DIGITAL 2023

Advances on Societal Digital Transformation

ISBN: 978-1-68558-115-2

September 25 - 29, 2023

Porto, Portugal

DIGITAL 2023 Editors

Petre DINI, IARIA EU/USA

DIGITAL 2023

Forward

The Advances on Societal Digital Transformation 2023 (DIGITAL 2023), held between September 25th and September 29th, 2023, continued a series of international events covering a large spectrum of topics related to digital transformation of our society.

The society is continuously changing at a rapid pace under digital transformation. Taking advantage of a solid transformation of digital communication and infrastructures, and with great progress in AI (Artificial Intelligence), IoT (Internet of Thinks), ML (Machine Learning), Deep Learning, Big Data, Knowledge acquisition and Cognitive technologies, almost all societal areas are redefined.

Transportation, Buildings, Factories, and Agriculture are now a combination of traditional and advanced technological features. Digital citizen-centric services, including health, well-being, community participation, learning and culture are now well-established and set to advance further on.

As counter-effects of digital transformation, notably fake news, digital identity risks and digital divide are also progressing in a dangerous rhythm, there is a major need for digital education, fake news awareness, and legal aspects mitigating sensitive cases.

We take here the opportunity to warmly thank all the members of the DIGITAL 2023 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to DIGITAL 2023. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the DIGITAL 2023 organizing committee for their help in handling the logistics of this event.

We hope that DIGITAL 2023 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of societal digital transformation.

DIGITAL 2023 Chairs

DIGITAL 2023 Steering Committee

Adel Aneiba, Department of Networks and Cybersecurity (NCS), School of Computing and Digital Technology, Birmingham City University, UK
Fernando Joaquim Lopes Moreira, Universidade Portucalense, Portugal
Yunpeng (Jack) Zhang, University of Houston, USA
Wanwan Li, University of South Florida, USA

DIGITAL 2023 Publicity Chairs

Laura Garcia, Universitat Politecnica de Valencia, Spain
Lorena Parra Boronat, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

DIGITAL 2023
Committee

DIGITAL 2023 Steering Committee

Adel Aneiba, Department of Networks and Cybersecurity (NCS), School of Computing and Digital Technology, Birmingham City University, UK
Fernando Joaquim Lopes Moreira, Universidade Portucalense, Portugal
Yunpeng (Jack) Zhang, University of Houston, USA
Wanwan Li, University of South Florida, USA

DIGITAL 2023 Publicity Chairs

Laura Garcia, Universitat Politecnica de Valencia, Spain
Lorena Parra Boronat, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

DIGITAL 2023 Technical Program Committee

Qammer Abbasi, University of Glasgow, UK
Aaroud Abdessadek, Chouaib Doukkali University, Morocco
Kawiwat Amnatchotiphan, Thai-Nichi Institute of Technology, Thailand
Daniel Amo Filvà, La Salle, Ramon Llull University, Spain
Mariia Andriyiivna Nazarkevych, Lviv Polytechnic National University, Ukraine
Adel Aneiba, Birmingham City University, UK
Sakthi Balan Muthiah, The LNM Institute of Information Technology, Jaipur, India
Mohamed Basel Almourad, Zayed University, United Arab Emirates
Louiza Bouallouche-Medjkoune, University of Bejaia, Algeria
An Braeken, Vrije Universiteit Brussel, Belgium
Roberta Calegari, University of Bologna, Italy
Jundong Chen, Dickinson State University, USA
Siming Chen, Fudan University, China
Marta Chinnici, ENEA, Rome, Italy
Javier Cifuentes-Faura, University of Murcia, Spain
María Elicia Cortés Cediel, Universidad Complutense de Madrid, Spain
José Paulo Cravino, Universidade de Trás-os-Montes e Alto Douro, Portugal
Mohamed Dahmane, CRIM – Computer Research Institute of Montreal, Canada
Babu R. Dawadi, Tribhuvan University, Nepal
Burcu Demirdöven, Pamukkale University, Turkey
Noella Edelmann, Danube University Krems, Austria
Zakaria Abou El Houda, University of Montreal, Canada
Nikos Fakotakis, University of Patras, Greece
Fadi Farha, University of Science and Technology Beijing, China / Aleppo University, Syria
Nasim Ferdosian, Cergy Paris University, France
Allan Fowler, University of Auckland, New Zealand
Ali Mohsen Frihida, University of Tunis El Manar, Tunisia
Andrea Gentili, Telematic University eCampus, Italy

Jing Gong, Uppsala University, Sweden
Teresa Guarda, CIST Research and Innovation Center | UPSE, Ecuador / ALGORITMI Research Centre of Minho University, Portugal
Stephan Haller, Bern University of Applied Sciences, Switzerland
Kuan He, Apple Inc., USA
Md Shafaeat Hossain, Southern Connecticut State University, USA
Chia-Yu Hsu, Arizona State University, USA
Hassan A. Karimi, University of Pittsburgh, USA
Michał Kawulok, Silesian University of Technology, Poland
Toshihiro Kuboi, Tastry, USA
Lahis Pasquali Kurtz, Institute for Research on Internet and Society (IRIS) | Federal University of Minas Gerais (UFMG), Brazil
Ahmed Lbath, Université Grenoble Alpes, France
Junghee Lee, Korea University, Korea
Azi Lev-On, Ariel University, Israel
Wenwen Li, Arizona State University, USA
Wanwan Li, George Mason University, USA
Adnan Mahmood, Macquarie University, Australia
Alberto Marchisio, Institute of Computer Engineering - TU Wien, Austria
Farhad Mehdipour, Otago Polytechnic - Auckland International Campus, New Zealand
Andrea Michienzi, University of Pisa, Italy
Gianluca Misuraca, Universidad Politécnica de Madrid, Spain
Fernando Moreira, Universidade Portucalense, Portugal
Mac Motsi-Omoijiade, RAND Europe, UK
Raghava Rao Mukkamala, Copenhagen Business School,Denmark /Kristiania University College, Norway
Diane R Murphy, Marymount University, USA
Mai Neo, Multimedia University, Cyberjaya, Malaysia
Mathias Nippraschk, Institute of Mineral and Waste Processing, Waste Disposal and Geomechanics - Clausthal University of Technology, Germany
Marcelo Iury S. Oliveira, Federal Rural University of Pernambuco, Brazil
Nuria Ortigosa, Universitat Politècnica de València, Spain
Hamza Ouarnoughi, INSA Hauts-de-France, Valenciennes, France
Pedro R. Palos-Sanchez, University of Sevilla, Spain
Giovanni Pau**,** Kore University of Enna, Italy
Sandra Milena Pérez Buitrago, Pontificia Universidad Católica del Perú, Lima, Perú
Paulo Pinto, Universidade Nova de Lisboa, Portugal
Filipe Portela, University of Minho, Portugal
Achim Rettberg, University of Applied Sciences Hamm-Lippstadt / Carl von Ossietzky University Oldenburg, Germany
Manuel Pedro Rodríguez Bolívar, University of Granada, Spain
Amirreza Rouhi, Drillmec SPA / Politecnico di Milano, Italy
Razak Seidu, Norwegian University of Science and Technology (NTNU), Norway
Atriya Sen, University of New Orleans, USA
Ecem Buse Sevinç Çubuk, Aydın Adnan Menderes University, Turkey
Pietro Siciliano, Institute for Microelectronics and Microsystems (IMM-CNR), Lecce, Italy
Anthony Simonofski, Namur Digital Institute | Université de Namur, Belgium
Rosario Soria, IAG Finance, New Zealand
Abel Suing, Universidad Técnica Particular de Loja, Ecuador

Do Duy Tan, Ho Chi Minh City University of Technology and Education (HCMUTE), Vietnam
Camel Tanougast, University of Lorraine, Metz, France
Najam ul Hasan, Dhofar University, Salalah, Sultanate of Oman
Giacomo Valente, University of L’Aquila, Italy
Washington Velasquez Vargas, Escuela Superior Politécnica del Litoral, Ecuador
Massimo Villari, University of Messina, Italy
Li Wang, University of North Carolina at Chapel Hill, USA
Olaf Witkowski, Cross Labs | University of Tokyo | Tokyo Institute of Technology, Japan
Marcin Wozniak, Silesian University of Technology, Poland
Seyed Yahya Nikouei, Kar Global, USA
Guillaume Zambrano, Nimes University, France
Tengchan Zeng, Virginia Tech, USA
Chi Zhang, The University of Glasgow, UK
Yunpeng (Jack) Zhang, University of Houston, USA
Jiayan Zhao, The Pennsylvania State University, USA
Zheng Zhao, Synopsys Inc., USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Could ChatGPT be Your New Project Management Assistant? 1
Claudia Hess and Sibylle Kunz

Verifiable Labels for Digital Services: A Practical Approach 8
Mael Gassmann and Annett Laube

Teachers, Reload Your Toolboxes: Using a Task Design Framework and ChatGPT to Generate Motivational Exercises for Female Learners in Computer Science 13
Sibylle Kunz and Claudia Hess

A Foveated Approach to Automated Billboard Detection 20
Sayali Avinash Chavan, Dermot Kerr, Sonya Coleman, and Hussein Khader

Deep Learning for Billboard Classification 29
Sayali Avinash Chavan, Dermot Kerr, Sonya Coleman, and Hussein Khader

Smart Contracts for Privacy-Preserving Identity Management: Ethics, Regulatory and Technical Challenges 36
Carmela Occhipinti, Tetiana Vasylieva, Luigi Briguglio, Alessio Bianchini, and Sayonara Crestani

Could ChatGPT be Your New Project Management Assistant?

Using AI Chatbots to Support Project Management Tasks

Claudia Hess and Sibylle Kunz
IT & Engineering
IU International University of Applied Sciences
Erfurt, Germany
e-mail: claudia.hess@iu.org; e-mail: sibylle.kunz@iu.org

***Abstract*—Over the past few years, people became accustomed to using a variety of digital tools in their work environment, e.g., for collaboration, knowledge management and task management. Recently, people have become aware of a new type of digital tool: Artificial Intelligence (AI) chatbots, such as ChatGPT, which are based on large-scale language models. These tools are expected to have a huge impact on the way we work. New application scenarios are constantly being explored and AI chatbots are being used in fields such as marketing, medicine, and education. This paper discusses another use case in the field of project management. It analyzes how AI chatbots could support project management tasks performed by project managers and project management assistants. It also discusses the new skills that these persons must acquire in order to interact with the tool in an efficient way and to use the results provided responsibly. A case study shows how selected project management tasks, i.e., stakeholder management, can benefit from the assistance of an AI chatbot. Special attention is given to the aspects of trust and control in such settings.**

Keywords-Large Language Model (LLM); project management; stakeholder management; prompt engineering.

I. INTRODUCTION

Project managers and project management assistants are familiar with using various digital tools in their daily work, ranging from project planning, and monitoring to stakeholder and risk management [1]. Some of these tools are highly specialized, such as project management software, while others are not limited to project management, such as video conferencing software, instant messaging applications, cloud-based document sharing, or digital whiteboards. These digital tools have fundamentally changed the way people work. Recently, a new type of tool has emerged that has the potential to have a profound impact on modern work culture, including project management: Artificial Intelligence (AI) chatbots [2]. They simulate human-like conversations and provide meaningful responses to user queries [3]. This is possible because they are based on Large Language Models (LLMs) that have been trained on vast amounts of text data from the Internet. Due to recent advances in AI, these AI chatbots are becoming increasingly powerful. Since their release to the public, people have explored their capabilities in a variety of application areas and shown how they can improve their productivity [3].

As the use of digital tools is already widespread in project management, it can be assumed that a new digital tool such as an AI chatbot will easily and quickly be adapted by many project managers. Sometimes, they are already considered as digital assistants [2]. AI chatbots can assist project managers in their daily tasks. Ideas can be found in blog posts, and first “cheat sheets” showcase examples of queries [4]. However, there are many differences in the way of interacting and interpreting the results obtained by an AI chatbot in comparison with classical tools in project management. The project manager and the project team should be aware of them and possess certain skills to use them in a productive way. Moreover, despite all the new and exciting possibilities that AI chatbots offer, there are risks involved that need to be mitigated, such as ethical challenges like discrimination, information hazards, misinformation harms and human-computer interaction harms [5][6].

To this end, the paper is structured as follows. Section II provides the required theoretical foundation on LLMs and the new skill “prompt engineering”. Section III discusses how persons responsible for project management tasks can benefit from AI chatbots in their daily work. Therefore, it structures AI-assisted project management tasks in six categories and analyzes the required skills. As a concrete case study, Section IV demonstrates how an AI chatbot can support the activities around stakeholder management. Useful prompts will be introduced. Based on the experiences with the case study, Section V discusses potential and risks of the use of AI chatbots in project management. Section VI concludes the paper.

II. LARGE LANGUAGE MODELS AND AI CHATBOTS

A. *AI Chatbots*

Since OpenAI released ChatGPT to the public in November 2022, AI chatbots gained much attention. Via their user-friendly interface users can benefit from the capabilities of LLMs. These systems use advanced Natural Language Processing (NLP) techniques to understand users’ questions, generate coherent answers and thereby simulate a conversation between human beings [3]. Their development has been made possible by the recent advances in AI, especially deep learning and transformer architectures, the rapid growth in computational power and the large amount of data available on the Internet that could be used as training data [7]. LLMs have learnt based on this extensive training set to generate text sequences that most likely continue a given text [8]. Nowadays, the quality of the text produced by these systems is so high that it is difficult to distinguish whether a text was written by an AI or by a human [9].

It is important to know the differences between LLMs and search engines. They are distinct technologies, as they are characterized by different capabilities. LLMs generate context-sensitive text based on a given input, formulated in natural language [3]. In contrast, search engines take a keyword-based query as input and retrieve information from indexed web pages or databases. They do not generate new content but provide access to existing information. This is not the case with LLMs which might "hallucinate" facts and references [3]. Moreover, current LLMs cannot provide information on recent events without being integrated into other tools that have access to up-to-date information (such as search engines) [3]. Users must be aware of these limitations.

B. Prompt Engineering

The instructions given to a LLM are called prompts. In the prompt, the user provides the LLM with the context of the conversation and defines how the intended output should look like in terms of content and format [10]. For instance, a user might specify the desired output as an essay with approximately 1500 words. The process of constructing these prompts is referred to as *prompt engineering*, i.e., "the art of fine-tuning the questions or commands provided to AI models in order to optimize their performance and guarantee that they produce the desired results" [11]. In consequence, to collaborate efficiently with LLMs, users need to have a good knowledge of prompt engineering [3].

III. AI CHATBOTS IN PROJECT MANAGEMENT

A. Patterns of AI-Assisted Project Management Tasks

AI chatbots can be applied to the discipline of project management in a variety of ways. In the following, AI-assisted project management tasks are organized into six patterns. Tasks as assisting in coding or creating images are not included because they are typically performed by team members, not the project manager. Given the rapid advances in LLMs, the following overview is not meant to be exhaustive. Rather, it aims to provide a structured framework for practical tasks.

1) Generate Ideas and Insights: An AI chatbot can support idea generation and brainstorming by suggesting topics or themes based on user input or some predefined categories. Combined with search engines, they can retrieve information from a variety of sources, such as articles, research papers, or databases. However, users need to check the output generated by the AI chatbot for their correctness [3], as a high likelihood of the generated content is no guarantee that it is correct [5]. A project manager could, for example, ask an AI chatbot to generate a list of risks or ideas for risk mitigation strategies.

2) "How to Do"- Assistance: Users can ask an AI chatbot for guidance on a certain task, for instance, how to write a project proposal or how to plan and estimate the project. The AI chatbot typically provides a list of steps for the user to follow, often like a recipe, e.g., certain aspects that need to be clarified for the project proposal. It should be noted that, if the prompt is too general, the result will also be unclear or inaccurate [12].

3) Text Summarization: AI chatbots are able to summarize or extract key aspects of large texts. This ability could save time for a project manager who needs to compile and summarize project data for a status update. Instead of reviewing numerous documents and reports, they could focus on interpreting the results. However, this task requires a significant level of trust in the AI chatbot.

4) Content Creation: AI chatbots can assist in generating content for various purposes, such as articles, blog posts, social media updates, or product descriptions [3]. Based on a given topic, they can, for example, offer relevant keywords, suggest headlines, or write complete articles. In project management, different types of content are relevant, such as official documentation like project proposals, one-pagers, reports, internal working documents, as well as any communication with stakeholders. In addition to content creation, AI chatbots can improve written text. On the one hand, a chatbot can correct a text in terms of spelling, punctuation, and grammar. This is especially helpful for non-native speakers. On the other hand, it can rewrite a text and change its writing style based on the desired tone and target audience. For instance, it could turn a formal and technical text into a conversational and engaging one. This could be useful when preparing two communications about a milestone achieved in the project, one for the project steering committee, and the other for future users.

5) Pretend to be Someone Else: An AI chatbot can help users explore different perspectives by simulating to be someone else. By pretending to be a particular persona ("role prompting") with certain opinions, experiences, or expertise, the chatbot can provide responses from that particular perspective. This allows users to gain insights that they might not have considered otherwise. In project management, this could be useful in stakeholder management. A variation would be a flipped interaction, where the AI chatbot asks questions and the user has to answer. This could be used to train difficult situations.

6) Sentiment Analysis: AI chatbots can identify sentiments or emotional tones expressed in a given text [13]. They can assist users in gaining insights into the overall sentiment or mood of the content. This capability could be used in stakeholder management, for instance, to analyze comments posted by stakeholders on social media or other online communications.

B. *Required Skills*

To interact successfully and effectively with AI chatbots, humans need certain skills. With the previously described tasks in mind, these skills are explored below.

1) *Ability to Write Concise Prompts*

When writing prompts, users should ask specific questions and be concise in their wording, as this helps chatbots to better understand user queries. In addition to the use of simple language, the questions, as well as the provided content should be clearly structured [11][12]. Various guidelines on prompt engineering can be found online. It should be noted, however, that these were compiled for a certain tool in a certain version, e.g., GPT-4. Because of rapid development, prompts that worked very well when these guidelines were written may not give good results today. Or prompts could be written in a much simpler way. This is also the case for all prompts discussed in the following. However, some prompting techniques that will be applied within the case study are presented.

