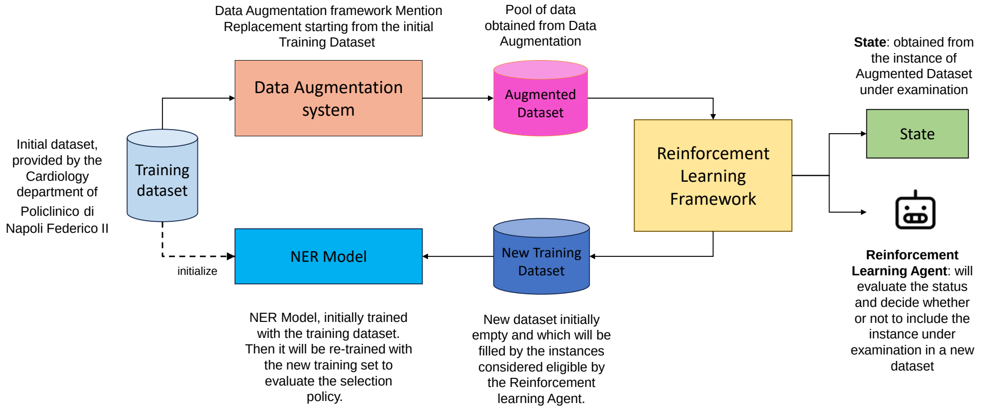
INTRODUCTION

Named-entity recognition is a Natural Language Processing (NLP) task whose purpose is to identify entities within a sentence.

The entities to be identified depend on the domain of interest: in our case, it is **Biomedical Named-entity recognition (BioNER)**, where the objective is to identify entities in the medical field such as diseases, genes, chemical agents.

Since (Bio)NER models require a lot of data to train on, the technique of Data Augmentation is used to obtain new samples by making modifications from the data already available.

The selection of augmented data for the re-training of the NER model is done through a Reinforcement Learning approach called Policy-based Active Learning.



Data Augmentation framework Mention Replacement starting from the initial Training Dataset

Pool of data obtained from Data Augmentation

State: obtained from the instance of Augmented Dataset under examination

Data Augmentation system

Augmented Dataset

Initial dataset, provided by the Cardiology department of Policlinico di Napoli Federico II

Training dataset

Reinforcement Learning Framework

State

NER Model

initialize

New Training Dataset

NER Model, initially trained with the training dataset. Then it will be re-trained with the new training set to evaluate the selection policy.

New dataset initially empty and which will be filled by the instances considered eligible by the Reinforcement learning Agent.

Reinforcement Learning Agent: will evaluate the status and decide whether or not to include the instance under examination in a new dataset

Figure 1: The graphic representation of research and explanation of methodology.

RESEARCH METHODOLOGIES

In this thesis will be proposed:

- A **Data Augmentation** system that performs Mention Replacement on a biomedical dataset provided by the Policlinico di Napoli Federico II.
- **Active Learning** is employed to select the most informative samples from the augmented pool.
- A **Reinforcement Learning** framework automatically learns a selection policy to augment samples during the Active Learning cycles.

ETHICAL ISSUES

Anonymization and sanitization methods are crucial for GDPR compliance.

- **Anonymization** renders data unrecognizable, removing personal identifiers and sensitive metadata.
- **Sanitization** minimizes unnecessary information, including sensitive data.

Eliminating irrelevant details ensures compliance and protects individual privacy.

CONCLUSION

This thesis aims to evaluate the effectiveness and the goodness of policy-based active learning approach for improving Bio-NER model, starting from biomedical Italian dataset and manipulating them via data augmentation, generating a new dataset of noisy samples.

References:

Dai, X., & Adel, H. (2020). An Analysis of Simple Data Augmentation for Named Entity Recognition. ArXiv, abs/2010.11683.

Bartolini, I., Moscato, V., Postiglione, M., Sperlí, G., & Vignali, A. (2022). COSINER: COntext SImilarity data augmentation for Named Entity Recognition. Similarity Search and Applications.

Fang, M., Li, Y., & Cohn, T. (2017). Learning how to Active Learn: A Deep Reinforcement Learning Approach. ArXiv, abs/1708.02383.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Università degli Studi di Napoli
Federico II

Consiglio Nazionale
delle Ricerche

hcaim
human centred
artificial intelligence
masters

GPT-3-based features extraction from EHR of cancer patients

Salzano, Chiara [1], Pallocca, Matteo [2], Marrone, Stefano [1]
[1]University of Naples Federico II, [2] IRCCS Istituto Nazionale Tumori Regina Elena, Rome

Abstract. Research hospitals can count on a large amount of data collected within clinical practice, but these are often stored as free-text, impossible to read by machines. We propose to apply data mining techniques, such as NLP for information extraction to our data. A final objective is to extract features such as patient anamnesis, comorbidities, and performance status, structure them in a database, and ensure data conform to FAIR principles to make it possible to use them for federated research.

Context

The use of **electronic health records** (EHR) is becoming more frequent in the majority of Italian hospitals.

Research hospitals generate a huge amount of data, but

lot of data does not mean lot of information.

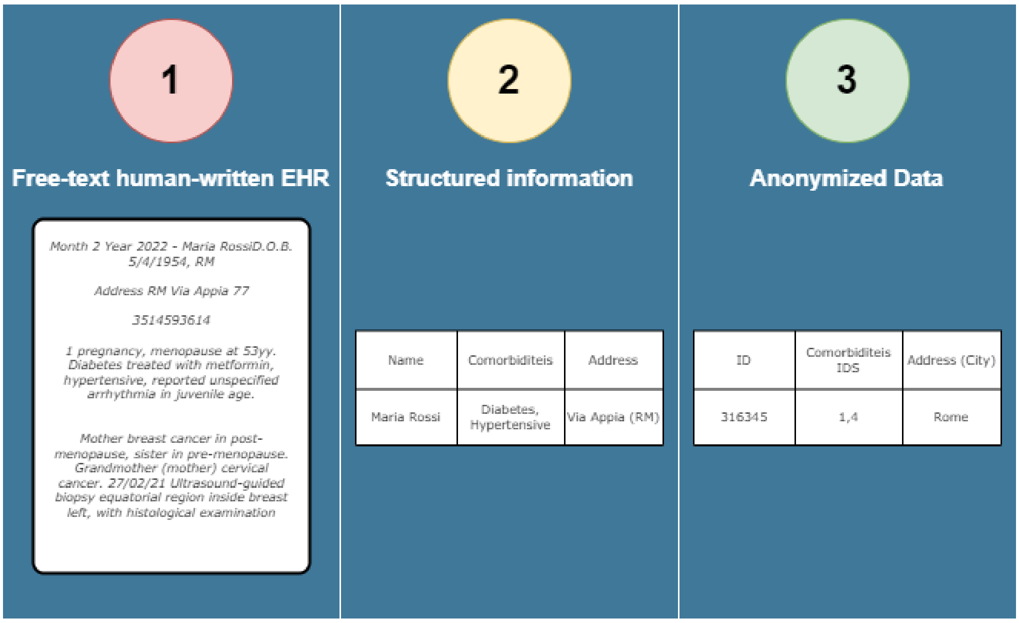
The process of extracting information from data is called **data mining** and today it is possible to achieve this task thanks to **Artificial Intelligence** tools.

From a "machine point of view" it is easier to understand data contained in a **database** instead of interpreting a free-text human-written note.
The desire to have information stored in a structured format is due to the possibility of using it to assess **data-driven decision-making**. Using this method, physicians are able to apply medicine that is patient-centric and tailored to each individual patient (**precision medicine**). [3]

Data

The dataset is a collection of **Real World Data** collected in this hospital in 2018. Three histologies have been selected:

Colon Breast Lung



Month 2 Year 2022 - Maria RossiD.O.B. 5/4/1954, RM

Address RM Via Appia 77

3514593614

1 pregnancy, menopause at 53yy. Diabetes treated with metformin, hypertensive, reported unspecified arrhythmia in juvenile age.

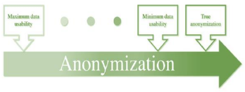
Mother breast cancer in post-menopause, sister in pre-menopause. Grandmother (mother) cervical cancer. 27/02/21 Ultrasound-guided biopsy equatorial region inside breast left, with histological examination

Name	Comorbiditeis	Address
Maria Rossi	Diabetes, Hypertensive	Via Appia (RM)

ID	Comorbiditeis IDS	Address (City)
316345	1,4	Rome

Motivating Use Cases

1. **Information extraction** from EHR
   A. Extract a **timeline** of events from textual records
   B. Extract a minimum dataset composed of patient anamnesis, comorbidities, performance status, etc.



2. **De-identification** of patient records and **Data FAIRirication**
   A. For data to be **Findable**, unique identifiers and indexes must be used.
   B. Accessible data should be stored in a place where access is governed by an open access protocol, and security roles should be implemented.
   C. Data must use vocabularies and metadata in order to be **Interoperable**.
   D. To be **Reusable**, the data should have a clear reusability license.

3. **Patient Search** for cohort discovery:

"I'd like to investigate how many patients are prescribed drug X for condition Y and return to hospital within time Z"

Methods

Natural language processing (**NLP**) refers to the branch of computer science regarding computers' ability to interpret text and spoken words.

Information Extraction (**IE**) is a task of extracting **pre-specified types** of facts from **written texts** or speech transcripts and converting them into **structured representations**. [1]

These techniques will be used in the first and second phases of the work to achieve information extraction and de-identification tasks.

References [1] Ji, H. (2009). Information Extraction. In: LIU, L., ÖZSU, M.T. (eds) Encyclopedia of Database Systems. Springer; [2] Sinaci AA, Núñez-Benjumea FJ, Gencturk M, Jauer ML, Deserno T, Chronaki C, Cangioli G, Cavero-Barca C, Rodríguez-Pérez JM, Pérez-Pérez MM, Laleci Erturkmen GB, Hernández-Pérez T, Méndez-Rodríguez E, Parra-Calderón CL. From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. Methods Inf Med. 2020 Jun;59(S 01):e21-e32. doi: 10.1055/s-0040-1713684. Epub 2020 Jul 3. PMID: 32620019. [3] Bhalla S, Laganà A. Artificial Intelligence for Precision Oncology. Adv Exp Med Biol. 2022;1361:249-268. doi: 10.1007/978-3-030-91836-1_14. PMID: 35230693.

hcaim human centred artificial intelligence masters

SISTEMA SANITARIO REGIONALE
IRCCS
ISTITUTI FISIOTERAPICI
OSPITALIERI

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.

Detecting Fake News using Machine Learning

Roberto Buono

Università degli Studi di Napoli Federico II, Naples

Abstract. Internet is a fundamental tool, whose importance has reached astounding levels, so much so that it has become essential in our lives.

One of its aim is information. Users, in fact, use Internet to inform and communicate with each other on the main available platforms.

However, one of the main problems is that these platforms do not exercise control over the available content. So some users try to spread fake news through these platforms, phenomenon that has reached dimensions out of control.

Fake news includes information presented in a biased or distorted manner, even in a manipulated or decontextualized way, or real information praising cyberbullying or hatred.

A human being may not be able to detect a fake news, and it is therefore necessary the help of machine learning that, thanks to the help of the machine learning classifiers, manages to find all possible fake news present.

Fig. 1. In the Era of Fake News, Teaching Media Literacy is a Must. Illustration by *Let Grow*.

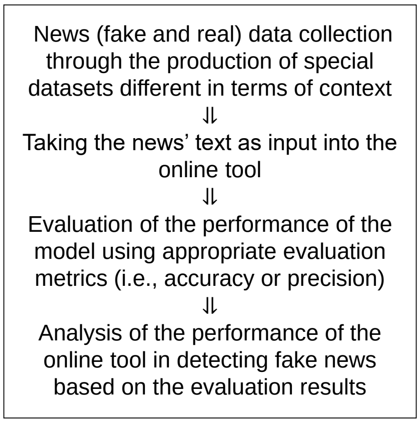
INTRODUCTION

Unfortunately nowadays it is very easy to spread fake news on the major online platforms (e.g. Facebook, Twitter, etc.).

Machine learning is the part of artificial intelligence that helps in making the systems that can learn and perform different actions [1].

The real challenge today is to find fake news [2] due to the fact that this is a difficult task.

Some researchers [3] claim that having recourse to machine learning has been very useful because algorithms of machine learning are trained to fulfil this purpose. Machine learning algorithms could detect the fake news automatically once they have trained.



News (fake and real) data collection through the production of special datasets different in terms of context

⇓

Taking the news' text as input into the online tool

⇓

Evaluation of the performance of the model using appropriate evaluation metrics (i.e., accuracy or precision)

⇓

Analysis of the performance of the online tool in detecting fake news based on the evaluation results

Fig. 2. Methodology.

RESEARCH METHODOLOGY

The objective of this research is to evaluate the effectiveness of an online tool in detecting fake news by analysing textual content using a NLP algorithm.

It's crucial to conduct a comprehensive review of existing literature on fake news detection, NLP algorithms, and online tools, trying to identify the strengths and weaknesses of previous approaches and tools.

Once that an appropriate online tool for fake news detection based on its capabilities, reputation, and availability has been selected (considering factors such as accuracy, scalability, ease of integration, and compatibility with the research objectives), it's important to gather different datasets of news articles from various sources, that includes textual content, ensuring the datasets cover different topics and reflect real-world scenarios, manually labelling the articles as either fake or real to create a ground truth for evaluation.

Subsequently, we will apply the NLP algorithms provided by the online tool to extract relevant features from the textual content.

Lastly we come to the evaluation and the analysis of the performance of the tool in detecting fake news based on the results, identifying strengths and weaknesses of the tool in differentiating between real and fake news.

CONCLUSIONS

Given the increasingly massive use of the internet, fake news are nowadays rampant.

Users can profit by sharing a large number of fake news on major online platforms, that may damage the reputation of some.

The online tool could effectively extract relevant features from the textual content, such as sentiment analysis, linguistic patterns, and contextual clues. These features contributed to a more comprehensive assessment of fake news and improved the detection process.

For future exploration a promising avenue could be the integration of Image Analysis, by exploring the functioning of image captioning with the aim of enhancing the accuracy of the fake news detection.

References [1] Donepudi, Praveen. (2019). Automation and Machine Learning in Transforming the Financial Industry. Asian Business Review. 9. 129-138. 10.18034/abr.v9i3.494.

[2] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

[3] William Yang Wang. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 422–426, Vancouver, Canada. Association for Computational Linguistics.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

A HC approach to abusive language detection: Efficacy of LLMs

Zaur Gouliev | HCAIM TU Dublin | June 2023 | Research Proposal | X00205702@mytudublin.ie | Supervisor: Rajesh Jaiswal, PhD

Abstract

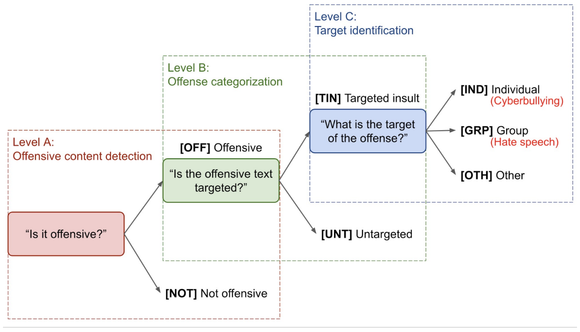


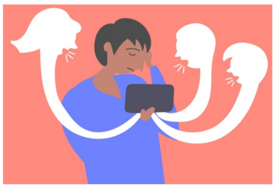
Fig 1: The three-level hierarchical taxonomy for categorizing offensive language, proposed by Zampieri et al. (2019).

This research **examines the efficiency of Large Language Models (LLMs) in detecting abusive language within conversations**. The study focuses on comparing the performance of three prominent LLMs: ChatGPT LLM, Facebook LLM, and Google's BERT/BART LLM. Our hypothesis (H_0) is that LLMs do not perform optimally and lack human alignment in identifying abusive language, while the alternative hypothesis (H_a) proposes that LLMs perform well in this task. The research further explores the architecture and machine learning techniques employed by each LLM and how this may affect abusive language dialogues. It also considers the social and ethical differences inherent in online conversations. Additionally, the explainability (XAI) of LLMs is investigated to gain insights into their decision-making processes and potential limitations in detecting this type of language. By analysing the explainabilty of LLMs, we can get a deep understanding of how LLMs and their inheirant architecture can be used to decrease harmful language online through the use of abusive language detection. Public datasets such as OffensEval (2019), ConvAbuse (2022) will be used in this research paper.

Background

The detection of abusive language in online conversations is a vital problem in society that requires attention. With the rapid growth of online platforms and social media, there has been an alarming increase in the prevalence of abusive language, hate speech, and cyberbullying. Such behaviour not only negatively impacts individuals but also undermines the inclusivity, safety, and well-being of online communities. Consequently, there is a pressing need for effective tools and strategies to identify and mitigate abusive language in real-time. Abusive language detection has received extensive attention for social media, but far less within the context of conversational systems.

Several works have investigated the application of machine learning algorithms and different features extraction techniques to define automatic approach for abusive language detection. Currently, the state of the art in abusive language detection primarily relies on the application of Natural Language Processing (NLP) techniques, particularly utilizing machine learning approaches such as Large Language Models (LLMs).



Research Questions

In the field of social media and online content moderation, it is important to acknowledge the existence of numerous forms of unsafe content, such as toxicity, abusiveness, hate speech, biases, stereotypes, cyberbullying, and identity attacks.

This means we recognize that these distinct content types may necessitate diverse approaches for effective mitigation. Moreover, it is worth noting that there is currently no universally accepted framework or consensus regarding the classification and characterization of unsafe behavior within pretrained language models. The absence of an established categorization is compounded by the fact that individual interpretations of such behavior can vary significantly due to differing social contexts and backgrounds. This prompts the research questions as follows:

1. To test our hypothesis on the efficacy of LLMs in detecting abusive language through the objective of comparing the performances.
2. Exploring the explainability (XAI) of these LLMs and understand the decision-making processes behind their identification of abusive language.
3. To investigate the different social and ethical differences in how abusive language detection models and algorithms work and their subsequent strengths and weaknesses.

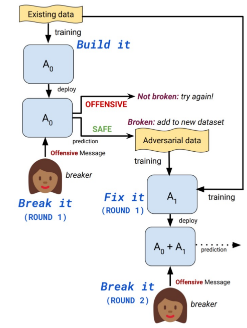
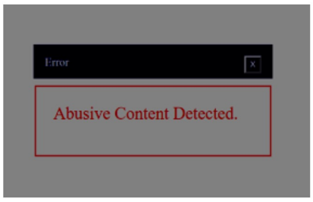
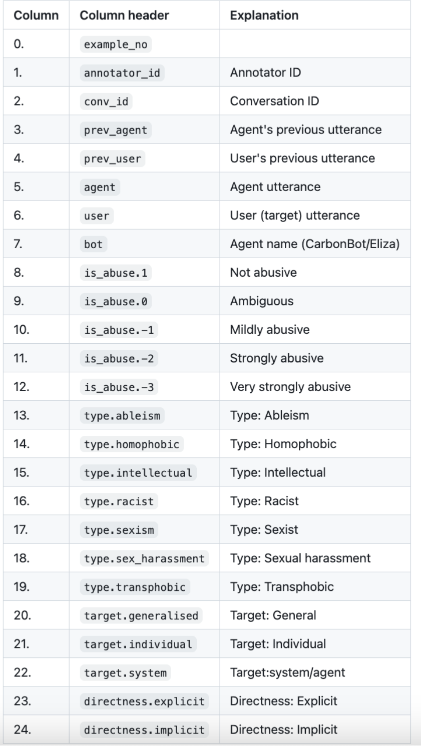


Fig 4: The illustration of iteratively improving a toxic content detection model via the "build it, break it, fix it" process. (Image source: Dinan et al. 2019)



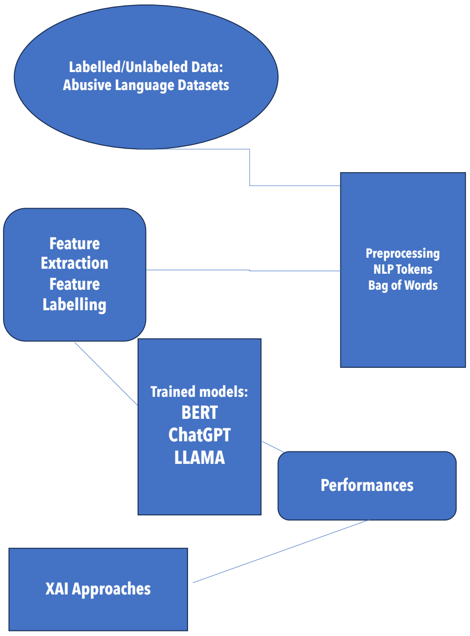
Abusive Content Detected.



Column	Column header	Explanation
0.	`example_no`	
1.	`annotator_id`	Annotator ID
2.	`conv_id`	Conversation ID
3.	`prev_agent`	Agent's previous utterance
4.	`prev_user`	User's previous utterance
5.	`agent`	Agent utterance
6.	`user`	User (target) utterance
7.	`bot`	Agent name (CarbonBot/Eliza)
8.	`is_abuse.1`	Not abusive
9.	`is_abuse.0`	Ambiguous
10.	`is_abuse.-1`	Mildly abusive
11.	`is_abuse.-2`	Strongly abusive
12.	`is_abuse.-3`	Very strongly abusive
13.	`type.ableism`	Type: Ableism
14.	`type.homophobic`	Type: Homophobic
15.	`type.intellectual`	Type: Intellectual
16.	`type.racist`	Type: Racist
17.	`type.sexism`	Type: Sexist
18.	`type.sex_harassment`	Type: Sexual harassment
19.	`type.transphobic`	Type: Transphobic
20.	`target.generalised`	Target: General
21.	`target.individual`	Target: Individual
22.	`target.system`	Target:system/agent
23.	`directness.explicit`	Directness: Explicit
24.	`directness.implicit`	Directness: Implicit

Fig 3: An example of one of the many datasets we will train test and train our model on (Implicit Hate Speech Dataset, 2022)

Methodology



Labelled/Unlabeled Data:
Abusive Language Datasets

Preprocessing
NLP Tokens
Bag of Words

Feature Extraction
Feature Labelling

Trained models:
BERT
ChatGPT
LLAMA

Performances

XAI Approaches

Fig 2: This methodology process of the proposed paper

We will be comparing the performance of three prominent LLMs (ChatGPT LLM, Facebook LLM, and Google's BERT/BART LLM) in detecting abusive language. Our dataset sources will utilise ConvAbuse, Wikipedia Abusive Conversations and other public datasets.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant **№**CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Multi-modal Fake News Detection on Twitter through User's Social Network

Wan Yit Yong, X00205680@mytudublin.ie, Research Proposal

Human Centred Artificial Intelligence Master's, School of Computing, Technological University Dublin, Ireland

Abstract

Fake news is a serious threat to the credibility and trustworthiness of news on social media platforms. Detecting fake news is challenging, especially when it involves multiple modalities such as text, image, and video. This research aims to **develop a framework for multi-modal fake news detection by leveraging the user's social network context**. The hypothesis suggests that the user's social network can provide useful cues for identifying the veracity and reliability of the news content, as well as the intention and reputation of the news sources. The research will use multi-model approaches to extract features and apply mutual attention to fuse them with social network features, then evaluate on a large-scale Twitter dataset provided by the industry partner. The expected contributions are: (1) a comprehensive analysis of the impact of different modalities and social network features on fake news detection, and (2) a multi-modal fake news detection framework that incorporates user's social network context.

Background

Fake news is a term that refers to false or misleading information that is presented as factual news, often with the intention of influencing public opinion or gaining political or financial advantage. Fake news can have serious consequences for individuals and society, such as eroding trust in institutions, sharing misinformation on health issues, and polarizing opinions. Fake news is not a new phenomenon, but it has become more prevalent and pervasive in the era of social media, where anyone can create and share content with a large audience, without any editorial oversight or accountability.

Social media platforms like Twitter also enable the use of multimedia content to enhance the credibility and appeal of fake news. Moreover, Twitter users can form echo chambers and filter bubbles, where they are exposed to information that confirms their existing beliefs and biases and avoid information that challenges them. Therefore, it is crucial to develop effective methods for detecting and combating fake news on social media platforms.

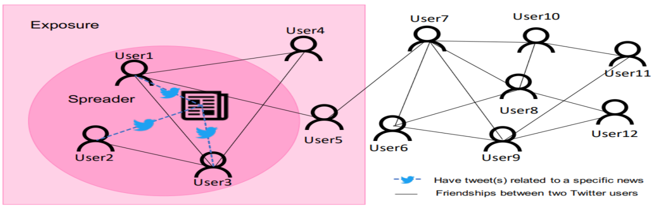


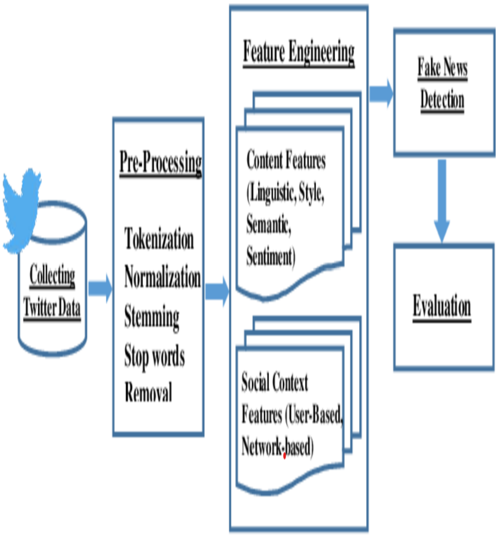
Figure: An example showing that users are connected on the Twitter platform.

Research Question

The main research questions for this study is how can multi-modal analysis and social network context be leveraged to improve fake news detection on Twitter. Besides that, there are 3 subordinate research questions as well:

- What are the distinctive features and patterns associated with fake news on social media platforms, considering both textual and visual content?
- How can the social network context, including user interactions and network structure, contribute to the identification and understanding of fake news propagation patterns?
- How can graph neural network models be applied to capture the complex relationships and dynamics within a user's social network, enhancing the accuracy of fake news detection?