First, users should think about whether they want to integrate several tasks into a single prompt or to split them into multiple prompts. For instance, an AI chatbot could be asked in a single prompt to generate four ideas about a certain topic and to discuss them in a short blog post. Alternatively, a first prompt could ask for four ideas, and a second one to write a blog post based on idea 1, idea 2, and so on. The advantage of using multiple prompts is that the output can be reviewed, and if necessary, adjusted or extended. A second important technique is to incorporate background knowledge into the prompt. These knowledge-augmented queries specify additional context that the LLM uses to generate more informed and accurate responses [12].

2) *Understand AI Chatbots' Capabilities and Limitations*

It is important that users are aware of the actual capabilities of AI chatbots, i.e., to know what tasks they can perform, what knowledge they use to answer queries, and how they differ from search engines. This will help users to frame their questions and expectations accordingly. Awareness of limitations and potential difficulties, such as bias in the underlying data, is also extremely important, as it is a prerequisite for users to write appropriate prompts and interpret the results provided by AI chatbots in a responsible manner. For instance, users should be careful to write prompts that are free of pre-existing biases or assumptions about gender, race, or other sensitive factors to avoid reinforcing bias [12]. Users should think critically, verify the results with their own expertise and conduct further research in trusted sources [12].

3) *Adopt an Iterative Approach*

Users should be prepared to take an iterative approach when interacting with an AI chatbot [3]. It may be necessary to phrase the prompt in a different style or add information. Being flexible and adapting to the AI chatbot's conversation flow will ensure a smoother conversation. Sometimes users need to be patient and try different approaches. While this can be considered a form of learning, users need to be aware of the chatbot's limitations and should consider whether not using the chatbot might be more efficient in specific situations.

4) *Awareness of Privacy and Security Issues*

When interacting with AI chatbots, users should critically evaluate what information they are willing to share. They need to be aware of whether the question or contextual information provided in a prompt reveals sensitive data. This applies to both personal and business-related data. For example, users should be careful about writing project-specific details, financial information, or proprietary data. They should always formulate such information in a generalized and anonymized manner, and consider the need to disclose sensitive information.

5) *Being Aware That You are Talking to an AI*

Users need to be aware that they are communicating with an AI, not a human. Because of the conversational interface, users tend to think of a chatbot as a human, or at least attribute some human-like characteristics to it. This can be observed in users' prompts where they use phrases like "please", "thank you", even though they know that they are talking to a bot. Anthropomorphizing the AI chatbot poses several dangers, such as shifting accountability to the chatbot or overestimating its capabilities [5]. Project managers using AI chatbots should be aware that AI chatbots may not be able to handle complex or nuanced situations that require human judgment and decision-making. They operate based on data and patterns, which may not always account for a specific situation or in unforeseen circumstances. Users should think critically and consider the limitations of chatbots when relying on their advice or recommendations. Project managers should also understand that not only can they be misled by an AI chatbot's response, but so can their team members.

IV. CASE STUDY ON INCORPORATING AI CHATBOTS IN PROJECT MANAGEMENT

The following case study demonstrates how AI chatbots can increase the efficiency and effectiveness of project management. It considers a digitalization project at a European manufacturing company which aims to introduce a new software for recording the working time. This system should simplify and automate current processes. The project encompasses the vendor selection, the customization of the selected solution, the process redesign, and the training of the future users. Note that scope and objectives of the project considered in the project management case study are already defined as it would be the case in any project in an enterprise context. A conversational AI could have been used in this preliminary step, too. For instance, it could support the project initiators to better formulate the objectives or assist other tasks in project definition.

Within the case study, concrete activities in project management should be performed with the help of an AI chatbot, namely one of the project performance domains described in the PMBOK Guide published by the Project Management Institute (PMI) [14]. A project performance domain is defined as "a group of related activities that are crucial for the effective delivery of project outcomes" [14]. The PMBOK Guide outlines eight performance domains that should be considered during the implementation of a project. These domains play a crucial role in driving the project towards its intended outcome. In the case study, the focus will be on the stakeholder performance domain.

This performance domain emphasizes that it is important to maintain alignment with the stakeholders of a project and to engage with them a positive relationship [14]. Stakeholders are persons or groups of people who affect the project and its outcomes, or who are affected by the project and its outcomes. They all bring their own ideas, values, qualifications, and prior experience [15]. The PMBOK Guide identifies several activities in the context of the stakeholder management as shown in Figure 1.

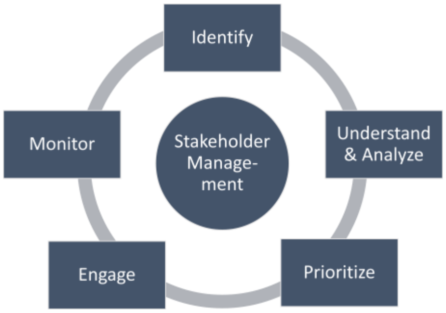


Figure 1. Activities to ensure stakeholder engagement based on [14].

A first try could be to write a single prompt that asks the AI chatbot to perform the stakeholder management, such as "Do the stakeholder management for a digitalization project in a German manufacturing company that aims to introduce a new digital system for recording the working time.". The result is, according to ChatGPT, a "comprehensive stakeholder management plan for your project". It contains eight steps that can be mapped to the activities named in the PMBOK Guide. For each step, a type of deliverable is presented, such as a list of stakeholders, or a communication plan. However, a closer look at the results reveals that they are rather general. For instance, one recommendation for managing stakeholders' expectations is to highlight the benefits of the new system, such as accurate time tracking, improved efficiency, and reduced paperwork. This is true, but too superficial to be useful in day-to-day work. This example highlights the need for more sophisticated prompts. For instance, each of the eight different steps could be refined with the help of individual prompts, each focusing on specific aspects of stakeholder management. Such prompts are shown below for each stakeholder management activity.

A. *Identify Stakeholders*

First, it is necessary to identify all potential internal and external stakeholders of a project. While certain stakeholders can readily be identified, there might also be stakeholders that are only indirectly affected by the project [14]. They are more difficult to identify. In practice, the activity of identifying stakeholder is often performed by the project manager and some team members in a brainstorming session. An AI chatbot can support this task. To enable the AI chatbot to generate a list of stakeholders that is specific to the project at hand, some context about the project must be provided. It is recommended to describe the project in a short but precise way. By doing this, it has to be taken care that no confidential information is passed to the AI-based system. Table I shows such a prompt.

TABLE I. EXAMPLE PROPMPT TO IDENTIFY STAKEHOLDERS

Intent	**Identify the stakeholders of a project**
Type of prompt	Brainstorming
Required input	Context information about the project
Prompt	Consider the following project and identify the different stakeholders involved in this project: The digitalization project in a German manufacturing company aims to introduce a new digital system for recording the working time. This new system should simplify and automate the process of recording working time. This project encompasses the vendor selection, the customization of the selected solution but also the process redesign and the training of the future users.
Result	List of stakeholders including a brief description of how they are connected with the project

Before further using the results within the project, it is necessary to carefully check the plausibility in the specific context of the project. As the AI chatbot generates the answer based on the information seen during the training of the LLM, relationships could be created that are not relevant in the current project. However, it may also be the case that the AI chatbot identifies stakeholders that one has not thought about oneself. So, one should carefully think about them before excluding them from the list of stakeholders. In addition, the list of stakeholders might be incomplete and might need to be extended "manually".

B. *Understand and Analyze Stakeholders*

To be able to effectively engage and communicate with stakeholders, it is important to understand their "feelings, emotions, beliefs and values" [14]. An AI chatbot can help to analyze the stakeholders' perspectives. In a first step, it could be asked in a rather general way to gain additional information about the stakeholders, their connection to the new system and to the project. Such a prompt could be "Provide more information about the stakeholder 'work council'.". Note that AI chatbots remember the information given within one conversation. So, the prompts can build upon each other.

Including good practices from project management in the prompt can improve the output. The PMBOK Guide, for instance, recommends considering different aspects in stakeholder analysis, especially their power, impact, attitude, beliefs, expectations, degree of influence, proximity to the project and their interest in the project [14]. These aspects can be added as desired output to the prompt as Table II shows. The project team should review the results as it might be necessary to refine them based on project specific information.

TABLE II. EXAMPLE PROPMPT TO ANALYZE STAKEHOLDERS

Intent	Understand the stakeholders of a project
Type of prompt	Generate insights
Required input	The stakeholder to be analyzed. If available, context information about the stakeholder or past experiences with this stakeholder should be added.
Prompt	Discuss for the stakeholder "work council" its power, impact, attitude, beliefs, expectations, degree of influence, proximity to the project and its interest in the project.
Result	Information about each of the key aspects required in the prompt. Note that without providing the key aspects in the prompt, the result would be more general.

To deepen the understanding of the stakeholders' perspectives, the AI chatbot can be asked to describe typical personas, i.e., fictional representations who have a name, a profession and who can describe the impacts that the new system would have on them. For instance, several personas could be created that represent the stakeholder “employees”. Then, the chatbot could be asked in a follow-up prompt what a certain persona would say about the project.

In our case study, it was necessary to ask explicitly for critical perspectives. The answers to our first prompts rather sounded coming from a marketing brochures of time tracking systems and every persona was said to be happy with the system. Table III shows a prompt that provides us with a more critical perspective.

TABLE III. EXAMPLE PROPMPT WITH PERSONAS

Intent	Explore different perspectives
Type of prompt	Generate insights
Required input	Additional information about the personas, such as their current way of working and potential concerns
Prompt	What would persona <A> say about the time tracking system when she was formerly used to start and stop working whenever she wished, only being judged by her results, not her time spent on them?
Result	Differentiated opinion from the persona's perspective

C. Prioritize Stakeholders

Due to limited resources, projects normally need to focus on the most important stakeholders [14]. Therefore, it is usual to evaluate all stakeholders with respect to their attitude to the project, i.e., whether it is positive or negative, and whether they have a high or low influence on the project's progression [14][15]. A so-called stakeholder matrix or stakeholder map can serve as a tool to provide a visual overview of all stakeholders with their interests and their influence [16]. It can be used as a basis when planning measures for engaging and communicating with stakeholders. Note that in the prompt showing in Table IV, the AI chatbot is already familiar with the concept of a stakeholder matrix.

TABLE IV. EXAMPLE PROPMPT TO PRIORITZE STAKEHOLDERS

Intent	Create a stakeholder matrix
Type of prompt	Create project documentation
Required input	List of stakeholders and their concerns
Prompt	Classify the stakeholders identified for this digitalization project in the different quadrants of a stakeholder matrix.
Result	List of stakeholders per quadrant (High Power, High Interest; High Power, Low Interest; Low Power, High Interest; Low Power, Low Interest)

The result generated by the AI chatbot is a list of stakeholders which could be used as the basis for creating a visual representation in the form of a stakeholder matrix. Current text-based AI chatbots are not yet able to generate such visualizations. They can only walk you through the process.

It is important to address any concerns that stakeholders with high power and high interest have. A prompt to generate ideas on how to address potential concerns might be “What incentives could be used to raise acceptance?”.

The list of stakeholders and the stakeholder matrix are important artefacts created within project management. They should be reviewed throughout the course of the project, and, if necessary, be updated because on the one hand, stakeholders might change their opinions and, on the other hand, influencing factors might change [14]. In addition, misevaluations can be corrected.

D. Engage Stakeholders

Successful stakeholder engagement requires different communication approaches [14]. Depending on the audience and the type of information, a verbal or written communication should be preferred. Moreover, the communication might be formal or informal. An example for a formal written communication is a progress report, while a social media post or an instant message is an example for a written informal communication. An AI chatbot can support project managers in preparing an appropriate communication artefact. This might include several steps in which the content of the communication is elaborated, the type of communication is determined, and the actual text is written. By doing this, users need to take care that the communication does not remain superficial and that it does not appear to be artificially generated. The time saved by utilizing an AI chatbot can be used to create customized communications for each stakeholder, which can also be supported by the AI chatbot. Table V shows such a series of prompts.

TABLE V. EXAMPLE PROPMPTS TO GENERATE CONTENT

Intent	**Engage stakeholders by a targeted communication**
Type of prompt	Step 1: Get advice Step 2: Create communication materials.
Required input	Additional information about the stakeholder and the topic to be communicated
Prompt	Step 1: How should the rationale behind the new system be communicated to stakeholder <A>? How should the communication be realized? Who should communicate in which way to these three personas? How often? Etc. Step 2: Write a formal e-mail with about 200 words about <topic>.
Result	Step 1: Step-by-step guidelines Step 2: Text

A further application scenario in stakeholder engagement is the training and simulation of difficult situations. For example, an AI chatbot can help the project manager to prepare a project presentation with stakeholders that express many concerns about the project. Table VI demonstrates such a prompt.

TABLE VI. EXAMPLE PROPMPT FOR TRAINING AND SIMULATION

Intent	**Train the conversation with a difficult stakeholder**
Type of prompt	Training and simulation
Required input	Information about the situation
Prompt	I'm the project lead of the following project: […] I have identified the work council as an important stakeholder in this project. Act as the stakeholder work council. I want you to do a dialogue with me. Start with asking a question about your first concern and wait for my answers.
Result	Flipped interaction with questions asked by the AI chatbot and answers given by the user

E. Monitor Stakeholders

During a project, stakeholders might change their opinions, new stakeholders might become relevant while the importance of others may decline. Therefore, it is important to monitor the stakeholders and their satisfaction with the project [14]. AI chatbots can support by analyzing the sentiments that stakeholders express in their communication. E-mails or any other communication in internal groups or communication platforms about the project could be analyzed by an AI chatbot with the help of a prompt such as "Given this text, what is the sentiment conveyed? Is it positive or negative?". However, it has to be noted that no confidential information is handed to a public AI chatbot.

V. POTENTIALS AND RISKS OF USING AI CHATBOTS IN PROJECT MANAGEMENT

The pattern of AI-assisted project management activities, as well as the examples within the case study have demonstrated that AI chatbots can indeed be a useful digital tool in project management. However, there are several issues that lead us to conclude that they are "just" a tool and not yet a digital assistant that is a full member of the project team. First, project managers need to trust that the chatbot has a solid knowledge of project management methods and good practices, for instance, when asking for guidance. To build up this trust, it would be helpful if the chatbot added a reference to established project management standards, such as published by PMI, ISO or PRINCE2 to the answer. As long as this is not the case, it is recommended to provide this guidance to the chatbot in the prompt, such as shown in the prompt in Table II, which explicitly mentions the different aspects to be considered in the stakeholder analysis.

Second, the project manager must be able to review any generated content. For instance, if the AI chatbot extracts a project status from various documents, traceability is required. In other words, it must be clear what information was used to set the status. To a certain extent, this can be achieved by explicitly asking the AI chatbot to provide its reasoning in addition to the answer. This can be achieved, for instance, by adding a specific response format that guides the AI chatbot such as shown in Table VII.

TABLE VII. EXAMPLE PROPMPT FOR PROVIDING REASONING

Intent	**Obtain the chatbot's reasoning in addition to the answer**
Type of prompt	Transparency
Required input	Information about the situation
Prompt	[Prompt] Response format: # Reasoning < provide a detailed reasoning here > # Response < Provide your response here >
Result	Answer splitted into two parts, the reasoning behind the response and the response itself

Third, as the example prompts have shown, the more specific details about the project, its environment, and stakeholders are provided, the better the AI chatbot's responses will be. However, this is likely to conflict with data protection and privacy. Users should be extremely careful about the disclosure of any sensitive information. Not only is there the problem of the AI chatbot adding this information to its knowledge base and using it to generate answers, but it also provides the chatbot with information to infer ("to guess") further sensitive information. A solution might be to further develop AI chatbots as business-internal solutions which are separated from the outside world. In this case, the underlying LLMs could be trained with company-internal data, e.g., on past projects. This would provide the AI chatbot with company-specific knowledge and push it a step further to a digital assistant.

VI. CONCLUSION

The paper explored the possibilities of using an AI chatbot, such as ChatGPT, to support project managers in their daily work. Patterns of AI-assisted project management tasks were identified, showing the wide range of possible collaborations between project managers and AI chatbots. Examples include using the AI chatbot to generate ideas and insights, or to act as a sparring partner. The skills required for successful collaboration were also presented. Special emphasis was placed on prompt engineering. The ability to write concise prompts is crucial to achieving results that are useful and not superficial. A case study demonstrated how an AI chatbot could be used. Examples of prompts were shown for stakeholder management activities. Finally, potentials, risks and solutions were discussed.

REFERENCES

[1] T. Wu, “Digital project management: rapid changes define new working environments,” Journal of Business Strategy, vol. 43(5), pp. 323–331, 2022, https://doi.org/10.1108/JBS-03-2021-0047.

[2] A. Nieto-Rodriguez and R. Viana Vargas, “How AI Will Transform Project Management,” Harvard Business Review. https://hbr.org/2023/02/how-ai-will-transform-project-management [retrieved: August, 2023].

[3] T. Teubner, C. M. Flath, C. Weinhardt, W. van der Aalst, and O. Hinz, „Welcome to the Era of ChatGPT et al.,” Business & Information Systems Engineering, vol. 65(2), pp. 95–101, 2023, https://doi.org/10.1007/s12599-023-00795-x.

[4] Institute of Project Management, “What Can ChatGPT Do for Project Management?, “ 2023, https://www.projectmanagement.ie/blog/what-can-chatgpt-do-for-project-management/ [retrieved: August, 2023].

[5] L. Weidinger et al., “Taxonomy of Risks posed by Language Models,” Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 214–229, https://doi.org/10.1145/3531146.3533088.

[6] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, “On the Dangers of Stochastic Parrots,” Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 610–62, https://doi.org/10.1145/3442188.3445922.

[7] A. Chernyavskiy, D. Ilvovsky, and P. Nakov, “Transformers: ‘The End of History’ for Natural Language Processing?,” Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2021, Lecture Notes in Computer Science, vol. 12977, 2021, https://doi.org/10.1007/978-3-030-86523-8_41.

[8] B. Agüera y Arcas, “Do Large Language Models Understand Us?,” Daedalus, vol. 151(2), pp. 183–197, 2022, https://doi.org/10.1162/daed_a_01909.