RESEARCH METHODOLOGY



Dataset:
"PolitiFact" provided by industry partner, collected from Twitter.

Preprocessing:

- Text data: BERT model
- Image data: EfficientNetBO model
- User profile data: BERT model
- Graph Representation: Tweets as nodes, user relations as edges

Model Construction:

- Graph Neural Network
- Graph Convolutional Network for feature aggregation

Model Evaluation:

- Accuracy, Precision, Recall, F1-score
- Analyze the impact of modalities and users on fake news detection

PRELIMINARY CONSIDERATIONS

Potential Societal Impact: The research aims to mitigate the harmful effects of fake news by developing effective detection techniques. By identifying and flagging false information, the research can help to create a healthier information ecosystem, promote informed decision making and reducing the spread of misinformation. As a result, people may trust online information sources and society as a whole will become better informed.

Advancements in ICT: The research has implications for the field of information and communication technology (ICT). The development and use of graph neural network-based methods for fake news detection can advance the state-of-the-art in machine learning and natural language processing. Beyond fake news identification, these developments may also be used for sentiment analysis, content suggestion, and customized user interfaces. By pushing the limits of existing procedures and techniques, the research can help progress ICT.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

References.

[1]. Caroprese, Luciano & Comito, Carmela & Zumpano, Ester. (2023). Fake News on Social Media: Current Research and Future Directions. 10.1007/978-3-031-31469-8_4.

[2] Alam, Firoj & Cresci, Stefano & Chakraborty, Tanmoy & Silvestri, Fabrizio & Dimitrov, Dimiter & Martino, Giovanni & Shaar, Shaden & Firooz, Hamed & Nakov, Preslav. (2021). A Survey on Multimodal Disinformation Detection.

DATA AUGMENTATION APPROACHES FOR LIP MOVEMENTS DETECTION

Daniele **Iuliano**
Università degli Studi di Napoli Federico II and CNR (Consiglio Nazionale delle Ricerche)

ABSTRACT

In this work, Data Augmentation approaches are considered to cope with a small and unbalanced video Dataset for improving lip movements detection in a legal context.

INTRODUCTION

Deep Convolutional Neural Networks have performed remarkably well on many Computer Vision tasks.

These networks typically rely on large amounts of training data to avoid overfitting. However, good quality labeled data for real-world applications may be limited. Data Augmentation can alleviate the problem, generating new data from a smaller initial Dataset.

Data Augmentation algorithms are usually specifically designed for single images. Recently, Deep Learning models have been applied for the Data Augmentation of video sequences [1].

RESEARCH METHODOLOGIES

The aim of this work is the evaluation of how different Data Augmentation approaches can improve the performances of Deep Neural Networks for real-life lip movements detection applications.

After a survey on Image [2] [3] [4] and Video [1] Data Augmentation techniques for Deep Learning, an analysis of the strengths and limitations of these methods is presented.

The most basic technique of Data Augmentation for images is noise injection: the Dataset is expanded creating duplicates of the original images injected with random values in the color space. Other common techniques are geometric transformations, cropping, flipping, rotating, translating, histogram alteration.

With the improvements in Neural Networks, more advanced Data Augmentation methods have been introduced.

Strategies based on generative modeling are able to create new input belonging to a similar distribution of the original Dataset. These strategies use Generative Adversarial Networks (GANs) to generate the new data. Neural Style Transfer is another DL based methodology able to augment the size of image Datasets. A common trait of all these methodologies is the use of images from the original Dataset as a base for generating the new images. A different approach is to generate the images for the augmented Dataset from physical models that approximate the world.

Data Augmentation of Video Datasets adds the temporal dimension to the images problem, resulting in a very complex challenge. In fact, in order to apply standard image Data Augmentation methods to video sequences, the time domain needs to be taken in consideration and the changes applied to each frame must be coherent through time.

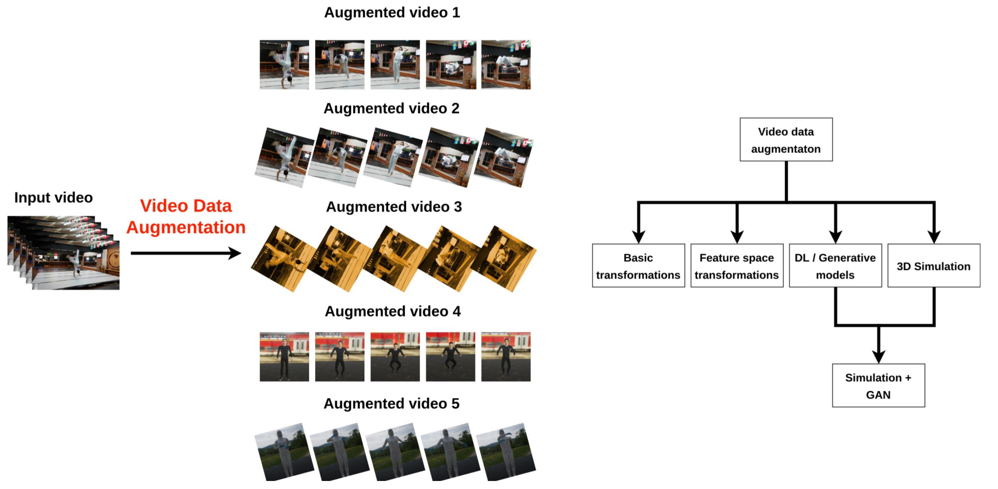


Figure 1: Graphical Abstract. [1]

In the Data Augmentation for Video Dataset, time series of images are usually organized in mini batches representing short clips. To guarantee a time coherence, geometric and color transformation must remain constant through the entire mini batch, and they can be randomized over different mini batches. In other scenarios, 2D motion models can be used to extend static image Data Augmentation methods to videos.

Other Data Augmentation approaches use DL models. DL models that perform better on videos are those that are able to keep a memory of the previous frames to generate the new ones. The past frames contain information about temporal variations in the scene, like object motions, dynamic light changes, and weather evolution, among others. RNN are widely used for the analysis of text and time series due to their ability to retain a memory of the past inputs through their internal loops. Recently, 1D RNN models (i.e., LSTM and GRU) have been integrated to CNN to perform video analysis and generation. Another approach used to analyze image temporal sequences is the use of 3D convolutions. In this case, the third dimension is used to stack several contiguous frames to obtain temporal information. Extending generator networks with a time series specific model like 3D convolutions or RNN is a promising solution. [1]

ETHICAL CONSIDERATIONS

The Dataset used in this work consists of video of speakers that vary in appearance, gender, skin tones, accents, glasses, facial hair, age and therefore represent a diverse sample. However, some imbalances still exist in the learning domain that need to be addressed. Thus, in order to fulfill the requirement #5 “Diversity, Non-discrimination and Fairness”, as specified in the Assessment List of the Ethics Guidelines for Trustworthy AI [5] presented in 2019 by the High-Level Expert Group on Artificial Intelligence (AI HLEG), Data Augmentation approaches should take in account also this aspect.

CONCLUSION

Various Data Augmentation techniques will be investigated and compared to improve the performance and to avoid unfair Bias (lack of diversity, non-representativeness) in an AI system used to detect lip movements in a real-life legal scenario.

References:

[1] Cauli, N.; Reforgiato Recupero, D. Survey on Videos Data Augmentation for Deep Learning Models. Future Internet 2022, 14, 93. https://doi.org/10.3390/fi14030093

[2] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[3] Suorong Yang, Weikang Xiao, Mengcheng Zhang, Suhan Guo, Jian Zhao, Furao Shen Image Data Augmentation for Deep Learning: A Survey, arXiv:2204.08610 https://doi.org/10.48550/arXiv.2204.08610

[4] C. Khosla and B. S. Saini, "Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 79-85, doi: 10.1109/ICIEM48762.2020.9160048.

[5] European Commission, Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI, Publications Office, 2019

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

MLOps: Using machine learning in the industry

Balázs Tibor **Morvay**[1]

1. balazsmorvay@yahoo.com, Department of Automation and Applied Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary.

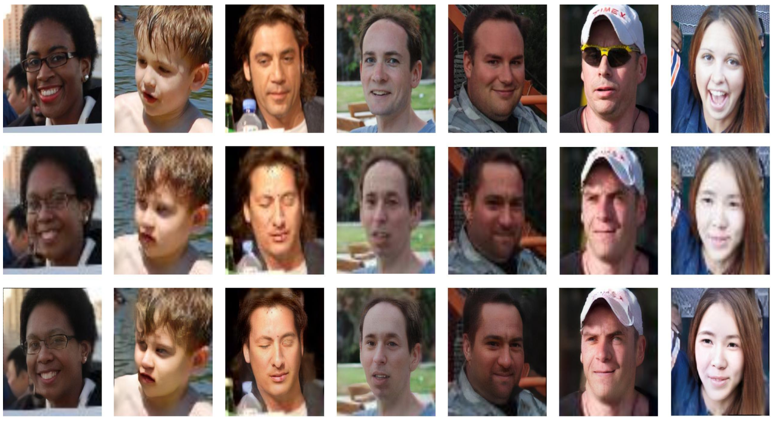


Fig. 2.Examples of Facediffusion anonymized images. The top images are the original ones, the middle row images are generated using the DDPM model, and the bottom row images are the super-resolved images.

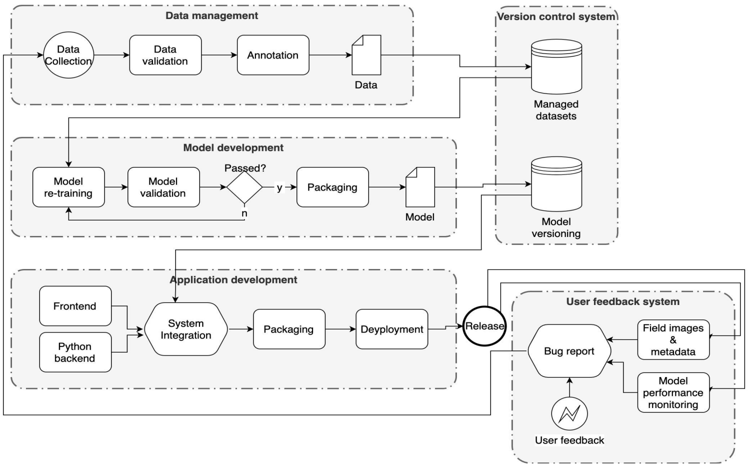


Fig. 1. The system architecture.

Abstract.

Machine learning (ML) models are starting to become widespread. This brings new opportunities and enable us to make algorithms that in some cases outperform humans. But it also brings new challenges.

In my thesis, I explored the field of ML and MLOps through an international industrial project. During this project, I designed and implemented parts of a **machine learning system** and used MLOps tools in the process. I also ported and deployed a machine learning model to an iOS application I made. For the AI system, I made a **risk assessment** and covered a bunch of **privacy-related questions**.

Introduction

The goal of my thesis was to examine the current state of MLOps through a real, industrial project, which I joined through my university (Budapest University of Technology and Economics). The project's name was IML4E[1]. It set industrial-grade ML integration as its main objective. This is an ITEA project that groups universities and experts from multiple countries (Germany, Finland, Hungary, etc.) and involves big corporations too (e.g. Siemens). Each country had a separate national project, and IML4E joined these into a European, large-scale project. The Hungarian consortium's goal involved making an ML system that can recognize and analyze children's posture, thus helping the work of school nurses, and allowing a more robust and precise diagnosis of spinal diseases. I contributed to the project from the beginning, in its early planning phase, to its implementation and release phase.

My contribution to the project was three-fold:

1. I implemented the backend component responsible for data management part of the whole system. The system architecture can be seen on Fig. 1.
2. Together with my colleagues, we conceived and implemented an anonymization pipeline called Facediffusion.
3. I ported an ML model to iOS using the Core ML framework to enable on-device model inference.

Backend system

I wrote the backend system in the FastAPI[2] framework in Python language. Its primary task was to expose three endpoints where data from three sources could arrive. Then, process the incoming data, perform validation, transformations, and save the resulting data in files. As these data were intended to use as a dataset to further train the pose estimation model of the system, the files were version controlled using DVC, to enable reproducible experiments.

Facediffusion

The Facediffusion[3] pipeline constructs 256x256 resolution anonymized images from matching resolution images. The pipeline first uses our Denoising Diffusion Probabilistic model to generate 64x64 resolution synthetic images with facial keypoint keeping. Then, a blind face restoration network, DFDNet[2], upscales the images to 256x256 resolution. Some results of Facediffusion can be seen on Fig. 2. We also measured the pipeline's performance using different metrics, and the results are promising.

iOS application

I ported the ESPCN[4] image super-resolution model to iOS using the Core ML framework. This enables iOS devices to perform on-device inference, which is a privacy-friendly ML setup, as sensitive data never leaves the device, and there is no need for an internet connection. I also measured the model's performance.

HCAI elements

At the core of the resulting AI system is privacy. Privacy was mainly achieved using a face anonymization system. Other efforts were made to make a differentially private pose estimation network, to preserve the privacy of dataset members.

I also categorized the finished system into the limited-risk AI category by eliminating the unacceptable and high-risk category: Our system clearly does not pose such danger to be considered as unacceptable risk, and it is not a safety-critical system, does not operate in a field where human rights are particularly affected, and the other high-risk properties are also not applicable. Therefore, I classified it as limited-risk category.

References.

[1] https://iml4e.org

[2] https://fastapi.tiangolo.com

[3] https://github.com/balazsmorv/facediffusion

[4] Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Application of Semantic Segmentation

Mátyás Pelle[1], Gábor Sörös[2], Luca Szegletes[1]
1 Budapest University of Technology and Economics, Budapest, Hungary | 2 Nokia Bell Labs, Budapest, Hungary

Abstract. The task is to gain deep knowledge in the field of semantic segmentation, which means exact pixel- or voxel-level delimitation and classification of objects, with the main goal to create a semantic map of an indoor space.

For explainability, we plan to deeper investigate the operations of a semantic segmentation graph neural network, to visualize its internals, and to understand why something goes wrong, and how an existing model could be supplemented with new objects.

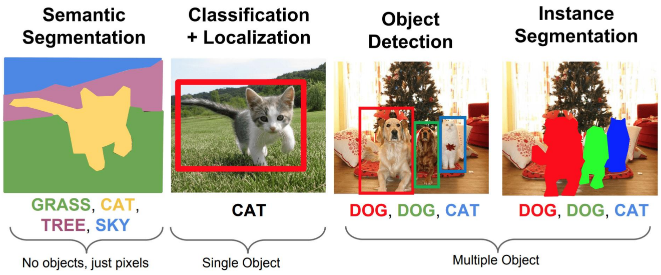


Fig. 1. Different types of visual image recognition [13]

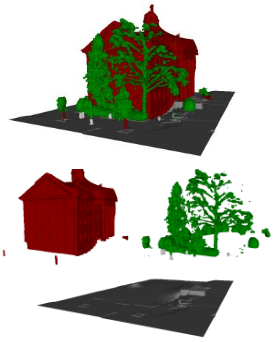


Fig. 2. 3D semantic reconstruction [14]

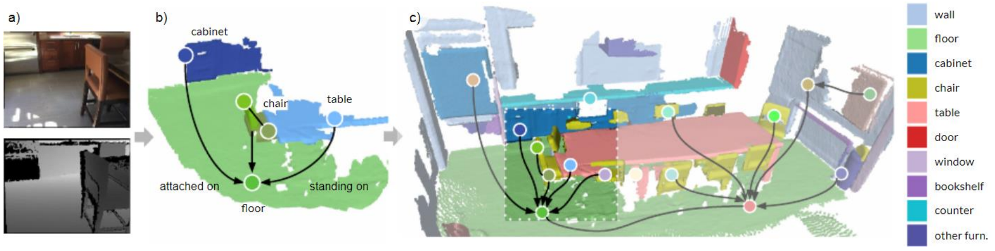


Fig. 3. SceneGraphFusion [9] builds a scene graph

INTRODUCTION

Deep learning based image processing techniques have made significant progress in the last decade and accordingly achieve outstanding results in classification, object detection and segmentation. The goal of this thesis project is to create a semantic map of an indoor space for robots and other smart devices.

Tasks to be performed:

- Group the most important directions of visual image recognition:
  - Classification
  - Detection
  - Segmentation
- Describe the most important algorithms and models in these fields
- Introduce CNN (Convolutional neural network), GNN (Graph neural network)
- Describe semantic segmentation, discuss state-of-the-art approaches
- Describe the concept of a semantic map and the possibilities of its implementation
- Create a 3D semantic map of an indoor environment

HUMAN-CENTRIC AI ASPECTS

- Explainable AI - Can we explain why a certain prediction result came out? Can we visualize partial results and the internal workings of a semantic segmentation neural network?
- Privacy - personal sensitive data must not be included during the creation of the semantic map
- Security - Data protection is important, e.g. store the maps locally in a safe way. Map can leave the robot (should be shared anyway), but never leave the local network. Verify who created and who modified the map
- Trustworthiness - e.g. How can we know whether a semantic map can be trusted? How can we verify a map? What is in the map is really what is there?

LITERATURE REVIEW

It is very important to learn the basics, we need to familiarize ourselves with the given topic. In order to learn the basics, we performed a deep literature review.

- Object detection:
  - Fast R-CNN [1]
  - Faster R-CNN [2]
  - YOLO [3]
- Semantic segmentation:
  - Mask R-CNN [4]
  - YOLACT, YOLACT++ [5]
  - ESANet [6]
- 3D semantic segmentation:
  - Kimera [7]
  - Hydra [8]
  - SceneGraphFusion [9]

SIMULATORS AND DATASETS

We perform our experiments in simulated environments for repeatability and for access to ground truth. Not only the methods, but also the data are very important:

- Matterport3D [10]
- ScanNet [11]
- 3RScan [12]

PRELIMINARY CONSIDERATIONS

The semantic map must be good, because if it is not suitable, then the robots will not be able to perceive their environment well, which can lead to many problems (incorrect obstacle avoidance, wrong object fetching).

CONCLUSIONS

The goal of this thesis is to try and compare different approaches for semantic mapping. We would like to combine the benefits of the segmentation neural network and the geometric reconstruction with graph neural network structure. We hope that the results of graph neural networks are more explainable than traditional deep methods.

References.
[1] Ross Girshick; Fast R-CNN; 2015
[2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun; Faster R-CNN; 2016
[3] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi; YOLO; 2016
[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick; Mask r-cnn; 2017
[5] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee; Yolact; 2019
[6] Daniel Seichter, Mona Köhler, Benjamin Lewandowski, Tim Wengefeld, Horst Michael Gross; Efficient rgb-d semantic segmentation for indoor scene analysisa; 2021
[7] Antoni Rosinol, Marcus Abate, Yun Chang, and Luca Carlone; Kimera; 2020
[8] Nathan Hughes, Yun Chang, and Luca Carlone; Hydra; 2022
[9] Shun-Cheng Wu, Johanna Wald, Keisuke Tateno, Nassir Navab, and Federico Tombari; Scenegraphfusion; 2021
[10] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang; Matterport3d; 2017
[11] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner; Scannet; 2017
[12] Johanna Wald, Armen Avetisyan, Nassir Navab, Federico Tombari, Matthias Niessner; RIO: 3D Object Instance Re-Localization in Changing Indoor Environments; 2019
[13] Rahul Agarwal; Object Detection: An End to End Theoretical Perspective
[14] Christian Häne and Marc Pollefeys; An overview of recent progress in volumetric semantic 3d reconstruction; 2016

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.

Mapping and navigation based on Neural Radiance Fields

Ágoston Csehi[1], Dr. György Józsa[2], Dr. László Lengyel[1]
[1]Department of Automation and Applied Informatics, BME, Hungary | [2]Nokia Bell Labs, Budapest, Hungary

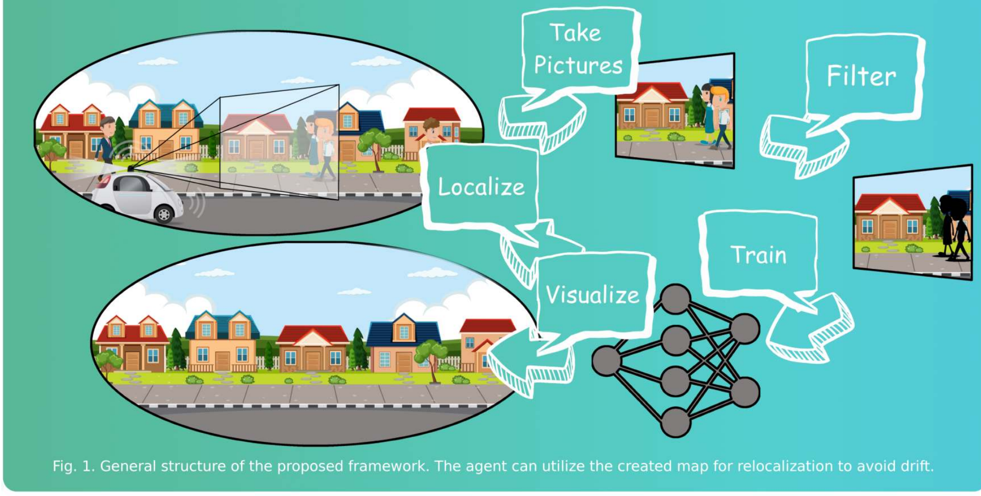


Fig. 1. General structure of the proposed framework. The agent can utilize the created map for relocalization to avoid drift.

Abstract

We investigate how Neural Radiance Fields could be used in SLAM problems as dense base space representations. The introduced framework implements the whole SLAM pipeline with the necessary modifications to be compatible with the suggested neural model. With this approach, we aim to achieve an efficient dense map of the environment, which could be used for localization to solve robotics navigation problems. Another possible use case would be visualization, which is essential for human inference.

Introduction

As we move, we constantly update our beliefs about our environments. This process incorporates many different types of sensory information. Even though a small child is able to solve this task inadvertently, the steps involved are rather complex to reproduce algorithmically. One of the many occurring challenges is the storage requirement for the learned maps. Since the popularization of neural networks, several NN-based space representation techniques have emerged. These might provide a more concise way to store all the necessary information about the surroundings for robotics applications.

One such novel approach is to use Neural Radiance Fields (NeRF). Being a disruptive technology, it is the focus of bleeding-edge research, trying to figure out its potential use cases. Previous work has already shown that NeRFs can be used not only for mapping but also during localization.

By having a compact, NN-based solution for simultaneous localization and mapping (SLAM) tasks, we could improve the safety of human-robot interactions in industrial or home applications, reduce costs by lowering the inherent storage requirements and simplify the needed sensors while maintaining a space representation, which can be easily visualized for verification and inference.

We expect that this approach can solve many problems that currently used SLAM solutions struggle with, like behavior in featureless and repetitive environments or the question of scalability.

Literature Review

NeRFs (Mildenhall et al., 2020) are neural networks representing color and density information for each point and view direction in a continuous bounded space. Through volumetric rendering based on these models, complete images can be synthesized for any given pose.

INeRF (Yen-Chen et al., 2020) takes advantage of the differentiability property of such model's to approximate the pose of images. This way, the direction of the original model's mapping is inverted so that they can be utilized for localization purposes as reference space representations.

Block NeRF (Tancik et al., 2022) shows how entire cities can be represented as a bundle of several NeRF models, enabling large-scale applications.

Research Methodology

Combining the referenced ideas, we can create a complete SLAM algorithm based on a scalable dense space representation. The research will be conducted during the spring of 2023.

First, a proof-of-concept version will be implemented to evaluate the idea against state-of-the-art solutions, such as ORB-SLAM2 (Mur-Artal et al., 2016).

Both synthetic and real-life environments will be tested concerning their storage requirements, performance, precision, and robustness. For real-life scenarios, drones or autonomous robots will be used to capture images.

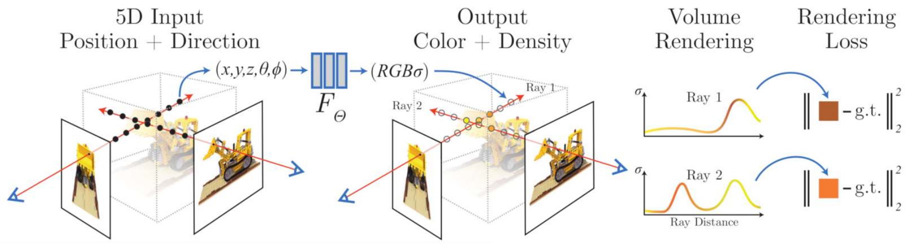


Fig. 2. NeRF algorithm (Mildenhall et al., 2020). To render colors for each ray, discrete points in the space are sampled where the neural network is evaluated. With volumetric rendering, we assign weights to each sample point, representing the probability of ray termination. Then, we get the final predictions in a form of a simple weighted sum of the color values.

Preliminary Considerations

Before any widespread applications, it is crucial to address all the arising undesirable implications. Dense maps of the world inherently contain sensitive information, many of which have to be obfuscated. Fortunately, NeRFs are trained using images, which gives us the opportunity to introduce filtering with the application of common computer vision techniques. This way, specific parts of the represented space can be omitted from the created map.

Although, during the research, only proof-of-concept implementations will be created, preventing the development of malicious versions htas to be considered. Misuses of the introduced ideas are possible, leading to a loss of civil privacy or even safety. With this in mind, final implementations for large scale environments will not be published.

Conclusions

By conducting this research, we expect to discover a novel use case for NeRFs while introducing an adequate alternative for state-of-the-art SLAM solutions. With the usage of neural space representations, we can tackle the inherent problem of storage limitations and the lack of visualization capabilities.

With the incorporation of the results from cutting-edge research, it will be possible to construct maps of even cities.

References

- Yen-Chen, L., Florence, P., Barron, J. T., Rodriguez, A., Isola, P., & Lin, T.-Y. (2020). INeRF: Inverting Neural Radiance Fields for Pose Estimation.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.
- Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P. P., Barron, J. T., & Kretzschmar, H. (2022). Block-NeRF: Scalable Large Scene Neural View Synthesis.
- Mur-Artal, R., & Tardos, J. D. (2016). ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras.

Acknowledgements.
The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.
The research will be conducted with the help of Nokia Bell Labs, Budapest.