[9] S. Mitrović, D. Andreoletti, and O. Ayoub, “ChatGPT or Human? Detect and Explain. Explaining Decisions of Machine Learning Model for Detecting Short ChatGPT-generated Text,” 2023, arXiv preprint available at https://arxiv.org/pdf/2301.13852 [retrieved: August, 2023].

[10] J. White et al., “A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT,”, 2023, arXiv preprint available at http://arxiv.org/pdf/2302.11382v1.

[11] L. S. Lo, “The CLEAR path: A framework for enhancing information literacy through prompt engineering,” The Journal of Academic Librarianship, vol. 49(4), 2023, https://doi.org/10.1016/j.acalib.2023.102720.

[12] L. Giray, “Prompt Engineering with ChatGPT: A Guide for Academic Writers,” Annals of Biomedical Engineering. Advance online publication, 2023, https://doi.org/10.1007/s10439-023-03272-4.

[13] Z. Wang, Q. Xie, Z. Ding, Y. Feng, and R. Xia, “Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study.”, 2023, arXiv preprint available at http://arxiv.org/pdf/2304.04339v1.

[14] Project Management Institute, “A guide to the Project Management Body of Knowledge (PMBOK® Guide; 7th ed.),” 2021.

[15] R. J. Martinelli and D.Z. Milosevic, “Project Management ToolBox: Tools and Techniques for the Practicing Project Manager,” 2nd ed., Wiley, 2016.

[16] O. Schibi, “Managing stakeholder expectations for project success: A knowledge integration framework and value focused approach,”, 2016, J. Ross Publishing.

Verifiable Labels for Digital Services: A Practical Approach

Maël Gassmann
School of Engineering and Computer Science
Bern University of Applied Sciences
Biel/Bienne, Switzerland
email: mael.gassmann@bfh.ch

Annett Laube-Rosenpflanzer
School of Engineering and Computer Science
Bern University of Applied Sciences
Biel/Bienne, Switzerland
email: annett.laube@bfh.ch

***Abstract*—Users often feel unsafe and unsecure when they use digital services. For normal users without technical backgrounds, it is difficult to recognize if a website is genuine. This makes them vulnerable to phishing attacks. In order to solve this issue, many organizations use corporate designs or logos to guide users through their websites. However, all this can be easily copied. More technical means are also advertised as solutions, like trusted Transport Layer Security (TLS) certificates or Extended Validation (EV) certificates, but they are too complicated for non-technical users and barely make any difference. Right now, users lack a way to easily verify that they are using the intended digital service. A pure visual indication, e.g., with simple graphic files or technical means users do not understand, is not sufficient. Using the TLS Public Key Infrastructure (PKI), verifiable labels will use these certificates to bind an entity's label to the certificate's key pair. Instead of trying to provide automated trust, verifiable labels acknowledge the presence of ill-intentioned entities. In order to differentiate them from trustworthy actors, cryptography is used to define facts, which allows a user client to form easily understandable recommendations and analyze a certain actor's reputation. Thus, allowing users to naturally develop an opinion and make an educated guess as to whether an entity is worthy of their trust or not. The end goal would be that most business websites that ask for some level of trust would use verifiable labels; this way, websites with bad or no labels would start to stand out.**

***Index Terms*—Trust; Anti-Phishing; Digital Label.**

I. INTRODUCTION

Nowadays, if website owners want to try and certify an accordance to a label, one sole option is at their disposal: The usage of copyable and thus untrustworthy digital representations, such as pictures or electronic documents. Without having to make any distinction between true and false claims, it can already be deduced that it has as much value as a self-proclamation and is at least hard and inconvenient, if not impossible, to verify. This is leading naïve Internet users to give their trust to a service unworthy of any. Moreover, it is far from affecting only a limited number of people, as since 2020, phishing attacks have become by far the most common type of attacks performed by cybercriminals [1]; 41% of security incidents begin with the initial access gained by a phishing attack [2]; approximately 1.385 million new phishing web pages are set up each month [3]; and overall, phishing is in the top three cybersecurity threat trends [4].

The real problem is there; a verifiable label would truly add value to anybody's Internet experience by directly reducing the impact of phishing. One standalone example of such a label is the 'Digital Trust Label' [5]. However, it is very limited in its range of action. Verifiable labels strive to establish a distributed framework for the development of labels in general and enhance user friendliness.

The rest of the paper is structured as follows: Section II analyses the current state of Internet related technologies; Section III describes the concept of verifiable labels, its underlying infrastructure and protocols; Section IV explains how the concept was adapted to a working prototype; finally, the work is concluded in Section V.

II. STATE OF THE ART

A. TLS Certificates

Based on Public Key Infrastructure (PKI) to establish chains of trust and using X.509 certificates to bind web-servers to key pairs and domain names, Transport Layer Security (TLS) certificates are nowadays widely used to encrypt communications on the Internet [6]–[8]. These so-called chains of trust are all built upon an entrusted third party—a root of trust—that certifies the trustworthiness of other entities, which in turn are sometimes allowed to do the same. Such entrusted third parties are called Certificate Authorities (CA), as shown in Figure 1.

Furthermore, the X.509 certificate itself can contain a variety of different claims. For instance, one way to bind a certificate to a server is to include its specific domain inside.

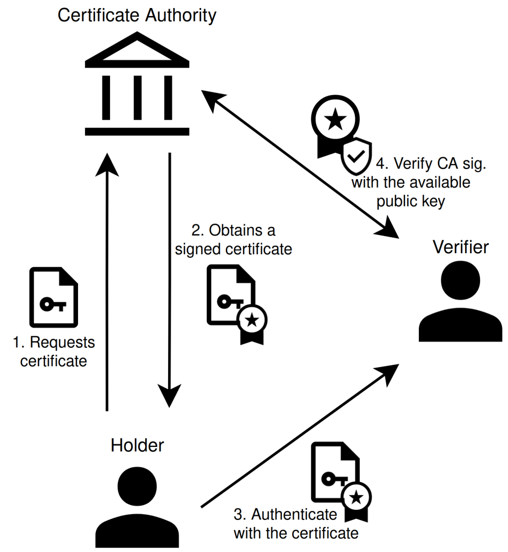

Certificate Authority
4. Verify CA sig. with the available public key
2. Obtains a signed certificate
Verifier
1. Requests certificate
Holder
3. Authenticate with the certificate

Figure 1. Minimalistic representation of a PKI.

In the case of TLS certificates, there are three major types of X.509 certificates that are used.

a) Domain Validated (DV) Certificate: These are the most basic types of certificates. The CA will only verify that the applicant has control over the requested domain name; this is typically done through email validation. More recently, the Automatic Certificate Management Environment (ACME) protocol allowed CAs to issue DV certificates without any intervention from their side [9]. When the ACME protocol is used, the certificate can be obtained free of charge.

b) Organisation Validated (OV) Certificate: Not only is the domain ownership verified, but also the legal existence and physical location of the applicant. Automation is, of course, out of the question. Such a certificate can be obtained for a range of 200 to +1000 USD per year [10].

c) Extended Validation (EV) Certificate: EV certificates undergo the most rigorous validation process; this includes all steps taken for OV certificates, including legal status, operational existence, and telephone verification [11]. The price range goes from 400 to +1700 USD per year [10].

OV and EV certificates were advertised as a way to prevent the customers' users from being prone to phishing, as the web browser, recognizing an EV certificate, used to display a green indicator containing the entity's legal name. Thus, users who knew of that distinction would change their behavior according to the level of certification displayed. However, studies showed that user behavior did not alter, and polls showed that the padlock's meaning was not understood correctly. Worse even, security researchers were able to prove that some EV certificates could be gotten with colliding organization names, which could be quite misleading as the domain would be hidden by the legal name in some browsers. That is why, in September 2019, most browsers stopped displaying any direct visual distinctions between DV, OV, or EV certificates, which invalidates the main selling point of these products [12].

Moreover, because CAs are private companies, the regulations are not always followed with the same rigor, as not all validation processes can be automated. A PKI infrastructure is always very sensitive to mistakes, and the verification process has proven to not be enough [13]. However, one thing is sure: TLS certificates do a good job of binding a domain name to its corresponding server, which holds the key pair. Especially with the help of the ACME protocol.

B. Decentralised Identifiers

In opposition to the traditional central authoritative system that PKI represents, Decentralized IDentifiers (DID), an open standard in active development, is part of a broader movement that strives towards decentralized identity. A DID resolves to a DID document—typically hosted on a decentralized network or infrastructure, e.g., a block chain or a distributed ledger—which contains a set of public keys, authentication methods, service endpoints, a time-stamp to keep an audit history, and a signature for its integrity [14].

A Verifiable Credential (VC) is a claim created from the key pair of a DID (the issuer) and is issued to a holder's wallet by using a holder proof. This holder proof varies greatly between implementations, and efforts are being made to standardize it. Self-Sovereign Identity (SSI) solutions strive to provide a way to assert, present, and verify claims in a decentralized manner [15].

The current limitations are that not everything is yet standardized. For instance, linking an existing TLS certificate key pair to a DID misses specifications. More than that, VCs need a wallet with a functional, universal holder proof. In the end, this solution is not yet defined and widely used enough to be applied in this specific use case.

III. CONCEPT

A. Different Perspective

The root of the verifiable label concept lies in a shift of perspective on what trust is and how can it be made identifiable to an end-user. As TLS EV certificates proved, a seemingly good concept will still need to be understood by anyone who uses the Internet in order to have any impact, especially by those who do not have any technical background. First, one must understand how trust is perceived as a concept alone; for this, a philosophical definition of trust is adequate.

'Trust is important, but it is also dangerous. It is important because it allows us to depend on others—for love, for advice, for help with our plumbing, or what have you—especially when we know that no outside force compels them to give us these things. But trust ***also involves the risk*** *that people we trust will not pull through for us, for if there were some guarantee they would pull through, then we would have no need to trust them.* ***Trust is therefore dangerous****. What we risk while trusting is the loss of valuable things that we entrust to others, ...'* [16]

That is, when a person *decides* to place their trust in someone else, they know about the risks—risks that can be clearly identified as they are based on facts.

Instead of distributing a trust people have to blindly believe in, verifiable labels proposes the idea of providing simple facts about Internet entities so that anyone with no technical background can, in a reasonable time, learn how to navigate the Internet with a valid sense of which entity deserves their trust. Trust is, after all, an individual decision, and users must be able to make that decision for themselves and not have to rely on a third-party organization they do not even know exists.

To do this, cryptography is paramount, as it is the sole option available to make any virtual information a tangible fact. The system must be implemented on top of the currently widely used Internet cryptographic technologies (e.g., TLS certificates) in order to have any chance of success, while also striving to be flexible and pushing towards more decentralized technologies (e.g., blockchains) because they provide a non-authoritative infrastructure.

B. Definitions

1) VERIFIABLE LABEL
   A verifiable label is a data structure that is bound to two domain names; the holder's and the issuer's. This is done by signing the label with both cryptographic identifiers

(e.g., TLS certificate). Other attributes will be present in order to allow users to derive a clear reputation for each label declared on a website and thus, the direct trustworthiness of the web entity.

2) VERIFYING USERS
Simple users that visit a website. If a valid label is detected, the user will be able to see it, list facts that concern it, and develop an idea of this label's reputation.
3) LABEL-WORTHY WEB-ENTITY
Such an entity can request a label from its corresponding issuer. If an issuance occurs, they can display their digital label on their website, which is visible and verifiable by anyone. It cannot be copied.
4) ISSUER
The entity that can verify and decide of its own accord who is worthy of being labeled. It will keep a record of who has been issued its label and can confirm it.
5) TIME STAMP AUTHORITY (TSA)
The backbone of the concept is here; this time stamp authority [17] will follow specific automated guidelines. While the automation makes sure that every issuer plays by the same rules, the guidelines aim at enforcing duplicate label prevention. As all issuers need an unexpired certificate, they will have to issue renewal requests, which, built upon one another, start to create a reputation. Every time-stamped issuer certificate will be stored in a publicly readable storage.

C. Protocol

a) Issuance of a label: Figure 2 depicts it. This is the least protocoled part of the system. A website must create a verifiable label and sign it with its TLS certificate. This ensures that the draft label certificate is bound to the domain name and also comes from the stated owner. The incomplete digital label can be sent to the issuer; no channel is specified. If the issuer decides to accept the request, it will sign it with its own TLS certificate, add the new signature to the now complete verifiable label, and send it back. Finally, the issuer save a copy of the signature and requester's domain in the list of its own draft certificate. In order to make a valid issuer certificate out of this draft, the issuer has to request a new time-stamp and signature from the TSA, as explained in Figure 3.

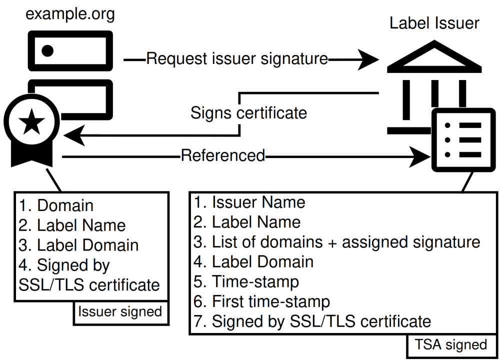


Figure 2. Certificate Issuance

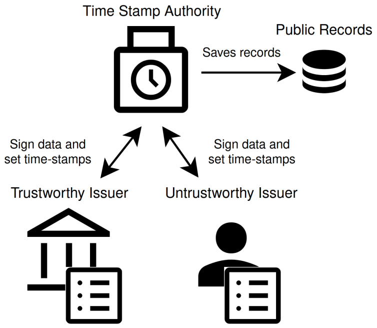


Figure 3. Time Stamp Authority

b) Issuer Certification: As stated before, an issuer's trustworthiness is defined by its own reputation. This reputation is built with time and the help of an automated time-stamp authority. The TSA's role is to reissue new signatures—necessary for the issuer's label to be considered cryptographically valid—and time-stamps to all requesting issuers that are on the brink of expiration. As it does so, it will first store a copy of the renewed certificate in a publicly readable storage and then send it back. However, if the issuer is new, i.e. does not possess a first time-stamp, the TSA will have a look at the requested label name, domain name, issuer name, and all fields that might be prone to confusing a human being if not filled with good intentions. If it is considered not to be confusing as well as not a duplicate of any existing labels, the web entity will receive its first time-stamp and signature, making it an issuer.

c) Validation and interpretation client: As a user with the verifiable label validation and interpretation client installed navigates the Internet, the client will try to detect if a digital label is present on the currently visited website. If it proves to be the case, the certificate validation process will begin, as shown in Figure 4. The first step consists of verifying the label's link with the domain and TLS certificate, that is, making sure the signature is correct and that the domain corresponds to the browser URL. On success, the next step will

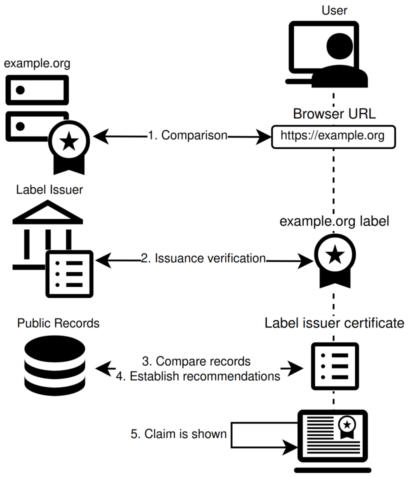


Figure 4. Certificate Validation

be to check the listed label domain for the issuer certificate. Records are to be compared, signatures are to be verified for the same reason as before, and the domain of the label must be found in the list. At last, if everything succeeds again, the client will search the label's public records for what could be considered pertinent to derive a reputation from. This could include refreshment regularity, time of existence, measurement of the movement of certified entities, activity ratio, and number of requests. How these measurements should be interpreted is now unknown; in order to gain more insights, the system would have to be put into place, and practical testing could allow recommendations to be derived. If a big enough dataset of reputation measurement could be collected, a machine learning algorithm could do a good job at identifying untrustworthy labels and trends in how attackers are trying to infiltrate the system. Staying flexible in this interpretation and allowing it to evolve is paramount; one fixed description of what an untrustworthy label is would never be enough; the system needs to be as resilient as the spoofers.

D. Purpose

Accurate protection is possible if we assume that a majority of web entities have adopted this digital label system. Only then would websites with bad or no certifications start to stand out, especially when they require trust, e.g., when they ask for credit card information or propose services.

IV. Implementation

A prototype has been implemented following a minimal working system approach. Furthermore, since different underlying technologies exist, extensibility is a top priority.

A. Verifiable Label Time Stamp Authority (VLTSA)

Starting from the very root of the system, the VLTSA is not a time stamp authority. Preferring a simple method, a library implementing a RFC 3161 [17] client interface was used to interact with an external TSA to provide the time-stamps. A TLS certificate was used as a way to warrant the need for issuers to still directly request the time-stamps from the VLTSA, as a third signature from that certificate is required to make the issuer certificate valid. This server software consists of a simple HTTP API with two paths: the POST method on '/sign' and the GET method on '/get_records'. Meaning it also acts as the publicly readable storage. All of this has been implemented in the most minimalistic way, with abstract interfaces of 'Storage', 'API', and 'Signer'. That is where flexibility is; the logical part of what makes the VLTSA is detached from all other components that could find better long-term alternatives (e.g., more resource-efficient or different time-stamp sources such as a blockchain).

B. Verifiable Label Issuer Client (VLIC)

The simple command-line client has persistent storage and saves all valid given arguments. If provided with a valid request, it will add a domain to its certificate and generate a valid verifiable label certificate (.vlcert), which can be sent back to the holder through any channel. It can issue a signing request to the VLTSA on demand. And, if successful, it will save the verifiable label issuer certificate (.vlicert). This vlicert has to be exposed on the label domain's web-server root as 'cert.vlicert'.

C. Verifiable Label Holder Client (VLHC)

This simple command-line client with no persistent storage can only be used to generate a vlcert without the issuer signature. It has to be manually sent to the issuer. Once a valid vlcert is in the holder's possession, it has to be exposed on its domain's web-server root as 'cert.vlcert'. This prototype thus only allows for one vlcert per holder.

D. Browser Extension Analyzer

A browser extension was a mandatory component of the client, as the active URL has to be accessed to perform the first cryptographic tests. However, the specific environment did not provide any way to download a TLS certificate for a specified domain, which blocked further development. More research showed that by using the native messaging interface, the browser extension can communicate data to an underlying program. Using this method, a cryptographic verifier was developed. It sends back the necessary data to perform a reputation analysis and is then displayed in a panel.

V. Conclusion

A solution that allows Internet entities to create verifiable labels, as well as aiming to reduce fraud was proposed. Based on TLS certificates and time stamp authorities, the current prototype stays flexible and, even if simplistic, already implements all the necessary cryptographic tools.

Future work could investigate the following directions:

- Conduct a field study of a live setup and user experience.
- Study relevant metadata for a good reputation evaluation.
- Provide a comprehensive User Interface (UI) for computers and phones.

References

[1] FBI Internet Crime Complaint Center, "Internet Crime Report 2020," 2020, *Online*. Available: https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf [retrieved: 08, 2023].

[2] IBM Security, "X-Force Threat Intelligence Index 2023," 2023, *Online*. Available: https://www.ibm.com/reports/threat-intelligence [retrieved: 08, 2023].

[3] M. Swindells, "How many phishing emails are sent daily in 2023? 11+ statistics," 2023, *Online*. Available: https://earthweb.com/how-many-phishing-emails-are-sent-daily/ [retrieved: 08, 2023].

[4] Cisco, "Cybersecurity threat trends: phishing, crypto top the list," 2021, *Online*. Available: https://umbrella.cisco.com/info/2021-cyber-security-threat-trends-phishing-crypto-top-the-list [retrieved: 08, 2023].

[5] Swiss Digital Initiative, "The Digital Trust Label," Swiss Digital Initiative, 2022, *Online*. Available: https://digitaltrust-label.swiss/ [retrieved: 08, 2023].

[6] T. Dierks and E. Rescorla, "RFC 5246: The Transport Layer Security (TLS) Protocol, Version 1.2," August 2008, *Online*. Available: https://www.rfc-editor.org/rfc/rfc5246 [retrieved: 08, 2023].

[7] IETF Trust, "The Transport Layer Security (TLS) Protocol Version 1.3," August 2018, *Online*. Available: https://datatracker.ietf.org/doc/html/rfc8446 [retrieved: 08, 2023].

[8] R. Housley, W. Polk, W. Ford, and D. Solo, “RFC 3280: Internet X.509 Public Key Infrastructure,” April 2002, *Online*. Available: https://www.ietf.org/rfc/rfc3280.txt [retrieved: 08, 2023].

[9] Internet Engineering Task Force (IETF), R. Barnes, J. Hoffman-Andrews, D. McCarney, and J. Kasten, “Automatic Certificate Management Environment (ACME),” March 2019, *Online*. Available: https://datatracker.ietf.org/doc/html/rfc8555 [retrieved: 08, 2023].

[10] DigiCert, “Compare TLS/SSL certificates,” 2023, *Online*. Available: https://www.digicert.com/tls-ssl/compare-certificates [retrieved: 08, 2023].

[11] CA/Browser Forum, “EV SSL Certificate Guidelines,” 2022, *Online*. Available: https://cabforum.org/extended-validation/ [retrieved: 08, 2023].

[12] Chromium Docs, “EV UI Moving to Page Info,” 2019, *Online*. Available: https://chromium.googlesource.com/chromium/src/+/HEAD/docs/security/ev-to-page-info.md [retrieved: 08, 2023].

[13] C. Cimpanu, “Extended Validation (EV) Certificates Abused to Create Insanely Believable Phishing Sites,” 2017, *Online*. Available: https://www.bleepingcomputer.com/news/security/extended-validation-ev-certificates-abused-to-create-insanely-believable-phishing-sites [retrieved: 08, 2023].

[14] W3C, “Decentralized Identifiers (DIDs) v1.0: Core architecture, data model, and representations,” W3C Recommendation 19 July 2022, *Online*. Available: https://www.w3.org/TR/2022/REC-did-core-20220719/ [retrieved: 08, 2023].

[15] W3C, “Verifiable Credentials Data Model v1.1,” W3C Recommendation 03 March 2022, *Online*. Available: https://www.w3.org/TR/2022/REC-vc-data-model-20220303/ [retrieved: 08, 2023].

[16] C. McLeod, “‘Trust’, The Stanford Encyclopedia of Philosophy (Fall 2021 Edition),” 2021, *Online*. Available: https://plato.stanford.edu/archives/fall2021/entries/trust [retrieved: 08, 2023].

[17] C. Adams, P. Cain, D. Pinkas, and R. Zuccherato, “Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP),” RFC 3161, August 2001, *Online*. Available: https://www.rfc-editor.org/rfc/rfc3161 [retrieved: 08, 2023].

Teachers, Reload Your Toolboxes: Using a Task Design Framework and ChatGPT to Generate Motivational Exercises for Female Learners in Computer Science

Sibylle Kunz and Claudia Hess
IT & Engineering
IU International University of Applied Sciences
Erfurt, Germany
e-mail: sibylle.kunz@iu.org; e-mail: claudia.hess@iu.org

***Abstract*—Challenging learners at school with interesting activities that foster competencies that are later required at university is an important part of a teacher's job. But creating these tasks is time-consuming, and often these activities are traditionally more adapted to the interests of male than female learners, especially in Science, Technology, Engineering and Mathematics (STEM) courses. This, among other factors, contributes to the phenomenon of "the leaky STEM-pipeline" – losing the girls over the course of time before they chose their university career or vocational education. This paper uses an iterative approach to describe how a Large Language Model (LLM), such as ChatGPT can be used to generate interesting and female-oriented learner tasks based on a formerly developed framework of categories in a morphological analysis that helps to design computer science exercises in a structured and consistent way. The LLM is provided with the relevant features of the framework and iteratively asked to generate exercises with respect to girls' interests and corresponding grading criteria, wrapping the task in a motivating story. The results show how a generative artificial intelligence can be used as a productive approach to the creation of teaching material targeted at a specific audience.**

Keywords-STEM; ChatGPT; learners' task; computer science lessons; morphological analysis; prompt engineering.

I. INTRODUCTION

The lack of girl students in Science, Technology, Engineering and Mathematics (STEM) majors has been a persistent issue that has garnered increasing attention from researchers. One contributing factor to this phenomenon is the absence of self-confidence and a sense of inadequacy in meeting the requirements of universities. This leads to a continuous decline in the number of girls interested in pursuing STEM majors during their school years, often referred to as the "leaky STEM pipeline" [1]. This also counts for Computer Science (CS). One adjustment lever is to systematically promote the competencies needed for studying computer science which can be done by a conscious lesson and task design. The quality of school lessons and tasks in STEM subjects has a major impact on learners regarding their career choices [2]. But learning activities in computer science often follow similar patterns like programming tasks as a kind of "finger exercise", often designed by male authors or teachers: "The institutionalization of negative representations of women in CS demonstrates a severely unbalanced male influence in the field. […] Since CS has long faced a shortage of women, it is unsurprising that the materials reflect predominantly male-centered perspectives. This results in continuous generations of students learning from materials that are rooted in inequality. Subsequently, the continued use and advancement of the field using biased standardized materials, forms a vicious cycle." [3].

It seems that these tasks sometimes do not appeal to girls who are interested in other exercise contexts [4]. However, adapting exercises and their "contextual stories" is a time-consuming process for teachers. But not only context matters, it is also the way the exercise is done: for example, many girls seem to prefer group work over solitary training – which impacts task planning, distribution of work among team members, documentation, grading, etc. [5]. The overall goal – building up the competencies for a university career – must not be left out of sight.

This paper discusses how to use a Large Language Model (LLM) to design tasks for computer science lessons that target a specific group of learners. Due to the urgent need to better engage girls and to motivate them to pursue careers in STEM, we focus on creating exercises that are responsive to girls' interests. In the scope of a case study, concrete examples of the use of a LLM – in our case ChatGPT – are shown. We demonstrate current limitations of these technologies and discuss ethical considerations, especially with respect to reinforcing existing biases and gender stereotypes. Our goal is to provide teachers and educational staff with a blueprint for using Artificial Intelligence (AI) technology in a responsible and well-thought manner to create practical exercises in computer science lessons for a certain audience.

To this end, the paper is structured as follows: Section II provides a theoretical foundation on the morphological analysis containing the task dimensions. Section III outlines the necessary steps of planning, creating and testing female-oriented learner activities and shows how teachers can use the framework and/or AI Chatbots as new tools. Section IV presents an iterative approach using different prompting techniques to create new tasks on topics of interest to girls. A case study in Section V illustrates the implementation of this approach and the corresponding results, which can easily be adapted to teachers' needs. Section VI addresses the problem of gender bias in AI Bots and how it can be used for good to eradicate a real-world bias. The article is summarized in Section VII and Section VIII outlines further research.

Copyright (c) IARIA, 2023. ISBN: 978-1-68558-115-2

II. THEORETICAL FOUNDATION

Competency-based education has proven to be a promising approach in STEM education and teaching [6]-[8]. Therefore in [9], a framework for designing appropriate learners' tasks for female-responsive computer science lessons was developed with regard to the competencies most needed when starting a university career in the field of computer science. After identifying the most important dimensions needed to describe a task, those dimensions were grouped in a morphological analysis, resulting in a Zwicky box where instructors can choose one characteristic of each dimension to construct an exercise. Table 1 shows the Zwicky Box.

TABLE I. FRAMEWORK FOR DESIGNING FEMALE-RESPONSIVE LEARNING ACTIVITIES IN COMPUTER SCIENCE BASED ON [9]

Learning Activity Dimension	**Characteristics/Choices**			
Competency	Inventive	Thinking concretely	Analytic/ thinking in an abstract way	(Others)
Class organization	Single learner	Partnership work, two learners	Group-/ teamwork 3-6 learners	-
Gender pairing	Single-gender	Mixed-gender	-	-
Duration	Asynchronous during several lessons based on work packages	Synchronous during one lesson	Synchronous in more than one lesson	-
Task type	Focus on programming an algorithm	Focus on Human Computer Interaction (HCI-)design	Focus on physical construction	(Others)
Tools/ Media	Software tools	Real-world tools, hardware-based	Combination of hardware and software tools	-
Number of possible solutions:	There's only one correct sample solution	Many sample solutions can be correct	Unlimited number of correct solutions	-
Assessment (frequency)	No support, no interaction	Sporadic support during lessons	Regular support in and between lessons	-
Assessment (method)	Formative during creation	Summative when finished	-	-
Documentation required	Yes	No		
Grading	Yes	No		

Teachers can design tasks by choosing one item from each row and giving appropriate instructions for each dimension. But they still have to do some "storytelling" to describe a task subject, e.g., the kind of artefact to be developed and a context that is interesting to learners and connects the task to a real-world problem. This is where AI Chatbots based on LLMs come in. They have enough contextual background to come up with interesting exercise concepts, as will be shown in the next sections.

III. DEVELOPING LEARNERS' TASKS IN A CYCLE

To design and test exercises for computer science lessons, a cycle model can be used. It consists of activities for planning the tasks, letting the learners work on them, providing feedback, grading, and optimizing or updating the task, as shown in Figure 1.

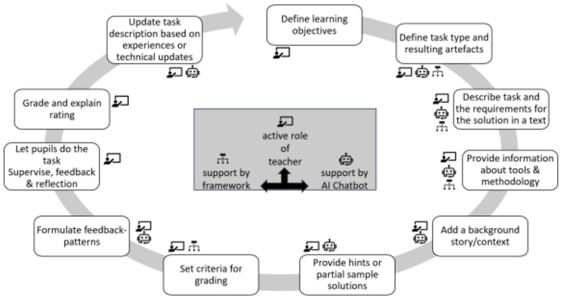

Define learning objectives
Define task type and resulting artefacts
Describe task and the requirements for the solution in a text
Provide information about tools & methodology
Add a background story/context
Provide hints or partial sample solutions
Set criteria for grading
Formulate feedback-patterns
Let pupils do the task Supervise, feedback & reflection
Grade and explain rating
Update task description based on experiences or technical updates
active role of teacher
support by framework
support by AI Chatbot

Figure 1. The cycle of exercise development

Many of the activities in this process can be supported by the framework and additionally by an AI Chatbot. Of course, the teachers themselves keep control over all activities and act as coaches for the pupils. (Note: The fact that the learners also will make more and more use of AI Chatbots for solving the tasks is not represented in the figure, since this is subject to negotiation processes in educational institutions on a meta level.) But it is important to state that one shouldn't solely rely on AI for the whole process: "Using AI-generated teaching materials (e.g., lesson plans) and information (e.g., content for mini-lecture slides) without critical examination can create additional harms for students and negatively impact teaching and learning" [10]. To do so would mean sacrificing a significant amount of valuable knowledge and understanding in the field of education.

IV. AN ITERATIVE APPROACH TO CREATING SUCCESSFUL PROMPTS BASED ON THE FRAMEWORK

We ask ourselves the question: how can the framework described in the previous section be used for automated generation of tasks by a LLM, like ChatGPT? Since LLMs are trained on a broad variety of textual materials, they are able to create answers from different perspectives, change sentiment and tone for a certain audience, and come up with background stories and information. This can be useful when teachers want to tailor their exercises to different ages and genders and provide a textual "hook" for learners to get them interested and started.

To enable a LLM to “create” a unique output, as much context as possible must be provided – so the basic idea is to feed the framework into the LLM first and then ask it to make use of it. In this case, the framework dimensions, as well as the possible items for each dimension, are needed as input. Apart from these details, the quality and usability of the resulting task depends also on the length of output, since the task description has to be specific about the problem, the desired form of results and the criteria for assessment and grading the exercises.

This means that a lot of prompt engineering has to be done before the results become usable [11]. Since prompt engineering often means starting with a rough draft and then fine-tuning the input, as well as asking the system to elaborate on certain aspects, while keeping the context, an iterative approach is needed. The design framework can support the quality of the generated tasks by considering several dimensions at the same time.

Several approaches can be distinguished in prompt engineering: zero-shot prompting, one-shot prompting and few-shot prompting. zero-shot prompting does not provide the AI Bot with any individual examples or additional information. Due to the broad training data base, the bot delivers answers in the correct context. One-shot prompting and n-shot-prompting (where n is any number greater than one) enrich the prompt with one or more examples or demonstrations, allowing the bot to make use of similarities. [12] [13]. Examples for these approaches can be found in the case study in Section V.

In general, the procedure for successfully developing interesting and female-oriented tasks based on the framework and an AI Chatbot can be outlined like this:

1. Preparation: Provide ChatGPT with the problem (i.e., create tasks for computer science lessons that might interest girls) and the framework. This can be done by listing all dimensions and the choices for each dimension. One difficulty is recognizing if two options exclude or impede each other, like solving a large task in a single lesson. Other requirements are that all dimensions are being addressed in the answer which requires a longer output size.
2. Design: Run the prompt and check the resulting suggestions (i.e., the task for computer science lessons) for completeness and usability. If things seem to be missing or do not make sense, add more contextual information and re-run the request. This might lead to more complex task descriptions. Ask ChatGPT to elaborate on the “story” of the task to make more understandable what is expected of the learners and to motivate them.
3. Evaluation: Check if enough information is included so that the task can be solved by the learners. Otherwise, learners might be frustrated at an early stage, worsening the problem of lack of interest in computer science.
4. Creating AI-based sample solutions: Depending on the type of task, AI Chatbots can also be used to generate parts of sample solutions. This refers to programming or interface-design tasks that do not require hardware to be manipulated (like in robot assembly or circuit design – but here the bot might be asked for part lists and assembly hints). Nevertheless, ChatGPT does (no longer) support the creation of large amounts of code. But it can outline basic structures, object classes or deliver code-snippets, as well as explain code segments in natural language. This can be used to draft the outline of a sample solution for teachers or to generate tutorials for learners.
5. Test: When finished, the resulting artefact is a task description that can be used in a computer science lesson. Before using it, the teacher should check it again for consistency and level of difficulty to ensure the learners can cope with it and have already sufficient theoretical background knowledge to get at least started.
6. Optimize: Since AI Chatbots can save conversations for a longer time, teachers can get back to their last state and re-work existing tasks or even provide feedback about the experiences made in the classroom and add more context.

The following sections demonstrate how exercises can be effectively created using this iterative approach. It presents AI Chatbot prompts that can easily be copied and evaluated or modified to obtain the desired results (e.g., for a certain age group or with respect to geographical or cultural aspects).

V. CASE STUDY: USING CHATGPT TO GENERATE LEARNING ACTIVITIES APPEALING TO GIRLS BASED ON A COMPETENCY-ORIENTED FRAMEWORK IN AN INTERATIVE APPROACH

A. A one-shot-prompt for simple learner acitivities

To get a feeling for the “knowledge” of the LLM about potential subjects suitable for girls, the first prompt (a zero-shot prompt) only asked for a task related to user-interface design. This is a task type suitable for pairs or teams, as well as single learners, and can be adapted to a variety of scenarios. Table II shows the prompt and the corresponding result.

TABLE II. EXAMPLE OF ONE-SHOT PROMPT TO GET STARTED

Prompt: “Create a school task for a computer science lesson. The task should appeal to girls and the goal should be to create a user interface. Use up to 500 words.”
Result: Task Title: Design Your Dream Fashion App Objective: The goal of this task is for students to practice basic computer science principles by creating a user interface for a fashion-app. This activity aims to appeal to girls with an interest in fashion design while also fostering essential computer programming skills. Overview: […] In this task, students will take on the role of app designers, and their assignment is to create a unique fashion app that offers users an interactive and creative experience. The app should have features that girls would enjoy, such as: 1. Personal style profile creation

2. Outfit recommendations based on weather and events
3. A virtual wardrobe to mix and match outfits
4. Fashion news, trends, and […]

ChatGPT not only came up with a suitable task description, but also with a list of features, but presented no further details or hints how to solve the task.

B. Using the competency-based framework

In the next step, ChatGPT was given a short, but complete description of the ten categories forming the framework plus a hint regarding categories where only one could be chosen. Table III shows the resulting prompt.