Digital Twins for Personal Health

Improving well-being on a mass scale

Suciu Barnabás, BME Budapest | Alireza Dehghani, CeADAR Dublin

Abstract

In this research proposal poster the novel field of digital twin personal well-being assistants are explored. I outline the possible societal implications of this technology, its benefits, and the concerns it may raise to certain user groups. A literature review is conducted, and plans are developed for a system in which continuously improving machine learning models generate advice for the user and receive feedback for further improvement.

Introduction

In our increasingly busy society, neglecting personal health is becoming a more and more common phenomenon. What if there was a technological solution to the problem? By collecting data about yourself, a digital twin could make recommendations about lifestyle and how to improve your own well-being. By designing a continually learning, intelligent, trustworthy companion, great improvements could be made for the person conscious about their health, longevity and performance. The capability to collect, aggregate and process different kinds of personal data is a major challenge in the development of these systems. This work focuses on the exploration of ideas relating to this project, both from a data science and ethics perspective.

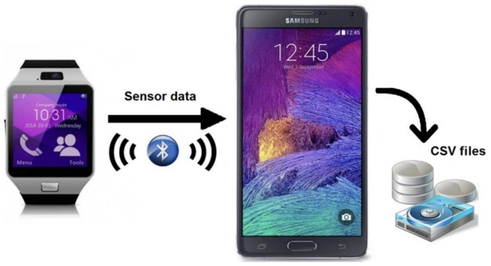


Figure 2: Data collection methodology for wearable device sensor data

Literature Review

While the field is quite novel, there are some worthy mentions when it comes to previous research. Digital twin technology has become a popular idea in recent years, a good demonstration can be found in [1]. Soma approaches, like [2], focus only on certain aspects of the human body, demonstrating the potential in these systems. Using wearable device health data is a crucial step for the success of a system like this, an example of this is outlined in [3]. Overall, the field is highly interdisciplinary, and it is still in the early stages of development. While there is great potential in the technology, lots of challenges have to be overcome before the solution can be adopted generally.

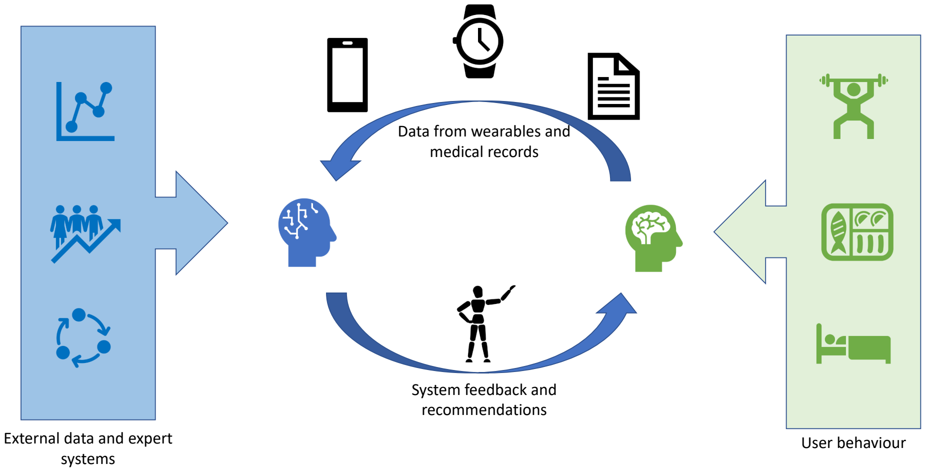

Data from wearables and medical records
System feedback and recommendations
External data and expert systems
User behaviour

Figure 1: Collenction and synthesis of personal health data of the user.

Methodology

In terms of data science, the following challenges require solving before a system like this can be implemented:

- Ethical and private data collection and storage
- Machine learning model selection based on performance
- Continuous and active learning
- Feedback incorporation

The technical aspects require further research; however, two potential approaches are logical at this stage. In the first phase of the implementation, a basic recommendation engine is developed using supervised learning on wearable device data. This model is then used to develop an application that can be distributed to users. From this stage, live data could be collected, and this can be used in combination with user feedback to personalize the model, create recommendations and improve performance.

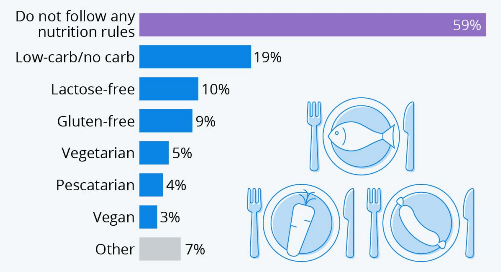


Figure 3: Dietary habits of the American population. Dietary recommendations are an important part of a healthy lifestyle.

Social Impact

From a human-centered perspective, personal medical advice has many implications which have to be taken into account. The collection of private data raises a lot of concerns, and in this case, the more personal the data, the better. However, if the benefits outweighed the trust concerns, the technology could revolutionize the lives of people. This technology can lead to improvements in well-being for a large number of people in several ways:

- **Personalized recommendations:** The digital twin can analyse data from various sources, such as fitness trackers, medical records, and dietary habits, to provide recommendations tailored to an individual's unique needs and goals. This can lead to more effective and sustainable changes in lifestyle, compared to generic recommendations.
- **Continuous monitoring:** The digital twin can continuously monitor an individual's health and lifestyle, providing real-time feedback and alerts to help them stay on track and make necessary adjustments. This can help individuals avoid health issues and improve well-being over time.

Conclusion

Overall, there is great potential in personal health recommender systems to improve the well-being and longevity of people. The use of digital twin technology has great potential in the field, as seen from the possibilities outlined in this proposal. The research and development is to be broken up into two phases, at the end of which a continuously learning agent is developed.

References

[1] Barricelli, B. R., Casiraghi, E. & Fogli, D. A. survey on digital twin: definitions, characteristics, applications, and design implications. *IEEE Access* **7**, 167653–167671 (2019).

[2] Coorey, G. et al. The health digital twin to tackle cardiovascular disease-a review of an emerging interdisciplinary field. *NPJ Digit. Med.* **5**, 126 (2022).

[3] King, R. C. et al. Application of data fusion techniques and technologies for wearable health monitoring. *Med. Eng. Phys.* **42**, 1–12 (2017).

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.

Human-Centred Regulated Artificial Intelligence (HCRAI), BME, Budapest 2023

Balancing Innovation and Ethics: AI Digital Twins in Remote Working Hubs

Yilin Li, Alireza Dehghani
CeADAR, Ireland's Centre for Applied AI, UCD

Abstract

This work explores the ethical considerations when implementing AI Digital Twins (AIDT) for remote working hubs (RWHs). As the remote working lifestyle continues to grow, innovative solutions like AIDT have the potential to optimise the remote working experience, improve resource allocation, motivate the city's green transition, and provide valuable insights for RWH operators and city planners. However, the implementation of this technology raises important ethical considerations.

This poster reviews current literature and research findings on the ethical aspects of AIDT implementation, including user consent and transparency, data security and privacy, equity and bias, sustainability, user autonomy, and accessibility and inclusiveness. The aim is to highlight the importance of addressing these ethical considerations when developing and using RWH's AIDT, building trust with users and ensuring the sustainability of AIDT.

Introduction

The rapid growth of remote working has created a demand for innovative solutions that optimise the remote working experience, improve resource allocation, trigger the green transition towards zero net, and provide RWH operators, their users, and urban planners with valuable insights. One such innovative solution is using AIDT, virtual copies of physical entities that can simulate, predict and optimise the functionality of their physical counterparts. In the context of RWH, AIDT can effectively monitor, analyse, and optimise the use of these facilities by processing users' real data and synthetic data that are closely related to real-world usage patterns as shown in **Figure 1**. However, implementing AIDT in RWH raises important ethical considerations that need to be addressed to ensure the successful and responsible use of this technology.

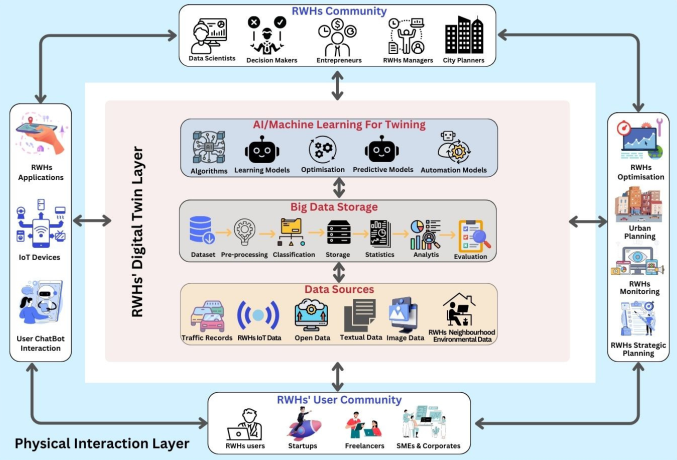


Figure 1: The structure of RWH AIDT

> " In the pursuit of innovation, we must not lose sight of our ethical responsibilities. Implementing AI digital twins in remote working centres is not only about optimising resources, taking care of the green transition and improving the user experience but also about respecting users' rights and ensuring their privacy.

AIDT Ethics Matter

Ethical considerations when implementing AIDT for RWH were the subject of several research papers and articles. Rainey [1] explored the ethical aspects of brain digitisation through a digital twin case study, highlighting the importance of user consent and transparency, data security and privacy, and fairness and bias. Macnish discussed how to operationalise digital twin ethics, highlighting the need for sustainability, user autonomy, accessibility and inclusiveness [2]. BIM+ argues that the successful adoption of digital twins lies in the understanding of ethics, reinforcing the importance of the ethical considerations discussed by Rainey and Macnish [3]. Huang, Lu and Chen provide a preliminary mapping of digital twins for personalised healthcare [4], a preliminary mapping study on the ethical issues of AIDT provides a process-oriented ethical mapping, and these can be used to identify and address potential ethical issues in the use of AIDT by RWHs. Finally, the UK Centre for Digital Architecture [5] presents the Gemini Principles that guide the development and use of information management frameworks and national digital twins, highlighting the value of data and the need for a common set of definitions and principles.

AIDT Ethical Considerations

Implementing AIDT for RWHs brings up several important ethical concerns that need to be carefully addressed. Ensuring user agreement and transparency is most important in the implementation of AIDTs. Users should be careful about the methods used to collect, use, and secure their data. This includes giving detailed explanations of the methods used to collect the data, the types of data that will be collected, and the expected applications of the data. Additionally, users should be allowed to refuse to collect data if they so choose. Transparency at the data collection and processing stages is essential to build trust with users and ensure their continuous engagement.

Users' privacy should be protected, and any unauthorised access to sensitive data should be prevented by adopting strong data security measures. This involves ensuring that data is transmitted and maintained securely and that only authorised people can access it. In addition, there should be privacy policies that clearly explain how users can use and disclose their personal data. In addition to anonymisation, privacy-preserving machine learning should be applied to AIDT when developing ML methods such as predictive analytics based on user data. It ensures that the ML models developed are not hacked and thus violate the user's data [6].

AI/ML algorithms that analyse data and generate insights should be fair and unbiased, preventing algorithms from discriminating against certain groups of users [7]. Algorithms should be able to provide accurate and unbiased conclusions, so it is vital that they are routinely checked for bias.

Users should have control over their data to regulate how it is used. This covers giving users the ability to view, edit and delete their data and to decide who has access to-their data. Users should also have the option to opt out of having their data used for certain purposes, such as marketing or research purposes.

Last but not least, the AIDT also needs to be created such that everyone can utilise it, regardless of their situation or ability. Ensuring that the AIDT is simple to use, offers advantages to all users, and does not exclude particular user groups. The AIDT should be designed with the needs of users with disabilities in mind and should adhere to accessibility guidelines is key.

Conclusion

AIDT implementation for RWH can transform the remote work environment and also can offer insightful information to RWHs operators and urban planners, enhance resource management, and trigger the green transition. However, this technology must be used in a way that respects user rights and interests and adds to RWHs sustainability. To do this, it is necessary to address several crucial ethical issues. We should increase user trust, ensure ethical AIDT use, and support the ongoing development and optimisation of RWHs by addressing these ethical issues.

References

[1] Rainey, S. (2022). Datafied Brains and Digital Twins: Lessons From Industry, Caution For Psychiatry. Philosophy, Psychiatry, & Psychology 29(1), 29-42.
[2] Macnish, K. (2021). Operationalising Digital Twin Ethics. Sopra Steria.
[3] BIM+. (2021). Digital twin adoption: success lies in understanding the ethics.
[4] Huang, P. H., Lu, H. C., & Chen, C. S. (2022). Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study. Journal of Medical Internet Research, 24(1), e33081.
[5] Walters, A. (n.d.). Gemini Principles. Centre for Digital Built Britain Completed Its Five-year Mission and Closed Its Doors at the End of September 2022.
[6] Shu, X., Yao, D., & Bertino, E. (2015). Privacy-preserving detection of sensitive data exposure. IEEE transactions on information forensics and security, 10(5), 1092-1103.
[7] Wachter, S., Mittelstadt, B., & Russell, C. (2020). Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. W. Va. L. Rev., 123, 735.