TABLE III. PROVIDING THE FRAMEWORK

Prompt: “I will give you a framework for designing interesting tasks for lessons ins computer science for school girls aged 12-19. I will list the important properties and the possibilities for each category. You will create a task according to these categories. Make sure the properties you choose fit well together.
The properties are:

- Fostered Competency should be one or more of these: like “thinking concretely”, "analytic”, “thinking in an abstract way”, “being able to visualize”, “thinking holistically”
- Class organization should be either individual work, working in pairs or working in groups
- Gender pairing should either be girls-only or mixed groups
- Duration should be either 1 lesson or several lessons
- Task type should be either programming, HCI-design, constructing a switch board, building a robot or any other computer science related work
- Tools and Media should be used, list sample tools to fulfill the task
- Number of possible solutions: either only one solution is correct or the task can have different solution variants or even an infinite number of solutions
- Assessment: the task can be graded or not graded
- Documentation: the task could require a written documentation or just a self-explaining artefact
- Grading: describe if the task can be easily graded concerning certain quality properties

Chose topics that girls are generally interested in. The task description should contain of a maximum of 2000 words and contain suggestions for all ten properties.”

The result provided by ChatGPT is shown in Table IV. It is noticeable that it makes use of all given dimensions and comes up with a lot of details concerning requirements, software tools and hints for grading. It even “explains” the goals to the learners in the end.

TABLE IV. A GENERATED TASK

Result: Task: Designing a Mobile App for a Sustainable Fashion Community

Fostered Competency: Thinking holistically, thinking in an abstract way, being able to visualize.
Girls will need to think about the overall concept of a sustainable fashion community, consider abstract ideas related to fashion and sustainability, and visualize the app's design and functionality.

Class Organization: Individual work.
This task can be completed individually, allowing each girl to explore her creativity and problem-solving skills independently.

Gender Pairing: Girls-only.
Girls can work on this task in a girls-only setting, creating a comfortable environment that encourages participation and collaboration among female students.

Duration: Several lessons.
This task requires multiple lessons to give students enough time to research, plan, design, and prototype their mobile app concept.

Task Type: HCI-design (Human-Computer Interaction design) and programming.
Girls will design the user interface and user experience (HCI-design) of a mobile app for a sustainable fashion community. They will also have the option to implement a basic prototype using programming tools.

Tools and Media:
HCI-design tools: Figma, Adobe XD, Sketch, or any other design software the school has available.
Programming tools: Thunkable, App Inventor, or any other beginner-friendly app development platform.
Research materials: Access to the internet or school library for researching sustainable fashion practices, existing fashion communities, and mobile app design inspiration.

Number of Possible Solutions: Different solution variants.
Girls can come up with different designs and functionalities for their sustainable fashion community app. While there may be common elements, each student's app will reflect her unique perspective and creativity.

Assessment: Graded task.
This task can be graded based on specific criteria, such as the overall concept and functionality of the app, the effectiveness of the user interface and user experience design, and the quality of the prototype (if applicable).

Documentation: Written documentation required.
Girls will need to provide a written documentation explaining the concept, features, and design choices of their app. This documentation will serve as a guide for evaluating their work.

Grading: The task can be easily graded concerning certain quality properties.

The task can be graded based on the clarity and creativity of the app concept, the usability and visual appeal of the user interface, the functionality and effectiveness of the app features, and the completeness and coherence of the written documentation.

Task Description (within the maximum word limit):

In this project, you will have the exciting opportunity to design a mobile app for a sustainable fashion community. The app will provide a platform for fashion enthusiasts to connect, share ideas, and promote sustainable practices in the fashion industry.
You will be working individually on this project, allowing you to explore your creativity and problem-solving skills. Throughout several lessons, you will research, plan, design, and prototype your unique mobile app concept.

To foster your holistic thinking, you will need to consider various aspects of sustainable fashion and abstract ideas related to fashion communities. Visualizing your app's design and functionality will be crucial to creating an engaging and user-friendly experience.

You will utilize HCI-design tools such as Figma, Adobe XD, or Sketch to create the user interface and design the user experience of your app. These tools will enable you to bring your creative ideas to life visually. Additionally, you will have the option to implement a basic prototype of your app using beginner-friendly programming tools like Thunkable or App Inventor.

Your app design can take different forms, reflecting your unique

perspective and creativity. While there may be common elements, such as user profiles, discussion forums, and sustainability resources, you have the freedom to explore various features and functionalities

To find out more about the variety of topics, the prompt was repeated several times, resulting in the following tasks: “Design a Virtual Wellness Retreat”, “Plan an Environmentally-Conscious Community Garden App” and “Designing an Interactive Virtual Reality Experience for Exploring Marine Life”. All tasks contained a similar variety of detail and suggestions for the solution.

C. Adding a storyframe with role models

To get the learners “hooked” on to the task, the next prompt requested a background story – but since no additional information was given on the girls’ age (set to 12-19 in the first prompt), the result very much resembled the style of a children’s book and was not really convincing (see Table V):

TABLE V. CREATING A BACKGROUND STORY

Prompt: “Create a background story for the sustainable fashion app you suggested first. The story should raise the girls' interest and contain a maximum of 250 words.
Result: Once upon a time, in a world where fashion reigned supreme, a group of passionate young girls embarked on a mission to change the industry for the better. Their names were Mia, Lily, and Ava. […]

Since role-models are an important catalysator in fostering the girls’ interest in computer science [5], it makes sense to customize the background. Iterative prompting leads to better (but still quite flowery) results, like in the following example (see Table VI):

TABLE VI. ADAPTING THE BACKGROUND STORY TONE AND CONTENT

Prompt: Rewrite the background story using a famous female computer scientist as a role model. Give a few biographical details on the woman and make it sound less pathetic.
Result: In the realm of technology, where innovation thrived, a group of aspiring young girls found themselves captivated by the remarkable story of Dr. Carol Shaw, a pioneering computer scientist who left an indelible mark on the gaming industry. Inspired by her groundbreaking achievements, they set out on a mission to combine their passion for fashion with their love for technology. Dr. Shaw, an exceptional programmer and game designer, blazed trails in a male-dominated field. Her talent and perseverance led her to become one of the first female video game developers, shaping the landscape of interactive entertainment. The girls admired her technical brilliance, creative vision, and the barriers she shattered along her journey.

Depending on the context, the idea of female founders instead of female computer scientists might also help to raise the learners’ motivation (see Table VII):

TABLE VII. ADAPTING THE BACKGROUND STORY TONE AND CONTENT

Prompt: Create a background story for the sustainable fashion app you suggested first. The story should raise the girls' interest and contain a maximum of 150 words. Use a female founder role model to inspire the girls.
Result: In the bustling city of creativity, a young girl named Maya found herself captivated by the vibrant world of fashion. But she couldn't ignore the environmental toll the industry was taking. Determined to make a difference, Maya sought inspiration from a remarkable role model - Stella McCartney. Driven by Stella's fearless pursuit of sustainable fashion, […]

This is a good example how iterative prompting can lead to better results. Since ChatGPT allows reference to previously processed prompts and increasingly more memory is available for this, many iterations can be performed. Moreover, users can store pre-defined instructions. This can be used for prompts describing the framework instead of repeating the details every time. When the task creation is finished, the development of a sample solution or additional hints can be generated. Table VIII shows an example for an Android code snippet with explanations generated from a simple prompt in natural language:

TABLE VIII. CREATING PARTS OF A SAMPLE SOLUTION

Prompt: “Create a code snippet for a function in the sustainable fashion app registering a new user.”
Result: Certainly! Here's a code snippet for a function that registers a new user in a sustainable fashion app:

```
// Assuming you have a User class to represent the user data

public class User {
    private String username;
    private String email;
    private String password;

    // Constructor, getters, and setters
    // ...
}
// RegistrationActivity.java
import android.os.Bundle;
import android.support.v7.app.AppCompatActivity;
import android.view.View;
import android.widget.Button;
import android.widget.EditText;
import android.widget.Toast;

public class RegistrationActivity extends
AppCompatActivity {
    private EditText usernameEditText;
    private EditText emailEditText;
    private EditText passwordEditText;
    private Button registerButton;
        @Override
    protected void onCreate(Bundle
savedInstanceState) {
        super.onCreate(savedInstanceState);
[…]
```

Please note that this code snippet assumes you have an activity_registration.xml layout file that contains the necessary UI elements such as EditText and Button to capture user registration information. Additionally, you would need to implement the logic to save the user data to a database or an API, which is beyond the scope of this code snippet.

Remember to adapt the code to fit your specific application structure and requirements.

Thus, sample solution code generation becomes less time-consuming.

VI. MAKING USE OF GENDER STEREOTYPES “FOR GOOD”?

Generative AI models are often accused of containing certain biases based on the data they have been trained with

[14]. Especially gender bias is a widely discussed topic [15] and teachers using LLMs to generate learning material need to be highly aware of it. On the other hand, the approach taken in this paper actively makes use of the fact that a LLM has information about “typical” interests of boys and girls. But in this case, the original problem stemmed from a bias in the real world, namely that the exercises developed by teachers were too often dealing with topics that were more attractive to boys than to girls. Our approach actively uses the AI models’ gender bias in the short run to fight the leaky STEM-pipeline in the long run, because it is important to get the girl students “hooked” on the technical topics. There’s evidence that girls solving gender-stereotypical mathematical tasks can outperform girls in a control group [16]. Therefore, if this leads to more interesting learners’ tasks and in result to more active participation of girls in computer science lessons, it is a legitimate use of such an AI chatbot.

VII. DISCUSSION AND CONCLUSION

An AI Chatbot, such as ChatGPT, can be a useful tool in creating learners’ tasks for computer science lessons, especially with a focus on content that female learners are interested in. But to successfully create exercises, a framework is needed to ensure that certain properties are considered and that the task descriptions are comparable. This can be done by providing categories of a framework to the AI Chatbot. Prompt engineering can be steered in an iterative way, providing more and more context or using storytelling to create more appealing exercises. The task descriptions can even be enriched by using role-models from computer science or the business world. It is also possible to generate sample solution or hints, depending on the type of task. Evaluation and grading should remain in the hands of the teachers, since these more complex tasks that require group work can – at least at the present time – not be automatically done. But if teachers give written feedback to learners’ artefacts, an AI Bot can be used to fine-tune the mode of address and to help (re-)formulate objective criticism.

Due to the current LLMs’ large scopes of texts and facts, creating interesting and innovative exercises by this approach can save teachers lots of time and effort and cover many stages of the learning-material lifecycle. Student teachers should therefore be familiarized with such possibilities of task creation already during their studies and encouraged to use them. In this respect, the active use of LLMs should be included in the teacher training curricula.

VIII. OUTLOOK

Teachers can use the case study to gather some experience in AI-supported task design and test the results with regard to a rise in motivation especially for female learners. The next steps could then be to use the approach described above to generate female-, as well as male-oriented lesson tasks and test whether girls and boys prefer one over the other using quantitative research approaches. A suitable study design could use an A/B-Testing approach to let students decide which type of task they would prefer to work on. Here, the reasons could also be asked in a questionnaire. Parallel to this, the approach should be tested within a larger third-party funded research project by a group of female computer science professors from the IU International University of Applied Sciences (including the authors), which is currently in the application phase. It would consist of a series of interventions and a ring lecture for female students on IT tasks at a large number of German secondary schools.

As long as the factual and methodological knowledge is successfully transported equally in these exercises (which can also be evaluated quantitatively), nothing speaks against this idea of structured and automated exercise-design.

REFERENCES

[1] J. D. Speer, “Bye Bye Ms. American Sci: Women and the Leaky STEM Pipeline,” Economics of Education Review, Volume 93, 102371, 2023, https://doi.org/10.1016/j.econedurev.2023.102371

[2] C. Hess et al., “How does school lesson design contribute to girls’ perception of their future-readiness for STEM majors? – An empirical study using critical incidents technique”, The 4th International Conference on Teaching and Education Management (ICTEM 2023), Berlin, Germany, June 2023 (in press).

[3] P. Medel and V. Pournaghshband, “Eliminating gender bias in computer science education materials” [Paper presentation], Proceedings of the 2017 ACMSIGCSE Technical Symposium on Computer Science Education,. New York, NY: Association for Computing Machinery, pp. 411–416, March 2017.

[4] A. Steffen et al., “An Exploration of Critical Incidents Impacting Female Students’ Attitude Towards STEM Subjects”, Vol. 6 No. 1: Proceedings of the 6th International Conference on Gender Research, 2023, pp. 215-223, https://doi.org/10.34190/icgr.6.1.994.

[5] L. Happe, B. Buhnova, A. Koziolek, and I. Wagner, “Effective measures to foster girls’ interest in secondary computer science education”, Education and Information Technologies, Vol. 26(3), pp. 2811–2829, 2021, https://doi.org/10.1007/s10639-020-10379-x

[6] CC2020 Task Force, “Computing Curricula 2020: Paradigms for Global Computing Education,” ACM, 2020, https://doi.org/10.1145/3467967.

[7] V. Thurner, A. Bottcher, and A. Kamper, “Identifying Base Competencies as Prerequisites for Software Engineering Education,*”* EDUCON – 2014 IEEE Global Engineering Education Conference, pp. 1069–1076. https://doi.org/10.1109/EDUCON.2014.6826240.

[8] D. Zehetmeier et al., “Self-assessment of freshmen students’ base competencies,” *EDUCON – 2014 IEEE Global Engineering Education Conference*, pp. 429–438. https://doi.org/10.1109/EDUCON.2014.6826130, 2014.

[9] S. Kunz and C. Hess, “How to catch’em all: Designing attractive learning activities for girls in computer science. A systematic framework”, The 4th International Conference on Teaching and Education Management (ICTEM 2023), Berlin, Germany, June 2023 (in press).

[10] T. Trust, J. Whalen, and C. Mouza, “Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education”, Contemporary Issues in Technology and Teacher Education, vol. *23*(1), pp. 1-23, 2023.

[11] T. Teubner, C. M. Flath, C. Weinhardt, W. van der Aalst, and O. Hinz, „Welcome to the Era of ChatGPT et al.,” Business & Information Systems Engineering, vol. 65(2), 2023, pp. 95–101, https://doi.org/10.1007/s12599-023-00795-x.

[12] DAIR.AI. Prompt Engineering Guide: Prompting Techniques https://www.promptingguide.ai/techniques, 2023. [retrieved: 08, 2023]

[13] J. White et al., “A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT,”, 2023, arXiv preprint available at http://arxiv.org/pdf/2302.11382v1. [retrieved: 08, 2023]
[14] C. O’Neil, “Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy”, Penguin Books, 2017.
[15] A. Nadeem, B. Abedin and O. Marjanovic, "Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies", *ACIS 2020 Proceedings*. 27., 2020, https://aisel.aisnet.org/acis2020/27. [retrieved: 08, 2023]
[16] J. M. Hermann and R. Vollmeyer, "Gender stereotypes: implicit threat to performance or boost for motivational aspects in primary school?", Soc Psychol Educ vol. 25, pp. 349–369, 2022, https://doi.org/10.1007/s11218-022-09693-8.

A Foveated Approach to Automated Billboard Detection

Sayali Avinash Chavan   Dermot Kerr   Sonya Coleman   Hussein Khader*
Intelligent Systems Research Centre
University of Ulster
Londonderry, United Kingdom
**The Neuron*, Amman, Jordan
email:{chavan-s, d.kerr, sa.coleman}@ulster.ac.uk, hussein.khader@theneuron.com

***Abstract*—Understanding billboard visibility is vital when considering the value of each billboard to advertisers, hence the growing demand for artificial intelligence based approaches to visibility measurement. Addressing this need, this research paper presents a comprehensive approach to billboard detection using street-view images. We have developed a robust billboard detection system by leveraging state-of-the-art object detection models, such as You Only Look Once (YOLOv8), YOLOv5, Faster-Region-based Convolutional Neural Network (RCNN) and CenterNet resulting in high model accuracy. We have introduced an innovative foveated approach, based on the human visual systems, that applies a Gaussian function to assign weights to billboards to determine which is the most significant billboard based on a combination of confidence and location with respect to the image centre. The approach demonstrates an improvement in overall accuracy of the detection process. In particular YOLOv8 experienced a high accuracy increase from 63.40 to 82.71 percent. This research provides valuable insights and practical solutions for billboard detection in real-time.**

***Index Terms*—Object Detection; Deep Learning; YOLO; Convolutional Neural Network.**

I. INTRODUCTION

Billboards play a significant role in outdoor advertising, aiming to capture the attention of individuals and deliver brand messages effectively. Understanding the visibility of a billboard can help advertisers place value on those which are more desirable considering the likelihood that the advertisement can be clearly observed. Detecting billboards automatically with computer vision can provide valuable insights for this purpose. It allows advertisers and marketers to assess the impact and reach of their advertisement campaigns. It also helps them to evaluate the effectiveness of their strategies and facilitates data driven decision-making tools. Such insights can guide businesses in making informed decisions and optimise their advertising strategies for maximum impact and return on investment. With the ever-increasing presence of billboards in urban areas, the need to detect and analyse billboard visibility has become essential [1].

However, accurately identifying billboards remains a challenging task due to their diverse shapes, sizes, types and the environment within which they reside. Additionally, occlusions, obstructions, and lighting variations further complicate the detection process for the network. Developing a universal billboard detection algorithm that works accurately across different types of billboards requires training on a diverse dataset for robust detection [2]. Hence, in this paper, we present a novel approach to billboard detection using state-of-the-art object detection models, combined with a function that prioritises billboards in desired locations.

This paper makes significant contributions in the field of billboard detection using street-view images with a comprehensive dataset, development of a robust detection algorithm and an approach to prioritising billboards dependent on their location within a scene. Through extensive experimentation and evaluation, we demonstrate the effectiveness of this approach in achieving higher accuracy in comparison to popular models, such as CenterNet [3], Faster-RCNN [4], YOLOv5 [5] and latest version of YOLO [6]- YOLOv8 [7]. The results indicate that our proposed method achieves significant performance using data in the wild, reaching 82.71% correctly detected billboards on unseen data. Our work offers valuable insights and practical solutions that can be utilised in various applications, including urban planning, advertising analysis, and outdoor media management.