A Deep Learning approach for identifying sexual harassment

Beatrix Tugyi and Drs. Tarry Singh, Dr. Tannistha Maiti

Abstract, *Sexual Harassment* is an unwelcoming behaviour of sexual nature, that can happen *anywhere*. It is a widespread problem all around the world, even in the workplaces. However, it's difficult to identify, because it's a multimodal process that can include many factors.

Armed with **Artificial Intelligence** technology, organizations can protect the employees, the company, and the culture from malicious employees, who would otherwise be toxic to the whole environment.

Introduction

What?

- *Unwanted touching*
- *Action movements*
- *Aggressive talks*
- *Facial emotions like anger, fear, obsession*

Why?

- **52%** *of women* and **10%** *of men* have experienced unwanted sexual behavior at work.[1]
- **75%** of sexual harassment cases typically *go unreported.*[1]

How?

- These problems can be solved with a combination of various **Deep learning** models. *RealAI* already has a working architecture, but it's accuracy needs to be improved.
- They have a big dataset, mostly based on films. This dataset can be improved with annotations and more videos.
- It's important to make sure nor the model nor the data have any biases.

Literature Review

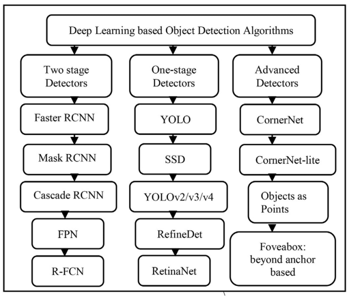
Sexual harrasment ***not only harm*** the victims but also has toxical effects on the culture, company and environment.

AI may be the HR solution to help identify instances *before* the situation got any worse. There were several previous attempts to solve this *complex problem*, but these solutions are not comprehensive and accurate enough for video detection.

Models that detect warning signs **in text or emails**, on the other hand, work quite well. (e. g. Aware app[2])

Research Methodology

For video detection I want to try out a combination of different Deep Learning methots to detect all the factors.



Deep Learning based Object Detection Algorithms

Two stage Detectors → Faster RCNN → Mask RCNN → Cascade RCNN → FPN → R-FCN

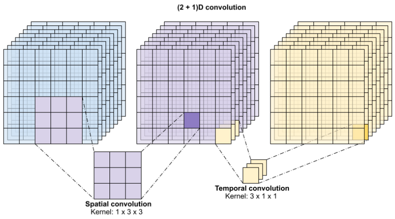
One-stage Detectors → YOLO → SSD → YOLOv2/v3/v4 → RefineDet → RetinaNet

Advanced Detectors → CornerNet → CornerNet-lite → Objects as Points → Foveabox: beyond anchor based

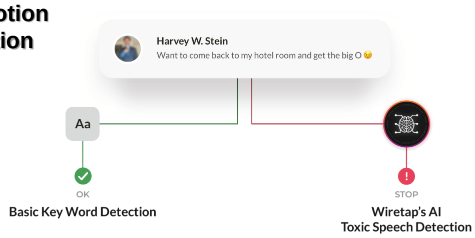
Object detection[5]

Facial Emotion Recognition

Movement detection[4]



3D CNN for image recognition[3]



Harvey W. Stein
Want to come back to my hotel room and get the big O 😉

Text and speech detection[2]

Preliminary Considerations

The definition of sexual harassment covers a big ground, it's not trivial, when the signs needs to be reported. The appropriate **boundaries must be established**.

Since most of the workspaces maintain *CCTV cameras*, it can be used to detect such occurrences, but the continuous monitoring of employees can lead to **ethical problems**.

Deep Learning models tends to be biased against a group of people, so it is crucial to find ways to reduce this effect.

Timeline

I will have **one year** for this project. This is my plan for a time schedule, each point should be maximum 2 months long.

- Data analysis and collection.
- Get to know the current architecture.
- Literature review of Deep learning techniques that can improve the result.
- Architecture upgrade.
- Consider the etichal implications and check the biases and intervene where necessary.
- Propose an etical way to use this with real life surveillance cameras.

Conclusion

It is extremely important to find a way of the early detection of sexual harassment in workplaces, because it can cause enormous damage, both emotionally and financially.

It's a challenging problem, but not impossibe to solve. There are already exitsting, good approches.

I am sure, that with the help of AI we can create a solution for this problem.

References

[1] According to an Everyday Sexism Project and the Trades Union Congress (TUC), https://www.globaltechoutlook.com/how-artificial-intelligence-helps-fight-sexual-harassment/

[2] Aware, Using AI to Identify and Reduce Sexual Harassment at Work, *https://www.awarehq.com/blog/identifying-and-reducing-workplace-sexual-harassment-with-ai*

[3] Tensorflow Video classification with a 3D cnn, *https://www.tensorflow.org/tutorials/video/video_classification*

[4] Kemtai, The Complete Guide to Human Pose Estimation, *https://kemtai.com/blog/the-complete-guide-to-human-pose-estimation/*

[5] Mittal, Payal & Singh, Raman & Sharma, Akashdeep. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. Image and Vision Computing. 104. 104046. 10.1016/j.imavis.2020.104046.

[6] Maiti, A., & Singh, U. N. (2021). The Impact of Brazenly Glorifying Sexual Abuse in Indian Film. Journal of Media Ethics, 1-3.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.

Revalidation and Explainability of PreSS

Cloë van Geest
Technological University Dublin
x00204826@myTUDublin.ie

Abstract

In 2006, a prediction model called Predict Student Success (PreSS) was developed to predict success early in introductory programming modules (CS1). So far, the model's accuracy has been revalidated twice. The proposed research will revalidate the accuracy of this AI model on a larger, international dataset from institutions from six continents. Performance measures for different target groups will be investigated to ensure transparency and sensitivity of the model. Additionally, the research will look into Explainable AI (XAI) approaches in order to develop the PreSS model further. The study aims to determine the level of explainability of the model and the trade-off with performance will be discussed. This way, the research contributes to building trustworthiness of PreSS for educational institutions to consider adoption of the model and guide appropriate interventions.

Background and Aim

Computer Science Education research is a relatively young field of study, with many gaps to fill. Identifying students at risk of dropping out or failing has been investigated, yet studies are rarely revalidated or tested for generalizability [1]. The computational prediction model called **Predict Student Success (PreSS)**[2], has shown to achieve high accuracy for predicting success early in introductory programming modules (CS1). The model uses a Naive Bayes algorithm, and takes in different predicting factors such as age, high school mathematics grades, and programming self efficacy [3]. So far, the model has been revalidated by two studies; one with institutions in Ireland and Denmark [4], and one study with institutions in Ireland and the US [3]. These findings laid the foundation for a larger scale international study. Hence, the first research question is: ***Is it possible to revalidate the previous studies done on PreSS, using data from multiple international contexts and generalize the results?***

The existing PreSS model is black box model. The outcome of the model is not easily explained or understood by the user. To offer correct support to the student, the explainability of the AI plays a role. Additionally, article 6 of the current proposal for the AI act, published in April of 2021, already states that **AI in education should be considered as high-risk AI** [5]. This is because the AI may determine one's educational course and ability to secure livelihood.

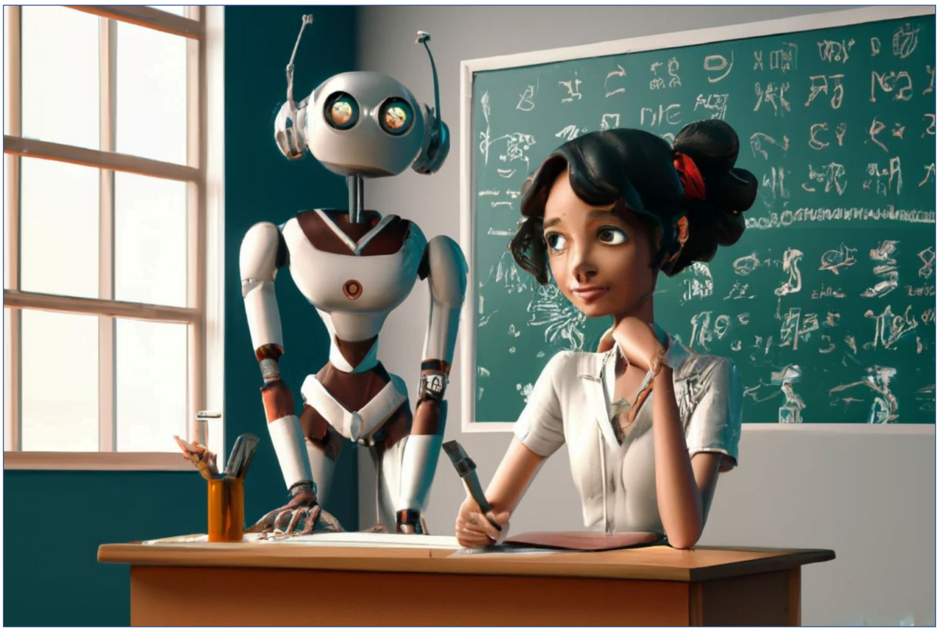


Figure 1. Teacher in classroom getting help from AI. Picture produced by Dall-E 2.

Even more, the Commission Expert Group created a specific use-case with ethical considerations for the use of AI and data to predict student progress and dropout[6]. Following this, the second research question of this thesis is: ***To what extent is PreSS explainable to build trustworthiness for adoption and what is the trade-off with performance?***

Methodology

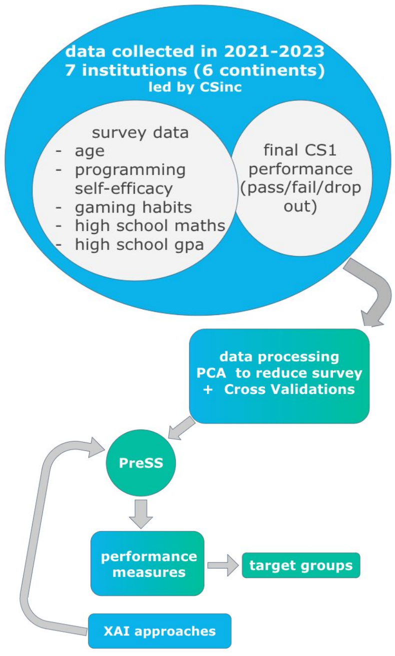


Figure 2. Representation of method steps

Ethical considerations

Data will be anonymized and is converted to non-personal data. Different institutions have different levels of education (e.g., QS world ranking), which may influence the data by creating bias towards higher achieving students and grades. This may influence the generalizability of results. Participating institutions raised concern about the study results being used as comparison for the quality of teaching between the institutions. It should be clear that the current research is no reflection of the teaching quality, as this should be determined by many factors[7]. Last, it can prove to be difficult to measure the explainability and transparency of a model. The explainability of this AI is particularly important as it was designed for education, which is considered a high-risk environment[5].

References. [1] Quille, K., & Bergin, S. (2019). CS1: how will they do? How can we help? A decade of research and practice. *Computer Science Education*, *29*(2-3), 254-282.
[2] Bergin, S. (2006). A computational model to predict programming performance (Unpublished doctoral dissertation). Department of Computer Science, Maynooth University.
[3] Quille, K., Nam Liao, S., Costelloe, E., Nolan, K., Mooney, A., & Shah, K. (2022, July). Press: Predicting student success early in cs1. a pilot international replication and generalization study. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1* (pp. 54-60).
[4] Quille, K., & Bergin, S. (2018, July). Programming: predicting student success early in CS1. a re-validation and replication study. In *Proceedings of the 23rd annual ACM conference on innovation and technology in computer science education* (pp. 15-20).
[5] Madiega, T. A. (2021). Artificial intelligence act. *European Parliament: European Parliamentary Research Service*.
[6] Fiok, K., Farahani, F. V., Karwowski, W., & Ahram, T. (2022). Explainable artificial intelligence for education and training. *The Journal of Defense Modeling and Simulation, 19*(2), 133-144.
[7] Little, O., Goe, L., & Bell, C. (2009). A Practical Guide to Evaluating Teacher Effectiveness. *National Comprehensive Center for Teacher Quality*.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Explainable AI to understand machine learning models

Szabolcs K. Weyde

Budapest University of Technology and Economics

Abstract

This thesis will contain an overview of the relevant literature on Explainable Artificial Intelligence. Furthermore, it will present a set of tools designed to help understand the mechanics of machine learning models. The main goal of the thesis is to provide a guideline to help people choose the right tools according to their model and the data they are willing to use.

Introduction

Nowadays, as AI is applied to more and more areas, the demand for XAI (Explainable AI) is rising. Authorities have already started working on acts and regulations, which further increases the need for tools and processes that are able to verify the trustworthiness of machine learning models. On the other hand, it is also important to make the decisions made by AI understandable and explainable to people, so the negative narratives surrounding it can be ended. The purpose of this thesis is to review the literature on XAI and to present some tools belonging to given topics.

Literature review

Besides the experiments I am willing to make, the thesis is using three works on the domain. The first one, "Explainable AI Methods - A Brief Overview" [1] contains a set of tools and methods designed to explain AI. It also contains information on these tools and methods, such as the idea behind them, and a brief review of them. The set contains elements such as LRP, Lime, Shap, Integrated Gradient, and much more. The second one, "Explainable AI Methods - A Brief Overview" [2], is about the history of Explainable AI. It tells the reader how the need for literature on the topic has arisen.