This paper is structured as follows: In Section II, we examine the state-of-the-art in billboard detection, identifying key gaps that our study aims to fill with new insights. We outline our chosen model, dataset, and training process in Section III, followed by the implementation details in Section IV which highlights our efforts to enhance each model through hyperparameter tuning, showcasing the refinement of each network architecture. The introduction of the Gaussian weighting algorithm further improves model accuracy. Finally, in Section V, we present the outcomes of our study through an extensive analysis of the results, providing valuable insights for future research in this field, as summarised in Section VI.

II. STATE OF ART

For object detection, feature-based detection methods widely utilise approaches such as Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), Speeded-Up Robust Features (SURF), and Oriented FAST and Rotated BRIEF (ORB) [8]. However, these methods possess inherent limitations. They have low detection rates,

Copyright (c) IARIA, 2023.   **ISBN: 978-1-68558-115-2**

are sensitive to changes in illumination and struggle with complex backgrounds, occlusions, and variations in rotation. Moreover, these approaches are slow and encounter difficulties when dealing with multiple objects in low-quality images. As a result, Convolutional Neural Network (CNN) have emerged as a popular solution, offering high-speed object detection and recognition [9]. CNNs have revolutionised the field by enabling automatic feature learning directly from raw image data. Specifically designed CNN models for object detection can effectively detect patterns in image data and leverage transfer learning for improved performance [10].

Over the past two decades, numerous ground breaking object detection models have been published and extensively studied [11]. These models include SSD (Single Shot Detector) [12], CenterNet [3], RCNN [13], Fast RCNN [14], Faster RCNN [4], YOLO [6], Feature Pyramid Network (FPN) [15], Retina-Net [16], RefineDet [17], Spatial Pyramid Pooling Network (SPPNet) [18], Deformable Part-based Model (DPM) [19], TridentNet [20], Fully Convolutional One-Stage Detector (FCOS) [21], Hybrid Task Cascade (HTC) [22], Deformable DETR [23], and many more. Recent cutting-edge advancement in the field of image segmentation is the Segment Anything Model (SAM) developed by Meta AI Research [24]. SAM is an instance segmentation model trained on an extensive dataset comprising 11 million images and 1.1 billion segmentation masks. Its state-of-the-art performance makes it highly accurate for real-time applications. However, despite these recent developments, the YOLO model remains widely popular in object detection applications due to its real-time high accuracy [6] [11]. YOLO has evolved from its initial version, YOLOv1, to the most recent version, YOLOv8 [5] [7]. YOLOv8 is capable of performing segmentation, classification, detection, tracking, and pose detection, making it a versatile and powerful model in the field.

For the specific application of billboard detection, previous studies have primarily concentrated on detecting advertising billboards using techniques, such as edge detection and planar object detection [25] [26]. However, these researches encounter localisation difficulties when confronted with multiple objects present in the scene. Additionally, research has targeted billboards specifically in soccer fields and sport TV broadcasts, employing methods like the Fast Fourier Transform and Hough transforms [27] [28]. Yet, it is crucial to acknowledge that their ability to accurately detect billboards was reliant on the use of a high-accuracy camera capable of capturing high-resolution pictures.

Further research has been conducted in the field of billboard detection, with a particular emphasis on the text and content displayed within advertisements. One machine learning-based approach aimed to identify illegal advertisements by analysing the extracted content from the advertisements themselves [29]. In another study, the detection of photo manipulation on billboards was accomplished using a novel forensic technique that assessed adherence to the rules of perspective projection [30]. Additionally, a study employed Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to recognise text present on billboards [31]. Compliance-related research explored the use of computer vision techniques to identify advertisements on buildings, as demonstrated by [32]. The billboard detection examples in [29]–[32] primarily focuses on text or content recognition, lacking a comprehensive approach to detect billboards regardless of their content. Therefore, in our research, we aimed to delve into the detection process irrespective of the billboard's content.

Shifting the focus to video frames, the development of the ADNet architecture allowed for the detection of advertising instances, with the utilisation of the Microsoft COCO (Common Objects in Context) dataset for training purposes [33]. A significant drawback of ADNet is its lack of scalability, which raises concerns about its ability to perform effectively in diverse real-world scenarios.

Research by Liu et al. [34] introduced combining attention-based multi-scale features with Faster RCNN for the purpose of billboard detection. They acknowledged billboard detection challenges such as small object sizes, cluttered backgrounds, and low resolutions impose limitations on accuracy improvement. A comparative study regarding urban billboard detection comparing SSD and YOLO models exhibited promising results [35]. However, both research studies in [34] and [35] demonstrated satisfactory performance within the constraints of their respective high-resolution limited datasets. Furthermore, the research conducted by Chavan et al. [2] demonstrated successful outcomes in billboard classification and detection. Overall, the implementation of state of art object detection models shows promise in overcoming real-world application challenges.

III. METHODOLOGY

The methodology section provides a comprehensive outline of the research approach employed in this study, offering a clear roadmap for model selection, data collection, data cleaning and training a custom model using a dataset containing ground truth bounding boxes for billboards. We fine-tune the models using a large number of images, annotations and hyperparameters. To further enhance the accuracy, we apply a Gaussian distribution-based weighting [36] to the centre of the detected objects during testing to deal with real world challenges such as multiple billboards within a scene.

A. Model Selection

After conducting extensive research on the current state-of-the-art models [11], we carefully selected four networks: Faster RCNN, CenterNet, YOLOv5, and YOLOv8.

- Faster RCNN stands out for its exceptional accuracy in object detection. By utilising a region proposal network, it generates potential object locations and then fine tunes the model for improved localisation and recognition, which is essential for applications requiring reliable results [4].
- CenterNet focuses on estimating the centre points of objects. It can accurately determine the position of the object, enabling precise centre detection and localisation of billboards. Its centre point estimation approach allows

it to handle and distinguish multiple objects efficiently, making it suitable for scenarios where multiple billboards may be present in the scene [3].
- YOLO is renowned for its impressive speed in object detection. It can process images in real-time. Moreover, YOLOv8, the latest version released in 2023, is designed to be computationally efficient enabling faster processing, which can be advantageous in scenarios where quick detection is crucial [6].

B. Image Dataset

We obtained longitudinal and latitudinal coordinates from an Out of Home (OOH) industry partner [37] for billboard locations in the United Kingdom, which were used as reference points for retrieving corresponding street-view images programmatically. A meticulous cleaning process was completed to ensure the dataset's quality and relevance. This involved removing duplicate images and filtering any irrelevant or low-quality images (as depicted in Figure 1) that did not accurately represent the billboard locations. After cleaning the dataset, we have 3,437 images (examples of which are shown in Figure 2), which were subsequently divided into three subsets: a training set comprising 2,500 images, a validation set consisting of 700 images, and a test set containing 238 images. The split of approximately 73% training, 20% validation, and 7% test optimises a substantial training dataset for model training, a validation set for precise hyperparameter tuning, and a carefully selected test set for an unbiased final evaluation [38].

The dataset consists of images obtained from various locations across the UK, with different types of billboards, including digital and static billboards, street furniture, spectacular billboards, and illuminated billboards. The billboards exhibit a wide range of designs, layouts, and content, showcasing promotional messages for products, services and events. Furthermore, the images also capture contextual elements such as trees, buildings, roads, pedestrians, and vehicles illustrates the difficulty in detection. By following this comprehensive data extraction process - created the custom billboard dataset, which involved obtaining billboard coordinates, extracting street view images surrounding each billboard to obtain an image dataset, cleaning and dividing the data and performing manual annotation. The manual annotation step entailed reviewing each image and accurately labeling the billboards by placing bounding boxes around them.

C. Training Process

The training process involves optimising the models parameters and weights through an iterative process to ensure accurate detection and localisation of billboards. The models (Faster RCNN, CenterNet, YOLOv5, and YOLOv8) were fine-tuned using the annotated dataset, allowing them to learn and adapt to the specific characteristics of billboard images in the UK regions. This is the general flow of the system:

- Input images - fed into the object detection model.
- Loss function - calculated based on the model predictions and ground truth.
- Model parameters - updated using gradient descent optimisation, minimising the loss.
- The training process iterates over the dataset multiple times depending on the number of steps/epochs to improve the model's performance.

This comprehensive training process enabled the models to achieve superior performance and effectiveness in detecting and recognising billboards in various real-life scenarios.

We trained Faster RCNN, CenterNet, and two versions of YOLO; these were compared based on their accuracy. We analysed their training and testing accuracy, and robustness to varying billboard images with single and multiple billboards in one image. In Section V, we conducted an evaluation of the models, focusing on their ability to generalise to unseen urban scenes (test set of 238 images).

Fig. 1. Example of a low-quality or irrelevant image (blurry image) filtered during dataset cleaning process.

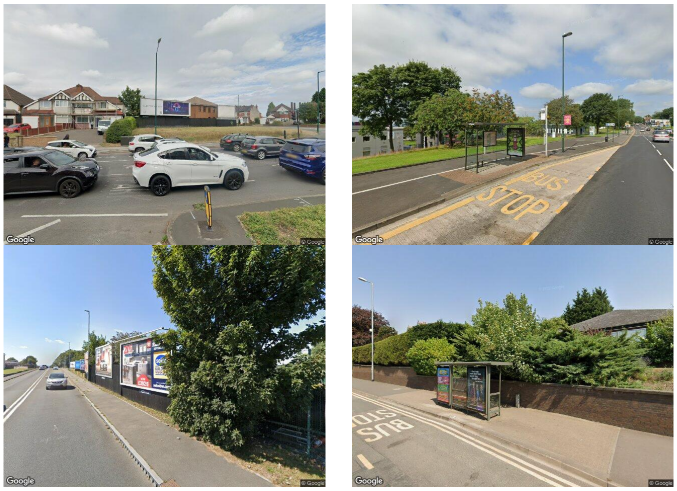


Fig. 2. Training sample from image dataset for UK region showcasing billboards in real-life environments with surrounding vehicles and background scenery.

IV. MODEL IMPLEMENTATION

The following section elaborates on the practical implementation of each network, elucidating the architectural details employed in this research.

A. Faster RCNN

Faster RCNN is a popular framework for object detection in computer vision. It consists of a convolutional neural network (CNN) backbone, such as ResNet or VGGNet, for feature extraction. As shown in Figure 3, Faster RCNN combines a Region Proposal Network (RPN) with a CNN-based object detection network to efficiently detect objects in an image.

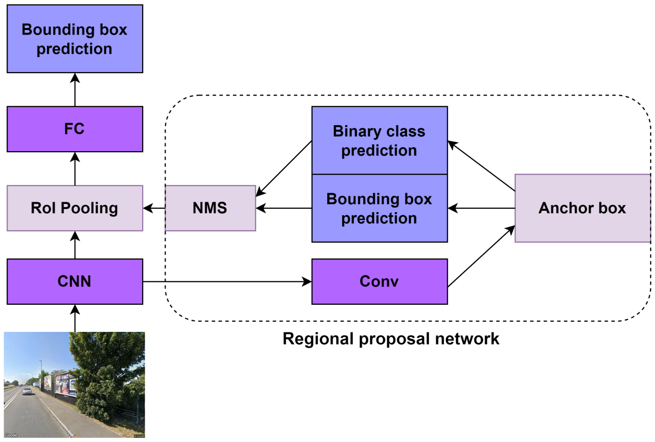


Fig. 3. Faster RCNN Architecture.

The RPN generates a set of region proposals. It generates candidate object proposals, which are then refined using a Region of Interest (RoI) pooling layer and passed through fully connected layers for classification and bounding box regression. RPN then shares full-image convolutional features with a detection network, thus enabling nearly cost-free region proposals. Faster RCNN merges the RPN and Fast RCNN into a single network by sharing their convolutional features 'attention' mechanism. The RPN component tells the unified network where to look. The key steps involved in an RPN for object detection framework cover the following aspects in sequence:

- Input: Image of size $W \times H$
- Convolutional feature map: $F = CNN(Image)$
- Anchor generation: Generate a set of fixed-size anchor boxes at different scales and aspect ratios.
- Anchor classification: For each anchor, predict the probability of it containing an object (foreground) or not (background).
- Anchor regression: For each foreground anchor, refine the coordinates of the bounding box to better fit the object.

The Object Detection Network processes RoIs and performs object classification and bounding box regression. Convolutional feature maps can be used for generating region proposals. This is constructed by adding a few more convolutional layers so that the model performs both localisation and regression tasks at the same time, thus a FCN (Fully Convolutional Neural Network) can be trained end to end specifically for the task of generating detection proposals. The main components of an object detection framework encompass the following aspects in sequence:

- Input: Convolutional feature map F from the RPN and region proposals.
- RoI pooling: Crop and resize the features within each region proposal to a fixed size.
- Fully connected layers: Pass the RoI-pooled features through a series of fully connected layers.
- Classification: Predict the class probabilities for each region proposal.
- Bounding box regression: Refine the coordinates of the bounding boxes.

In the framework, the loss function is used to train the model and optimise the network parameters. The loss function consists of two components: the classification loss (L_{cls}) and the regression loss (L_{reg}). The overall loss for the RPN is denoted as L_{rpn} and defined as the sum of the classification loss and the regression loss, with the regression loss multiplied by a balancing parameter (keeping λ = 10) [4]:

$$L_{\text{rpn}} = L_{\text{cls}}(p, t) + \lambda \cdot L_{\text{reg}}(t^b, t_i^b) \tag{1}$$

The classification loss (L_{cls}) is computed for each anchor and averaged over the number of classes (N_{cls}):

$$L_{\text{cls}} = \frac{1}{N_{\text{cls}}} \sum_c L_{\text{cls}}(p_c, t_c) + \lambda \cdot L_{\text{reg}}(t_b, t_i^b) \tag{2}$$

Here, t_i is the ground truth label for the anchor, t_i^b is the ground truth bounding box regression targets, $L_{\text{cls}}(p_c, t_c)$ is the binary logistic loss between the predicted probability p_c and the ground truth label t_c, and $L_{\text{reg}}(t_b, t_i^b)$ is the smooth $L1$ loss for bounding box regression. The term λ is a balancing parameter that determines the relative importance of the classification and regression components in the overall loss. By minimising this loss function during training, the Faster RCNN model learns to generate accurate region proposals and accurately classify and localise objects within those regions. Faster RCNN achieves impressive accuracy but is relatively slow compared to some newer models [39].

B. *CenterNet*

CenterNet is an innovative and efficient single-shot object detection model that deviates from traditional methods by focusing on predicting the center points of objects instead of explicitly predicting bounding boxes. By regressing the center point and object size, CenterNet achieves high accuracy and efficiency. Unlike other object detectors that rely on bounding box estimation, CenterNet adopts a keypoint estimation approach. It detects each object as a triplet of keypoints, specifically the object's center point and the two corners of its bounding box.

The CenterNet architecture comprises a backbone network, intermediate heatmaps for keypoint estimation, and offset regression maps for bounding box prediction, as shown in Figure 4. It incorporates two custom modules: cascade corner pooling and center pooling. Cascade corner pooling enriches information gathered from the top-left and bottom-right corners of objects, while center pooling provides more notable information from the central regions. This combination enhances the model's ability to capture both corner and center characteristics. The output includes corner heatmaps and center heatmaps, indicating the likelihood of object corners and

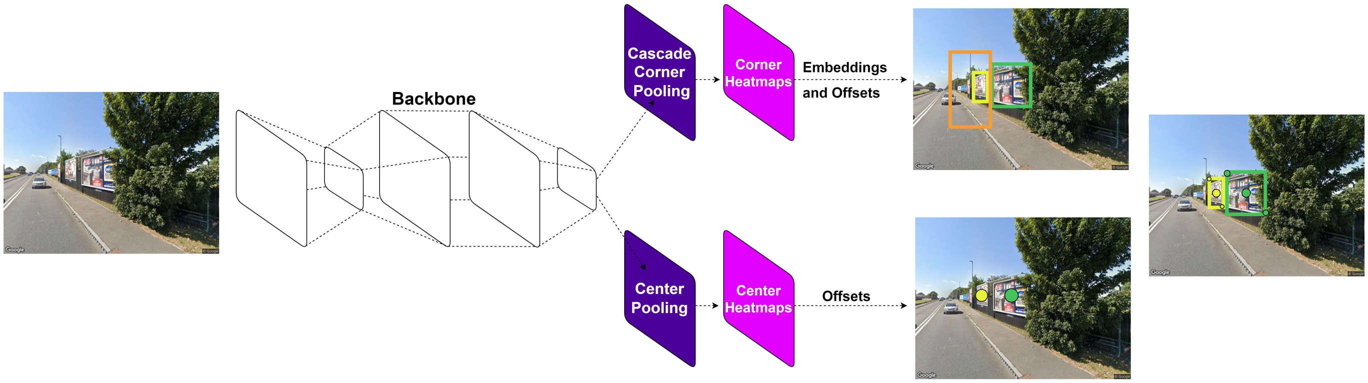


Fig. 4. CenterNet structure.

centers at different spatial locations. Additionally, embedding offsets refine localisation by providing displacement information. The final step involves localising the object center using the corner heatmaps, center heatmaps, and embedding offsets.

During inference, CenterNet utilises the predicted corner keypoints to generate proposals, and then checks if there is a center keypoint of the same class falling within the central region of each proposal. If a centre keypoint of the same class is found, it confirms the proposal as a valid object.

The loss function in CenterNet comprises the heatmap loss, which measures the dissimilarity between predicted and ground truth heatmaps using binary cross-entropy. Regression loss refines bounding box coordinates using the smooth $L1$ loss. The total loss combines the heatmap loss and regression loss, with a balancing parameter to control their relative importance. Mathematically, the loss function in CenterNet can be summarised as [3]:

$$L_{\text{total}} = L_{\text{heatmap}} + \lambda \cdot L_{\text{reg}} \tag{3}$$

where L_{heatmap} represents the binary cross-entropy loss between predicted and ground truth heatmaps, and L_{reg} denotes the smooth L1 loss for refining bounding box coordinates. λ is the balancing parameter [40].

CenterNet achieves accurate object detection while maintaining efficiency in terms of computational resources and processing time [3].