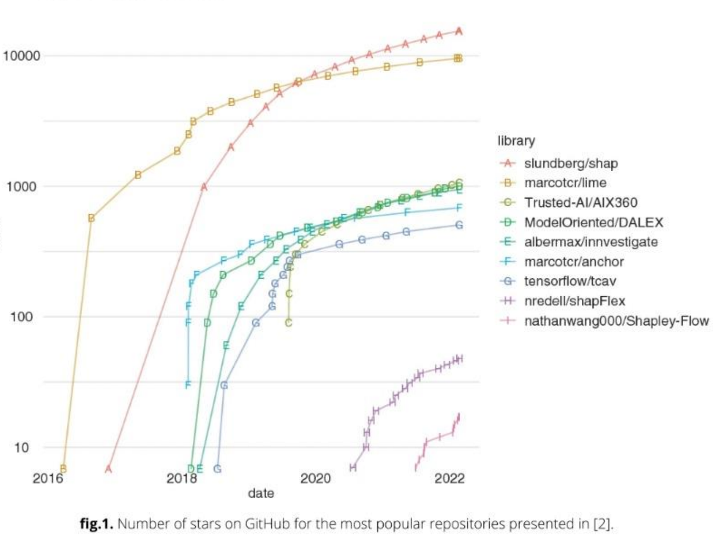


fig.1. Number of stars on GitHub for the most popular repositories presented in [2].

Research methods

During a thesis work, it is important to review literature in the given domain. Since XAI is currently a hot topic, there is a lot of literature in regard to it. This makes it necessary to set a set of goals, which can help you decide which are the right literary works for your research to progress. After that, the next step is to find these works. Once enough material on the topic has been processed and the tools, if there are any, have been gathered, it is time to test the tools. It is important to note that tests must be documented precisely enough so that they are repeatable. The more types of data and data set domains used during this phase, the more valuable the research can become.

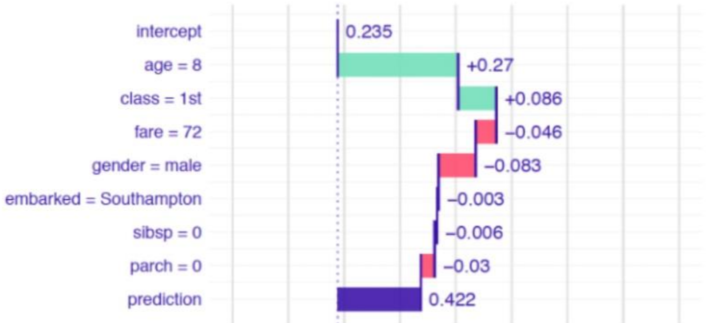


fig. 2. Dalex breakdown plot explaining the contribution of features to the result of the probability of surviving on the titanic dataset

Considerations

Explainable Artificial Intelligence can help people not only to understand, but also to accept AI. In the field of AI, it is an emerging problem that people are afraid of it. If the research is not well thought out enough, or the results are presented in a way that can be misleading for the wider audience, it can widen the gap between people and AI. It is crucial to make AI explainable for the science to progress, not only because it can allow it to be integrated into systems that need to be proven safe, but also because it needs to be accepted by people.

References

[1]: Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). **Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges.** In J. Tang, M.-Y. Kan, D. Zhao, S. Li, & H. Zan (Eds.), Natural Language Processing and Chinese Computing (pp. 563–574). Cham: Springer International Publishing.

[2]: Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). **Explainable AI Methods - A Brief Overview**. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K.-R. Müller, & W. Samek (Eds.), xxAI - Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers (pp. 13–38). doi:10.1007/978-3-031-04083-2_2

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant NoCEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.

Drone and Digital Twins for Crowded event management – A FIWARE-based approach

Dario De Dominicis
University of Naples «Federico II» and Meditech

Abstract. Managing crowded events (concerts, protests, marathons, etc.) is a challenge for any smart city. For officials in charge of the administration, forecasting attendance in advance, real-time situational awareness and predicting its evolution are critical to resource allocation. These large-scale events put pressure on public resources, organization, and security in smart cities. In this poster, I describe a FIWARE-based approach to model city layouts, deployed sensors, and citizen information (from social networks and smartphone apps) to handle these situations. We apply the concept of digital twins to cities by modeling different information flows integrated into a virtual 3D representation with predictive capabilities. The main purpose of my proposal is to facilitate the understanding and management of such incidents.

INTRODUCTION

Traditionally, large-scale event management has primarily focused on responding to disasters, and typically includes a static set of rules that dictate how to act when anomalous events occur [1]. The primary goal of any emergency plan is to prevent loss of life. Intelligent emergency response uses ICT technology, UAV and the Internet of Things extensively to evacuate during natural disasters (such as earthquakes, tsunamis, climate change impacts), crowd events (marathons, concerts), and fires, explosions, or floods. My initial hypothesis is mainly for crowded events. Furthermore, simulations can help predict the development of events and also help in solving organizational problems. 3D display and simulation functions are the basis for the digital twin concept. A digital twin [2] is a digital asset that looks and behaves like its connected physical counterpart (Figure 1). Every state change of a physical asset is reflected in its virtual counterpart, and every action in the virtual asset is propagated to the real asset. Because virtual assets behave similarly to physical assets, they can be used to simulate behavior and consequences, and predict the state evolution of their physical counterparts. One of the main challenges for the application of digital twins in urban spaces is information representation and situational awareness. The way we implement our digital twin approach is through augmented virtuality [3] and NGSI-LD API that is the current version of the core API of the FIWARE [4] open-source ecosystem.

LITERATURE REVIEW

The platforms that support the smart city concept originally came from IoT platforms, because smart cities are primarily applications of the Internet of Things. Snap4City is an example of platform originally developed for IoT and directly applied in the smart city domain. As a further development of such platforms, researchers and companies add services for citizens in specific subsectors of smart cities (traffic management, parking, security, pollution, waste management, etc.). The multitude of platforms proposed for smart cities has become a necessity to develop standards to unify protocols, layers and interfaces. Standards such as ISO 37100 defining a common vocabulary, ISO 37120 for key performance indicators, ISO/IEC 30182 for data interoperability, and IEC 63205 ED1 defining a smart city reference architecture are examples of standards developed to improve interoperability in smart city solutions. Smart city platforms face several challenges, including the difficulty of easily visualizing and understanding real-time information generated by the massive amount of data, and predicting the evolution of complex systems. The digital twin concept from Industry 4.0 offers a solution to these challenges by providing information visualization and situational forecasting. The concept is starting to be implemented in smart city platforms, such as the ongoing H2020 research project DUET (Digital Urban European Twins for smarter decision making), which aims to create virtual city replicas for smarter decision-making. The city of Shanghai has also created a virtual clone using Unreal Engine to support traffic flow monitoring. However, no additional information is provided about this platform.

Figure 1. Digital Twins and Urban Area – Abstract representation © *shutterstock/Jackie Niam*

RESEARCH METHODOLOGY

1. **Data Collection**: To test the efficacy of AI-based digital twins in crowd management, data will be collected from multiple sources, including sensors and cameras, and simulations run using digital twin technology.
2. **Data Analysis:** The collected data will be analyzed using various AI algorithms and machine learning techniques, such deep as learning, to evaluate the performance of the digital twins in predicting and managing crowd behavior.
3. **Model Development:** Based on the findings from the data analysis, a model of an AI-based digital twin for crowd management will be developed. This model will incorporate the best practices and techniques identified from the literature review and data analysis.
4. **Evaluation**: The developed model will be evaluated using real-world data and simulations to assess its performance in predicting and managing crowd behavior.
5. **Ethical Aspects**: The research will adhere to ethical standards and ensure that data privacy is protected and any personal information is kept confidential.

PRELIMINARY CONSIDERATIONS

1. **Data Availability:** To create an accurate Digital Twin simulation, it is crucial to have access to sufficient data. This can include data from sensors (IoT/UAV), crowd movement patterns, and other relevant information.
2. **Artificial Intelligence Techniques**: It is important to select the appropriate AI techniques to be used in the Digital Twin simulation.
3. **Ethical considerations :** The use of Digital Twins and AI in crowded environments raises several ethical concerns, including data privacy, the use of biometric data, and the potential for AI bias. It is important to consider these issues and implement appropriate measures to ensure ethical use of the technology.

CONCLUSIONS

The main contributions of this proposal, with regard to large-scale incident management, is a new perspective and an AI-based tool that can improve resource allocation, minimize response time in exceptional cases and improve user experience.

References.
1 W. O. et al., Crowd Management: Risk, Security and Health, G. Publisher, Ed. Goodfellow Publisher, 2019.
2 D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the digital twin: A systematic literature review," CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36–52, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1755581720300110.
3 J. Balaguer and E. Gobebetti, "Virtuality builder II: On the topic of 3D interaction"
4 J. Conde, A. Munoz-Arcentales, A. Alonso, S. Lopez-Pernas and J. Salvachua, "Modeling digital twin data and architecture: A building guide with fiware as enabling technology," *IEEE Internet Computing*, 2021.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Reinforcement learning in mechatronic systems

Pim Jansen - 30

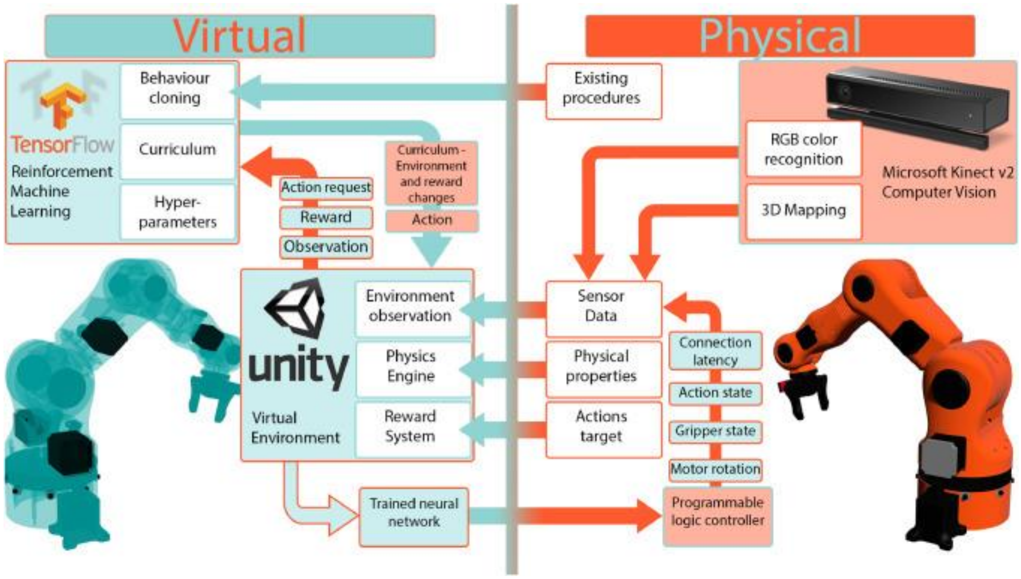
Abstract

Alten has proposed a project in which a robot learns to play four in a row. To give the project a greater social impact and generalize the problem, it has been chosen to conduct research to use reinforcement learning in combination with digital twinning to train robots digitally. The research will mostly consist of a literature study. Alongside this, a prototype will be made with the four in a row game.

Introduction

The project comes from Alten. Alten is a consultancy for large companies in the tech and IT sectors. The Eindhoven branch focuses on the Tech sector. The assignment comes from the mechatronics department of this branch. Unfortunately, they are not allowed to share the cool projects that Alten carries out with the world. This is why Alten works internally on projects with innovative techniques to share their expertise. One of these projects (and the one in question) is: making a robot play four in a row in a physical setup. This allows a person to play the game four in a row against a robot. The game and the mechanism that can place the pieces have already been realized, but an AI that can play the game has yet to be written.

Despite the project itself not being that complicated, I came up with a way to use the project itself to solve a larger social problem. A large part of mechatronic systems consists of robotics and industrial automation. These are often large machines that need to perform various tasks. Reinforcement learning can play a big role in learning these tasks, instead of the rule-based approach often used. The problem with reinforcement learning, and the main reason why it is not yet used, is the random actions taken by the agent. These random actions could cause the machine to damage itself or other machines or people. There is now a new technique to create a 1-to-1 digital copy of a machine, called digital twinning. This new technique makes it possible to train machines using reinforcement learning in a digital environment. This is what the research question is about: **Can combining digital twinning and reinforcement learning be an effective way to train robots?**



Ethical question

The concept that will be developed in this project will eventually have to be used by engineers. Besides whether the solution will work technically, the user, in this case the engineer, must also be considered. This could be done by making the model explainable by showing the training steps. This can be combined with the different levels created in the prototype. In this way, the solution is useful not only for Alten but for the entire industry. More opportunities to include engineers in the development will be explored during the project.

Prototype

To test the theory in a practical situation, a prototype will be made. In this prototype, a simple problem will be solved using a digital environment. For the environment, the game of connect four is used. To make the model more explainable along the training process, multiple models will be saved to understand the training process. These models will be connected to different age categories, so children can play against a computer that is challenging but not unbeatable. This prototype will mainly be used to look at ethical dilemmas by solving them in a simple situation. The technical requirements on the prototype have been kept simple on purpose to focus more on the issues around it. In an initial research phase, I came across several studies using the same techniques. [1] However, these studies focus only on the technical aspects of the problem.