C. YOLO

YOLO is a real-time object detection framework that aims to achieve high detection accuracy with fast inference speed. It divides the input image into a grid and predicts bounding boxes and class probabilities directly. The YOLO architecture [41] employs a CNN backbone (as shown in Figure 5), followed by a series of convolutional layers. These layers simultaneously predict the bounding box coordinates, objectness score, and class probabilities at multiple grid scales, allowing for accurate and efficient object detection. YOLO can be described as a deformable parts model that utilises a sliding window approach. The network performs multiple tasks concurrently, including feature extraction, bounding box prediction, non-maximal suppression, and contextual reasoning. This unified approach contributes to its efficiency and real-time object detection capabilities [6].

In contrast to traditional methods like RCNN, YOLO uses grid cells to propose potential bounding boxes and scores for objects. However, YOLO applies spatial constraints to these grid cell proposals, which helps mitigate multiple detections and leads to far fewer bounding box proposals. One of the notable advantages of YOLO is that it is a unified model for object detection. It can be directly trained on full images, unlike classifier-based approaches. YOLO models, such as YOLOv3 and YOLOv4, have shown good performance on various object detection benchmarks. We have chosen to use both YOLOv5 and YOLOv8 for billboard detection [42].

YOLOv5 has gained widespread recognition and popularity in the field of object detection. It stands out for its exceptional performance, achieving good results in terms of accuracy and speed using various benchmark datasets. This makes it a highly promising choice for object detection tasks. One significant advantage of YOLOv5 is its efficiency and lightweight nature. It utilises a CSPDarknet53 backbone, which enhances feature extraction capabilities and contributes to improved detection accuracy. The YOLOv5 network architecture consists of 20 convolutional layers, followed by an average-pooling layer and a fully connected layer. By incorporating both convolutional and connected layers, the ImageNet pre-trained YOLOv5 model has shown improved performance in object detection tasks [43].

YOLOv8 is the latest version of the YOLO model. Although it shares the same architecture as its predecessors, it introduces several improvements. These improvements include a new neural network architecture that combines the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN). The FPN in YOLOv8 gradually reduces the spatial resolution of the input image while increasing the number of feature channels. This results in the creation of feature maps that can effectively detect objects at different scales and resolutions. On the other hand, the PAN architecture aggregates features from multiple levels of the network using skip connections. By doing so, the network can capture features at various scales and resolutions

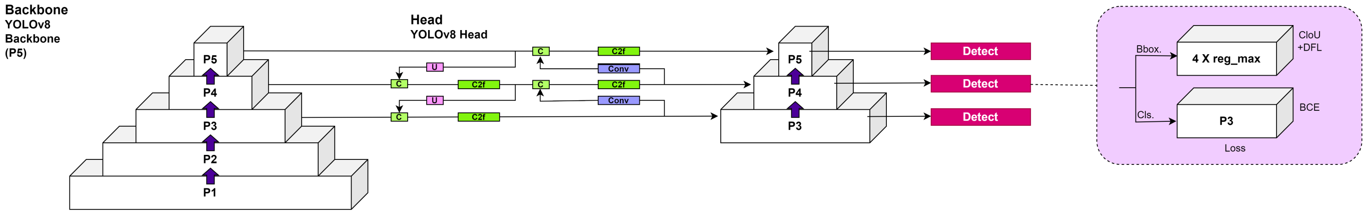


Fig. 5. YOLOv8 Architecture.

more comprehensively, which is vital for accurate detection of objects with different sizes and shapes [41].

The loss function in YOLO guides the optimisation process by aligning predicted bounding box coordinates, improving object localisation accuracy, and encouraging accurate object presence estimation and class prediction. The balancing parameters λ_{cls} , λ_{loc}, λ_{obj} provides the flexibility to fine-tune the significance of different components in the loss function. This allows for optimising the model's performance during training by adjusting the importance given to each specific loss term. The loss function helps the model learn and improve its object detection capabilities, leading to more accurate predictions during inference than predictions made without utilising the loss function [7].

$$L_{\text{yolo}} = \lambda_{\text{loc}} \cdot L_{\text{loc}} + \lambda_{\text{obj}} \cdot L_{\text{obj}} + \lambda_{\text{cls}} \cdot L_{\text{cls}} \tag{4}$$

where L_{loc} is the localisation loss, L_{obj} is the objectness loss, and L_{cls} is the classification loss.

D. Hyper-Parameters

It is important to note that the selection of hyperparameters is a crucial aspect of training deep learning models. The chosen values are influenced by factors such as the dataset characteristics, model architecture, available computational resources, and desired trade-off between speed and accuracy. Faster RCNN, CenterNet, YOLOv5 and YOLOv8 default settings served as a starting point for our experiments, and we further fine-tuned the hyperparameters to optimise the model's performance on our specific dataset using the values shown in Table I. The batch size is the number of samples processed at once during training. Larger batch sizes can lead to faster convergence, but it can also require more memory. The step size decides the total number of training iterations. Whereas the learning rate determines how quickly model learns and updates its parameters during optimisation. It affects both the speed at which the model converges and the accuracy achieved during the training process. An optimizer is a mathematical algorithm used in machine learning to adjust the parameters of a model in order to minimise an error function or maximise a desired output. It plays a vital role in improving the performance and efficiency of models. The following optimizers have been used:

- Momentum optimizer: The momentum optimizer incorporates a momentum term to accelerate convergence by considering the accumulated velocity from previous updates. The equations for the momentum optimizer are as follows:

$$v_t = \gamma \cdot v_{t-1} + \alpha \cdot \nabla J(W_t) \tag{5}$$

$$W_{t+1} = W_t - v_t \tag{6}$$

Here, v_t represents the velocity at time step t, γ is the momentum coefficient, α is the learning rate, $\nabla J(W_t)$ is the gradient of the loss function with respect to the weights W_t, and W_{t+1} is the updated weights [44].

TABLE I
HYPERPARAMETER

Model	**Faster RCNN**	**CenterNet**	**YoloV5**	**YoloV8**
Image Size	640	512	416	800
Batch Size	1	1	32	16
Optimizer	Momentum	Adam	SGD	SGD
Step Size / Epoch	0-2000	0-2000	0-100	0-50
Warm up Learning Rate	0.001	0.001	0.001	0.001
Step Size / Epoch	2000-25,000	2000-25,000	0-100	0-50
Final Learning Rate	0.004	0.004	0.001	0.001

- Adam optimizer: The Adam optimizer combines concepts from Momentum and root mean square propagation (RMSprop), adapting the learning rate for each parameter based on past gradients. The equations for the Adam optimizer are as follows:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \nabla J(W_t) \tag{7}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (\nabla J(W_t))^2 \tag{8}$$

$$W_{t+1} = W_t - \frac{\alpha}{\sqrt{v_t} + \epsilon} \cdot m_t \tag{9}$$

Here, m_t represents the first moment estimate (mean) at time step t, β_1 is the decay rate for the first moment estimate, m_{t-1} is the first moment estimate at time step $t-1$, v_t represents the second moment estimate (variance) at time step t, β_2 is the decay rate for the second moment estimate, $\nabla J(W_t)$ is the gradient of the loss function with respect to the weights W_t, W_t is the weights at time step t, W_{t+1} is the updated weights at time step $t+1$, α is the learning rate, and ϵ is a small value for numerical stability [44].

- Stochastic Gradient Descent (SGD) optimizer: SGD is a basic optimization algorithm that updates the parameters based on the gradient of the loss function computed using a small subset (batch) of training data. The equation for the SGD optimizer is:

$$W_{t+1} = W_t - \alpha \cdot \nabla J(W_t) \tag{10}$$

Here, W_t represents the weights at time step t, W_{t+1} is the updated weights at time step $t + 1$, α is the learning rate, and $\nabla J(W_t)$ is the gradient of the loss function with respect to the weights W_t [44].

E. Gaussian Weighting Algorithm

To further enhance the accuracy of billboard detection, a Gaussian weighting was applied to the detected object's centre. During testing, the bounding boxes generated by each network were filtered using a Gaussian function centred at the detected object's centroid. This additional step aimed to refine the object localisation and improve the model's accuracy [36].

To implement the Gaussian Distribution algorithm, the pre-trained billboard model is loaded, and a confidence threshold of 50% is set for detection. For each test image, the algorithm calculates the coordinates of the image center and the center coordinates of the detected object. The distance x between the object center and the image center is computed as the euclidean distance. The object's weight, denoted as σ, is set to one-fourth of the image width $\frac{w}{4}$. The Gaussian weight is obtained by applying the Gaussian function with the distance, where the mean μ is set to 0 for the center. The equation for the Gaussian Distribution is given by [45] [46]:

$$Gaussian\ Distribution = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{11}$$

The Gaussian weight is then multiplied by the confidence score of the detected object, resulting in the combined score. If the combined score surpasses the threshold of 50%, the algorithm identifies the billboard as the center one of interest and draws a green bounding box around it. Conversely, if the combined score is below the threshold, a red bounding box is drawn to indicate that the billboard is not located at the center. This can be observed in Figures 6 and 7. The Gaussian Distribution algorithm plays a crucial role in refining the results of billboard detection.

V. RESULT ANALYSIS

We conducted experiments using a dataset of 238 unseen images to analyse the generalisation capabilities of our model. The Mean Average Precision (mAP) is widely used as a metric to evaluate the object detection models [47], as it compares the ground-truth bounding box to the detected box and provides a score that reflects the accuracy of the model's detections. A higher score indicates more accurate detections [48].

TABLE II
MODEL COMPARISON RESULTS

Model	**Faster RCNN**	**CenterNet**	**YoloV5**	**YoloV8**
Step Size / Epoch	25,000	25,000	100	50
Training Accuracy (mAP.50)	59%	75.70%	64.50%	**87.40%**
Testing Accuracy (mAP.50)	54.30%	57.20%	52.60%	63.40%
With Gaussian - Testing Accuracy	65.43%	55.55%	80.24%	**82.71%**

Table II provides a comprehensive comparison of the performance of four different models: Faster RCNN, CenterNet, YOLOv5, and YOLOv8. The models were trained with different step sizes/epochs, and their performance was evaluated based on training accuracy and testing accuracy using mean Average Precision at 0.50 Intersection over Union (mAP@0.50). Furthermore, the testing accuracy was assessed through the application of Gaussian distribution weighting.

The training accuracy results for the models are as follows: Faster RCNN achieved a training accuracy of 59%, CenterNet achieved 75.70%, YOLOv5 achieved 64.50%, and YOLOv8 exhibited an impressive training accuracy of 87.40%. These results indicate the model's ability to learn and detect billboard features during the training process.

Using the testing dataset, the results are as follows: Faster RCNN achieved a testing accuracy of 54.30%, CenterNet achieved 57.20%, YOLOv5 achieved 52.60%, and YOLOv8 achieved 63.40%. These scores demonstrate the model's performance in accurately detecting billboards in unseen images. These results demonstrate that the YOLOv8 model obtains good performance compared with other models with respect to both training and testing.

In order to further enhance the accuracy of detecting billboards, we employed a Gaussian weighting technique specifically on the centre of the detected objects. This modification aimed to improve overall network performance by reducing incorrect classifications in a number of situations, namely those with multiple detections within a scene when wish to select the object that is closest to the centre of the image with a high confidence interval for the billboard class.

After applying the Gaussian weighting, we observed the following results in terms of testing accuracy for each model: Faster RCNN improved from 54.30% to 65.43%, CenterNet experienced a slight decrease from 57.20% to 55.55%, it performed relatively poorly during testing, indicating that although it had previously exhibited good generalisation during training by accurately detecting the object centres, it struggled to accurately identify the centres of billboard objects in the testing phase. YOLOv5 achieved a significant increase from 52.60% to 80.24%. Notably, YOLOv8 exhibited the most remarkable improvement in testing accuracy after the application of the Gaussian weighting technique. With an increase from 63.40% to an impressive 82.71%, it emerged as the top-performing model in terms of accurately detecting billboards for this dataset. These findings highlight the effectiveness of the Gaussian weighting technique in significantly enhancing the performance of the models.

Figures 6 and 7 illustrate examples of the application of the Gaussian distribution algorithm to detect the centre billboard in the case of multiple objects present in the scene and justifies the YOLOv8 testing accuracy increase. In Figure 6, the green bounding box highlights the selection of the center billboard, with a detection score of 80.60%. Significantly, even though there was an alternative billboard exhibiting a higher detection score of 94.66% (depicted by the adjacent red bounding box), its Gaussian weight of 15.06 was relatively lower, resulting in a

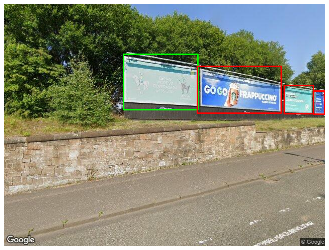


Fig. 6. Center Billboard Detection Results Using the Gaussian Distribution Algorithm.

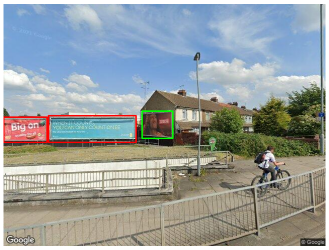


Fig. 7. Center Billboard Selection based on Combined Weights with Gaussian Algorithm.

combined weight that fell below the designated 50% threshold. Similarly, two additional billboards in the scene surpassed the 50% detection threshold with scores of 73.76% and 68.31%. However, due to their comparatively lower combined weights, they were not deemed as the center billboards of interest (indicated by the red bounding boxes).

Analogous findings were observed in Figure 7, reflecting the similar results depicted in Figure 6. The centre billboard was selected based on its confidence score of 84.89% and a Gaussian weight of 99.65%, outweighing another billboard that had a detection score of 92.45% but a lower Gaussian weight of 22.36% resulting in a combined weight of 20.67%, making it unsuitable for the center position. Furthermore, there was an additional billboard with a detection score of 70.00% and a Gaussian weight of 66.85%. However, the combination of this detection score and Gaussian weight fell below the threshold criteria of 50%, thus it was also not chosen as the centre billboard.

These results provide strong evidence supporting our hypothesis that applying a Gaussian weighting approach to objects based on their proximity to the centre of the image can greatly improve the accuracy of billboard detection. In summary, the comparison of the model's performance reveals that YOLOv8 consistently outperforms the other models in terms of both training and testing accuracy. However, it is important to consider other factors such as computational requirements, model complexity, and specific use case requirements when selecting the most suitable model for a given scenario.

VI. CONCLUSION

This research paper addresses the need for accurate billboard detection in advertising analytics. We have made notable contributions in several areas. Firstly, we developed a robust billboard detection system using advanced models like YOLOv8, YOLOv5, Faster-RCNN, and CenterNet. Challenges involved in this endeavor included finding optimal hyperparameters, mitigating over-fitting, and efficiently managing computational resources during the training process, all of which we adeptly addressed and resolved during the development of these networks. Hence, these models demonstrate high accuracy in detecting billboards in real-world scenarios.

Furthermore, we introduced an innovative approach by applying a Gaussian weighting technique to determine the most central billboards. This significantly improved the overall accuracy of the detection process, particularly in the case of YOLOv8, which achieved an impressive accuracy of 82.71% after the application of the Gaussian weighting. The combination of accurate detection models and the novel Gaussian weighting approach proves to be a promising direction for improving billboard detection in various domains. This advancement holds significant potential for applications such as urban planning, advertisement analysis, and traffic monitoring.

Future work could explore further refinements to optimise the proposed approach and extend it to real-time billboard detection systems, encompassing the task of verifying the relevance of displayed information. This would involve ensuring that billboards continuously present accurate and up-to-date content, addressing scenarios where some billboards may no longer convey valid information. Furthermore, exploring the billboard visibility based on environmental conditions, as well as the unique perspectives offered by different viewing angles, resulting in more effective outdoor advertising.

ACKNOWLEDGMENT

We wish to express our gratitude towards Digital Natives for providing the geo-coordinates for the dataset and funding the International PhD Scholarship at Ulster University.

REFERENCES

[1] R. T. Wilson and J. Casper, "The role of location and visual saliency in capturing attention to outdoor advertising: How location attributes increase the likelihood for a driver to notice a billboard ad," *Journal of Advertising Research*, vol. 56, no. 3, pp. 259–273, 2016.

[2] S. Chavan, D. Kerr, S. Coleman, and H. Khader, "Billboard detection in the wild," pp. 57–64 Irish Machine Vision and Image Processing Conference 2021, Sep. 2021. [Online]. Available: https://iprcs.github.io/IMVIP.html [retrieved: Aug, 2023]

[3] K. Duan et al., "CenterNet: Keypoint triplets for object detection," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 6568–6577, 2019.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." [Online]. Available: http://image-net.org/challenges/LSVRC/2015/results [retrieved: Aug, 2023]

[5] G. Jocher et al., "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7347926 [retrieved: Aug, 2023]

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 779–788, 2016.

[7] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond," pp. 1–31, 2023. [Online]. Available: http://arxiv.org/abs/2304.00501 [retrieved: Aug, 2023]

[8] K. Li and L. Cao, "A review of object detection techniques," *Proceedings - 2020 5th International Conference on Electromechanical Control Technology and Transportation, ICECTT 2020*, pp. 385–390, 2020.

[9] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020. [Online]. Available: https://doi.org/10.1007/s13748-019-00203-0 [retrieved: Aug, 2023]

[10] L. Alzubaidi et al., *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*. Springer International Publishing, 2021, vol. 8, no. 1. [Online]. Available: https://doi.org/10.1186/s40537-021-00444-8 [retrieved: Aug, 2023]

[11] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[12] W. Liu et al., "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer International Publishing, 2016, pp. 21–37.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jun 2014, pp. 580–587. [Online]. Available: https://arxiv.org/abs/1311.2524 [retrieved: Aug, 2023]

[14] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015, pp. 1440–1448.

[15] T. Lin et al., "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[17] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 5168–5177. [Online]. Available: https://arxiv.org/abs/1611.06612v3 [retrieved: Aug, 2023]

[18] K. He et al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," *Computer Vision – ECCV 2014*, pp. 346–361, 2014.