[1] https://hu.on.worldcat.org/search/detail/8981291479?queryString=A%20robot%20arm%20digital%20twin%20utilising%20reinforcement%20learning

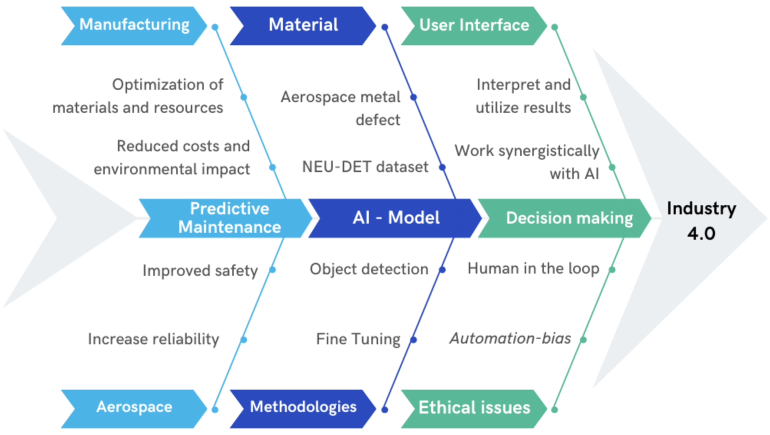
Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This Poster Was Created As Part of the Blended IP Event Organized Under Erasmus + Programme of the European Union.

Object Detection Model for Defect Inspection in Aerospace: Predictive Maintenance and Ethical Consideration in Industry 4.0

Flavia **Napoletano**
University of Napoli ‘’Federico II’’

ABSTRACT:

The research is focused on developing an AI object detection model for defect inspection in the aerospace industry. The model aims to detect six different types of defects on metallic surfaces, contributing to **predictive maintenance** and the principles of Industry 4.0. Ethical considerations will be addressed to ensure responsible implementation. The study highlights the potential of AI in optimizing defect inspection practices and improving operational effectiveness in aerospace.



INTRODUCTION:

Predictive maintenance is of paramount importance in the aerospace context and *Industry 4.0*, offering numerous benefits such as enhanced operational efficiency and cost reduction. It's alignment with green initiatives, promoting resource optimization and minimizing environmental impact.

In this study, we focus on implementing an **AI object detection model** for defect detection on metallic surfaces in the aerospace manufacturing sector. However, the deployment of AI models in the aerospace industry necessitates careful analysis and consideration of ethical implications. Balancing the advantages of AI-driven predictive maintenance with ethical considerations is crucial to ensure reliable and responsible implementation. This study aims to showcase the significance of predictive maintenance in aerospace, highlight the potential of AI object detection, and address the complex analysis and ethical risks associated with its use in the aerospace industry.

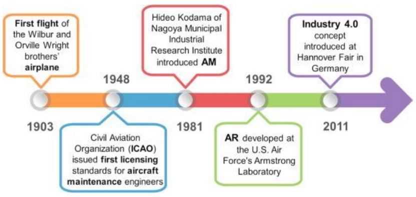

First flight of the Wilbur and Orville Wright brothers' airplane
1903
Civil Aviation Organization (ICAO) issued first licensing standards for aircraft maintenance engineers
1948
Hideo Kodama of Nagoya Municipal Industrial Research Institute introduced AM
1981
AR developed at the U.S. Air Force's Armstrong Laboratory
1992
Industry 4.0 concept introduced at Hannover Fair in Germany
2011

Figure 1: Evolution of aerospace maintenance to Industry 4.0 [1]

RESEARCH METHODOLOGIES:

For this thesis, the research methodology involves utilizing the **fine-tuning** technique by leveraging well-known pre-trained models (for example from the TensorFlow model zoo or the latest versions of YOLO). The objective is to perform object detection on the **NEU DET dataset,** [2] provided by the Northeastern University (NEU). This dataset consists of 1800 grayscale images, each measuring 200x200 pixels, and classified into six different **metal defect** categories: rolled-in scale, patches, crazing, pitted surface, inclusion, and scratches.

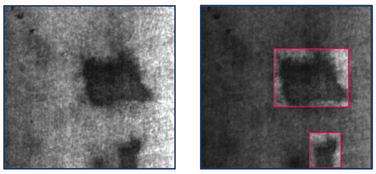


Figure 2: Comparison of a patches image and its corresponding defect detection version using object detection techniques.

ETHICAL CONSIDERATION:

A model for defect inspection in the aerospace context carries risks that should not be underestimated. Firstly, the model must ensure **resiliency**, meaning it should be *robust* and *reliable*. **Responsibility** and **accountability** come into play when using AI in predictive maintenance. The question arises: who is responsible if an AI algorithm fails to predict a critical failure? Another aspect not to be overlooked is the phenomenon of ***automation-bias***, [3] which refers to the tendency of humans to unquestioningly rely on automated systems or algorithms without critically evaluating their outputs or decisions.

CONCLUSION:

In conclusion, this study highlights the power of AI in predictive maintenance, particularly in object detection techniques within the aerospace industry. It emphasizes addressing ethical concerns and emphasizes responsible implementation of AI-driven solutions to maximize benefits while mitigating risks and upholding ethics.

References:

[1] - Alessandro Ceruti, Pier Marzocca, Alfredo Liverani, Cees Bil, Maintenance in aeronautics in an Industry 4.0 context: The role of Augmented Reality and Additive Manufacturing, Journal of Computational Design and Engineering, Volume 6, Issue 4, 2019, Pages 516-526, ISSN 2288-4300, https://doi.org/10.1016/j.jcde.2019.02.001.

[2] - Song, K.; Yan, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Appl. Surf. Sci. 2013,285, 858–864

[3] - Cummings, M. (2004). "Automation Bias in Intelligent Time Critical Decision Support Systems." AIAA 2004-6313. In AIAA 1st Intelligent Systems Technical Conference, september 2004. https://doi.org/10.2514/6.2004-6313

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Università degli Studi di Napoli Federico II
MEDITECH COMPETENCE CENTER
hcaim

hcaim

human centred
artificial intelligence
masters Budapest

Human-Centred Artificial Intelligence Master's Supplementary Programme

Faculty of Electrical Engineering and Informatics (VIK),
Budapest University of Technology and Economics (BME)

https://hcaim.bme.hu/en/msc-bme/msc-bme-phase2/

How to ensure human-centric and ethical operation of AI, its transparency and respect for the fundamental human rights while preserving the benefits of AI?

HCAI topics

- **HCAIM**
  - ***Personal and institutional autonomy and freedom:*** data security and privacy
  - ***Human-centric data analysis by design:*** MLOps, prior knowledge, explanation, active learning, machine teaching in cooperative intelligence, automated data analytics
  - ***Human-centric knowledge engineering:*** coding systems, ontologies, linked open data, summary statistics, automation of science (life sciences)
- **Classic ("existential risk")**
  - ***Benevolent AI***: AI solutions for humans, societies, and mankind
  - ***Trustworthy AI:*** value compatibility, understandability
  - ***AI safety:*** formal methods in systems engineering
- **Human-computer interaction**
  - ***Improved cognitive enhancers:*** personal assistants in education, intelligent citizen and costumer services, and decision support tools in personalized medicine
  - ***Improved sensorial man-machine interfaces:*** improved communication (speech, augmented reality) and human-computer interaction
  - ***Improved sensorimotoric man-machine cooperation:*** robotics, health-care assistance using wearable electronic devices, and automated driver assistance systems / autonomous vehicles
  - ***Smart devices, smart cities:*** IoT, sensor fusion for predictive maintenance

Didactic challenges

- How to mix engineering and human-centric (i.e. ethical, legal, social) knowledge in a single MSc course?
- How to demonstrate the human-centric approach in the whole design and development process of AI systems?
- How to apply the human-centric approach in practical exercises?
- How to prepare the students to comply with and monitor the legal regulations?

Goal: Deliver a 60 ECTS-credit Master's degree programme in AI with a human-centred focus and strong legal, ethical, sociological and other social science perspectives.
Realisation: As a supplementary programme *embedded* into our current 120 ECTS-credit Computer Science Master’s programme. The majority of courses are held in Hungarian.

Course mapping and credit requirements at BME VIK

Course type			Course group	Course title	Neptun code	ECTS	Specialization				C	Elective	ECTS req.	
HCAIM	BME	BME					S1	S2	S3	S4			min.	max.
I.														
Basic	Common	From each group, “ECTS req. min.” must be completed. From each group, at most “ECTS req.. max.” may be recognized within the 60 HCAIM credits.	1	Applied algebra and mathematical logic (autumn)	TE90MX75	5	5	5	5	5			5	5
				Mathematical statistics (autumn)	VISZMA11	5								
				Stochastics (autumn)	TE90MX77	5								
	Specialization-dependent		2	Machine learning (autumn)	VIMIMA27	5	5						5	5
				Machine learning (spring)	*VIMIMA27*	5						4		
			3	Deep learning (autumn)	VITMMA19	5	5						4	14
				Deep learning in practice with Python and LUA (autumn)		4						4		
				Deep learning in visual computing (spring)	VIIIMB10	5			5					
				Neural networks (spring)		4						4		
			4	The security of machine learning (spring)	VIHIMB09	5					5		5	10
				Trusted artificial intelligence and data analytics (spring)	VIMIMB10	5					5			
	Elective		5	Artificial intelligence and ethics (spring; also autumn in 2023)	GT41V105	2						**2**	**2**	**2**
			6	Artificial intelligence and law (spring and autumn)	GT55V106	2						**2**	**2**	**2**
			7	Artificial general intelligence (autumn)	VIMIAV22	2						**2**	**2**	**2**
	Common		8	Project lab 2 (with AI content)		5	5	5	5	5			**5**	**5**
				Thesis work 1 (with AI content)		10	10	10	10	10				
			9	Thesis work 2 (with HCAI content)		20	20	20	20	20			**15**	**15**
				A. HCAIM basic, total		94	50	40	45	40	10	18	**45**	**60**
II.														
Optional	Common	Mandatory completion depending on specialization	10	Project lab 1 (with AI content)		5	5	5	5	5			0	5
	Specialization-dependent		11	Intelligent data analysis and decision support (spring)	VIMIMB09	5	5						0	5
				Business intelligence (autumn)	VIAUMA24			5						
				AI-based human-machine interaction (autumn)	VITMMA23					5				
			12	Machine learning case studies (autumn)	VITMMA18	5	5						0	5
				Business intelligence lab (spring)	VIAUMB09			5						
				UX laboratory (spring)	VITMMB14					5				
			13	Advanced data analysis methods lab (spring)	VITMMB10	5	5						0	5
				B. HCAIM optional, total		20	20	15	5	15	0	0	0	20
		HCAIM basic + specialization optional, total											**45**	**80**
III.														
Optional	Elective	Recognizable	14	Any HCAI related course (after prior arrangement)										
		HCAIM optional, elective courses to the minimum 60 ECTS											**15**	**0**

Specializations and responsible departments			
S1	Major	Data science and artificial intelligence	Dept. of Measurement and Information Systems (MIT) - Dept. of Telecommunications and Media Informatics (TMIT)
S2		Software development	Dept. of Automation and Applied Informatics (AAIT)
S3		Visual informatics	Dept. of Control Engineering and Information Technology (IIT)
S4	Minor	User experience – UX and interaction	Dept. of Telecommunications and Media Informatics (TMIT)

How to obtain the 60 ECTS-credits* at BME VIK?

I. HCAIM basic courses: min. 45, max. 60 credits by completing a number of common, specialization-dependent or elective Computer Science courses, at least one course from each course group.

II. HCAIM optional courses: max. 20 credits by completing a number of common or specialization-dependent Computer Science courses, at most one course from each course group.

III. HCAIM optional courses: the missing credits (max. 15) may be obtained by completing a few HCAI-related elective courses (prior arrangement with the Dean’s Office is required).

* ECTS: European Credit Transfer and Accumulation System

What you get (in addition to the knowledge gained)? A ***HCAIM certification*** in your Diploma, Section 6.1.1 with the following text:

The student completed the requisite learning outcomes of the Human-Centred Artificial Intelligence Master's (HCAIM) programme, defined by the EU project INEA/CEF/ICT/A2020/2267304.

Common HCAIM-approach

4 modules, composed of technical, practical and human-centric components, defined by

- Bodies of Knowledge and Skills
- Lesson Plans,
- Learning Events, and
- Learning Outcomes.

Are you interested, would you like to join the HCAI Master’s at BME VIK? Then register at **https://forms.office.com/e/H4CQFA5JWY** (with a BME Sharepoint / Directory account).

Completing the form does not imply any obligation and does not count as an application for the Master's programme.

Those who fill in the form will be added to an MS Teams group, and will be regularly informed about the latest developments in the HCAI Master's programme, and the project advancements that might concern them.

human centred
artificial intelligence
masters

Register here!

Acknowledgement. Human-Centred AI Master's (HCAIM) project, grant № CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068

Co-financed by the Connecting Europe Facility of the European Union