[19] R. Girshick, F. Iandola, T. Darrell, and J. Malik, "Deformable Part Models are Convolutional Neural Networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 437–446, sep 2014. [Online]. Available: https://arxiv.org/abs/1409.5403v2 [retrieved: Aug, 2023]

[20] Y. Li, Y. Chen, N. Wang, and Z. X. Zhang, "Scale-Aware Trident Networks for Object Detection," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 6053–6062, jan 2019. [Online]. Available: https://arxiv.org/abs/1901.01892v2 [retrieved: Aug, 2023]

[21] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 9626–9635, apr 2019. [Online]. Available: https://arxiv.org/abs/1904.01355v5 [retrieved: Aug, 2023]

[22] K. Chen et al., "Hybrid Task Cascade for Instance Segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 4969–4978, jan 2019. [Online]. Available: https://arxiv.org/abs/1901.07518v2 [retrieved: Aug, 2023]

[23] X. Zhu et al., "Deformable DETR: Deformable Transformers for End-to-End Object Detection," oct 2020. [Online]. Available: https://arxiv.org/abs/2010.04159v4 [retrieved: Aug, 2023]

[24] A. Kirillov et al., "Segment Anything," 2023. [Online]. Available: http://arxiv.org/abs/2304.02643 [retrieved: Aug, 2023]

[25] R. F. Rahmat, Dennis, O. S. Sitompul, S. Purnamawati, and R. Budiarto, "Advertisement billboard detection and geotagging system with inductive transfer learning in deep convolutional neural network," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 5, pp. 2659–2666, 2019.

[26] P. Liang et al., "Planar object tracking in the wild: A benchmark," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 651–658, 2018.

[27] A. Watve and S. Sural, "Detection of on-field advertisement billboards from soccer telecasts," *IET International Conference on Visual Information Engineering*, pp. 12–17, 2006.

[28] G. Cai, L. Chen, and J. Li, "Billboard advertising detection in sport TV," *Proceedings - 7th International Symposium on Signal Processing and Its Applications, ISSPA 2003*, vol. 1, pp. 537–540, 2003.

[29] H. Liu, L. Wang, W. Zhang, and W. Wang, "An illegal billboard advertisement detection framework based on machine learning," *ACM International Conference Proceeding Series*, pp. 159–164, 2019.

[30] V. Conotter and G. Boato, "Detecting photo manipulation on signs and billboards," *Proceedings - International Conference on Image Processing, ICIP*, no. 1, pp. 1741–1744, 2010.

[31] S. Anbukkarasi, V. E. Sathishkumar, C. R. Dhivyaa, and J. Cho, "Enhanced Feature Model based Hybrid Neural Network for Text Detection on Signboard, Billboard and News tickers," *IEEE Access*, vol. 11, no. April, pp. 41 524–41 534, 2023.

[32] K. Bochkarev and E. Smirnov, "Detecting advertising on building façades with computer vision," *Procedia Computer Science*, vol. 156, pp. 338–346, 2019. [Online]. Available: https://doi.org/10.1016/j.procs.2019.08.210 [retrieved: Aug, 2023]

[33] M. Hossari, S. Dev, M. Nicholson, K. McCabe, A. Nautiyal, C. Conran, J. Tang, W. Xu, and F. Pitié, "ADNet: A deep network for detecting adverts," *CEUR Workshop Proceedings*, vol. 2259, pp. 45–53, 2018.

[34] G. Liu, C. Wang, and Y. Hu, "RPN with the attention-based multi-scale method and the adaptive non-maximum suppression for billboard detection," *2018 IEEE 4th International Conference on Computer and Communications, ICCC 2018*, pp. 1541–1545, 2018.

[35] Á. Morera, Á. Sánchez, A. B. Moreno, Á. D. Sappa, and J. F. Vélez, "Ssd vs. Yolo for detection of outdoor urban advertising panels under multiple variabilities," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–23, 2020.

[36] L. Hou, K. Lu, X. Yang, Y. Li, and J. Xue, "G-Rep: Gaussian Representation for Arbitrary-Oriented Object Detection," *Remote Sensing*, vol. 15, no. 3, pp. 1–21, 2023.

[37] "The neuron intelligent connections: Programmatic exchange for dooh advertising." [Online]. Available: https://theneuron.com/ [retrieved: Sept, 2023]

[38] I. O. Muraina, "Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts," *7th International Mardin Artuklu Scientific Researches Conference*, no. February, pp. 496–504, 2022.

[39] Y. Chen et al., "Domain Adaptive Faster R-CNN for Object Detection in the Wild," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3339–3348, 2018.

[40] Y. Guo and X. Lu, "ST-CenterNet: Small Target Detection Algorithm with Adaptive Data Enhancement," *Entropy (Basel, Switzerland)*, vol. 25, no. 3, 2023. [Online]. Available: https://doi.org/10.3390/e25030509 [retrieved: Aug, 2023]

[41] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-Time Flying Object Detection with YOLOv8," 2023. [Online]. Available: http://arxiv.org/abs/2305.09972 [retrieved: Aug, 2023]

[42] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[43] J. Doherty, B. Gardiner, E. Kerr, N. Siddique, and S. Manvi, "Comparative study of activation functions and their impact on the yolov5 object detection model," *International conference on pattern recognition and artificial intelligence, ICPRAI*, pp. 40–52, Jun. 2022.

[44] S. Ruder, "An overview of gradient descent optimization algorithms," pp. 1–14, 2016. [Online]. Available: http://arxiv.org/abs/1609.04747 [retrieved: Aug, 2023]

[45] T. Beckers, "An Introduction to Gaussian Process Models," *arXiv preprint arXiv:2102.05497*, 2021.

[46] X. Zhang, "Gaussian distribution," *Encyclopedia of Machine Learning*, pp. 425–428, 2010.

[47] R. Padilla et al., "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics (Switzerland)*, vol. 10, no. 3, pp. 1–28, 2021.

[48] J. Revaud, J. Almazán, R. S. Rezende, and C. R. d. Souza, "Learning with average precision: Training image retrieval with a listwise loss," pp. 5107–5116, 2019.

Deep Learning for Billboard Classification

Sayali Avinash Chavan Dermot Kerr Sonya Coleman Hussein Khader*
Intelligent Systems Research Centre
University of Ulster
Londonderry, United Kingdom
**The Neuron*, Amman, Jordan
email:{chavan-s, d.kerr, sa.coleman}@ulster.ac.uk, hussein.khader@theneuron.com

***Abstract*—Advertising is essential to increase product awareness and foster a positive outlook, which in turn helps sales. To promote the brand and its products, billboard advertisements are widely used. This paper presents a novel approach for classifying billboards. The proposed method utilises Convolutional Neural Network (CNN) architectures to extract features from the images to enable classification. The model is trained on a dataset of billboards collected from various locations and achieves results that demonstrate high classification accuracy. The system is trained and evaluated using the CIFAR10 dataset, which includes 10 classes of objects and an additional 11th class - 'billboard', is included. The experiment utilises five different CNN architectures: Basic CNN, ResNet, Visual Geometry Group (VGG), MobileNet, and DenseNet. The performance and evaluation of each architecture are presented in detail, and extensive experiments and comparisons are conducted to determine the most effective model for classifying billboards. The results indicate that a CNN and its architectural designs are a promising solution for automating the classification of billboards in the wild.**

***Index Terms*—Classification; CNN-Architecture; Billboard; Image-Processing; CIFAR10**

I. INTRODUCTION

Out of Home (OOH) advertising is one of the most powerful channels, helping connect brands with a large and diverse audience when they are outside of their homes. This type of advertising offers several advantages including the ability to reach specific audiences using visually striking images that can increase sales and maintain brand visibility in the market. It is considered less intrusive than other forms of advertising as consumers choose whether or not to engage with it. Such advertisements can be found on roadside billboards, transit stations, street furniture, retail outlets, health and beauty, point of care and office buildings just to name a few [1]. The visibility of OOH advertisements can vary. For example, a billboard's size, shape, location and viewing angle affect visibility and in some cases, obstructions such as trees, traffic signals and vehicles can partially block an onlooker's view of a billboard. As a result, detecting billboards in images can be challenging due to the dynamic nature of their content [2] [3].

The purpose of this study is to create a classifier for billboard images using a CNN model that can accurately categorise billboards amidst other types of images in natural environments. The goal is to train and test 5 CNN architectures using billboard images merged with a subset of the CIFAR10 dataset [4], evaluate their effectiveness and determine the most efficient deep learning model for real-time billboard classification [5] [6]. The complexity lies in discerning billboards, which often blend seamlessly into the urban landscape, demanding the model's heightened perceptual acuity. Moreover, the efficacy of our proposed method in classifying billboards becomes evident when juxtaposed with classification against 10 distinct categories present in the CIFAR-10 dataset. Billboards, despite being distinct from these classes, can share similar visual characteristics with urban structures, causing intricacies in differentiation [2].

Computer vision techniques are driving a revolution across diverse applications, encompassing object and pattern recognition, precise image segmentation, robust facial detection, and even pioneering advancements in robotics and autonomous driving [7] [8] [9]. Object classification is a fundamental problem in computer vision and has been extensively studied in the machine learning and computer vision communities [10]. One of the earliest and most successful approaches to object classification is the use of hand-crafted features, such as edge detection and texture analysis [11]. These methods rely on expert knowledge and domain-specific heuristics to extract informative features from images and use them to train a classifier [12]. Some of the well-known classifiers, logistic regression, naive bayes, k-nearest neighbors, decision tree etc., were utilised for classification purposes [13]. These classifiers are widely recognised and frequently employed in various machine learning tasks due to their simplicity, interpretability, and reasonable performance. However, these methods are limited in their ability to capture complex and abstract patterns and require a significant amount of manual effort to design and implement which led to advancements in deep learning [14] [15].

Recent advances in machine learning and deep learning have led to the development of more powerful and flexible object classification algorithms [14] [16]. These methods use CNNs to automatically learn rich and complex features from raw images, without the need for manual feature engineering. CNNs have achieved state-of-the-art performance on many benchmarks [17] and are widely used in practical applications. Current studies have centred on addressing these challenges and advancing the state of the art in object classification. However, there are still many challenges and open questions in the field of object classification [15]. These include the need for large amounts of labelled training data, the sensitivity of deep learning models to small perturbations in the input and the difficulty of explaining and interpreting the decisions made by these models [18]. This includes the development of CNN architectures, such as the VGG and ResNet architectures [16] [19], the use of transfer learning and other techniques to improve the efficiency and robustness of these models and the exploration of more interpretable and explainable approaches to object classification [4] [14] [20].

Hence, in this research, we investigate well-known architectures, such as ResNet, VGG, MobileNet, and DenseNet to build a billboard classifier. The study provides a detailed analysis of each architecture's performance, facilitating a comprehensive understanding of the achieved results. To illustrate our research trajectory, this paper adheres to the designated structure in the forthcoming sections. In Section 2, we delve into the Methodology, addressing dataset creation and billboard classifier design. Following that, Section 3, sheds light on Network Architectures, laying out the specific layers, training details, hyperparameter optimisation and provides comprehensive information about the five distinct network architectures employed in the research. Subsequently, Section 4 is dedicated to the Results and Analysis of the applied methods, with a focus on evaluating the performance using F1 scores on datasets CIFAR11 and CIFAR2. Lastly, the final section encapsulates the conclusion.

II. METHODOLOGY

This section discusses the dataset used and the methodology employed for the image classification experiment using CNN architectures.

A. *Dataset*

The CIFAR10 dataset [21] [22] is often used as a benchmark for evaluating the performance of different image classification algorithms. The 10 in CIFAR10 dataset represents 10 classes with 6000 images per class, each image is of size $32*32$ pixels. In total this dataset is comprised of 60,000 colour images. The 10 classes are as follows in alphabetic order: Airplane, Automobile, Bird, Cat, Deer, Dog, Frog, Horse, Ship, and Truck.

For our experiment, we have integrated a new class denoted 'Billboard' in the dataset. This class of images was collected by the OOH industry partner. The Billboard images shown in Figure 1 were cropped and resized to $32*32$ pixels to be consistent with the CIFAR10 standard. Several observations are made for the billboard class such as images including different types of billboards, different locations, different backgrounds, obstructions caused by trees, vehicles or people, and other background clutter, etc. A total of 778 images were deemed suitable for use in the experiment. However, this created class imbalance while merging this class into CIFAR10 which has 6000 images per class.

Imbalanced datasets can negatively affect the performance of machine learning models by biasing the training towards the majority class and reducing the ability of the model to accurately classify the minority classes. Previous studies have shown that balancing the dataset can improve the performance of CNNs in image classification tasks [23]. Hence, we selected approximately the first 700 images of each of the 10 classes in the CIFAR10 dataset. By selecting an equal number of images for each class, we aimed to improve the accuracy and robustness of the CNN models used in this experiment [24]. Billboard images were pre-processed to match the CIFAR10 standards creating two separate datasets, denoted as CIFAR11 and CIFAR2 for nomenclature purposes, as briefly stated below:

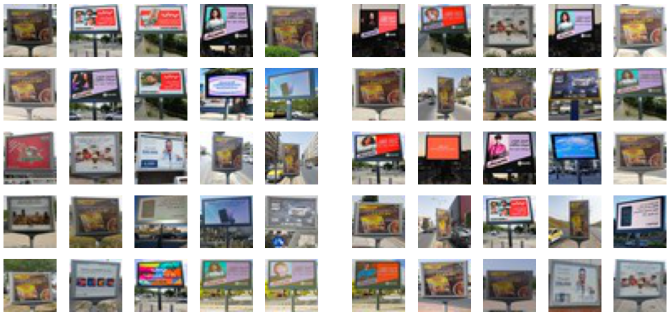


Fig. 1. Billboard class images.

- CIFAR11: Billboard class combined with all 10 classes of CIFAR10 – 6188 images used for the training set, 1546 images for validation set and 220 images for the testing set. Total 7,954 images belonging to 11 classes.
- CIFAR2: Based on the outcomes in Table III, the 'Billboard' class was merged with only 'Ship' class from CIFAR10 - 1230 images used for the training set, 306 images for the validation set and 40 images for the testing set. Total 1,576 images belonging to 2 classes.

B. *Billboard Classifier*

To conduct the research, five CNN architectures, namely a basic CNN, ResNet, VGG, MobileNet and DenseNet, have been utilised [16] [19] [25] [26]. Our approach involved building upon pre-trained models. The transfer learning was applied for ResNet, VGG, MobileNet, and DenseNet architectures [27].

The basic building blocks used for this experiment are shown in Figure 2. The standard methodology for image classification using CNNs involves several steps. Firstly, a dataset comprising images and their respective labels is split into training and validation sets. These images are then pre-processed by performing data normalization to make them compatible with the CNN architecture. The preprocessed data are fed into the classifier for training, which is one of the five CNN architectures used in the experiment. The classifier is trained on the validation and training datasets to extract features from the images using convolutional layers, activation functions, and pooling layers. The output is then passed through fully connected layers to classify the images into their respective classes. Backpropagation is applied to adjust the

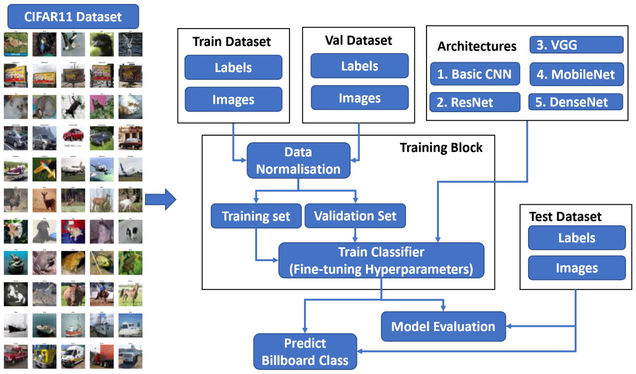


Fig. 2. Billboard classifier building blocks.

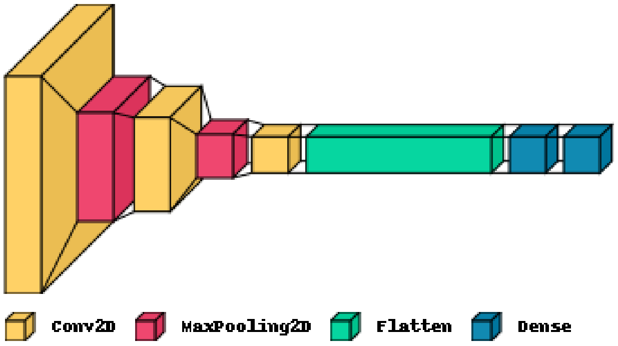


Fig. 3. Basic CNN Visualisation in Layered View.

weights and biases to train the network. Finally, the training process is repeated and fine-tuned using hyper-parameters such as a number of epochs, learning rate, batch-size, etc., to improve the accuracy. The model's performance is evaluated on an unseen test set [12].

III. NETWORK ARCHITECTURES

Each of the five CNN architectures has its own strengths and weaknesses, and the choice of architecture depends on the specific requirements of the image classification task at hand [7]. For example, ResNet [16] and VGG [19] are good choices for large, complex datasets, while MobileNet [25] is a good choice for deployment on mobile devices with limited computational resources. DenseNet [26] is a good choice for image classification tasks that require a high degree of accuracy and fast convergence.

A. *Basic CNN*

A basic CNN is a simple feed-forward network that is designed to provide a foundation for understanding more complex models. The basic structure of a CNN for training on the dataset involves several key components: the input layer, convolutional layers, pooling layers, fully connected layers, and an output layer with an activation function. These components work together to learn to recognize and classify the objects in the images [28]. The convolutional layer is used for extracting features from the input image. The basic formula for a single convolutional layer can be represented as:

$$f(x) = W * x + b \tag{1}$$

where $f(x)$ is the output of the convolutional layer, W is the set of weights filters, x is the input image, and b is a bias term.

B. *ResNet*

ResNet is a deep residual learning framework that was introduced in 2015 by He et al. [16]. It is designed to address the vanishing gradient problem in deep networks by adding shortcut connections that bypass one or more layers. This allows for easier training of much deeper networks to learn residual mappings, which can help to improve performance for image classification tasks. ResNet is known for its high accuracy and ability to handle large datasets. For our experiment, we have employed the ResNet 50 model, chosen for its remarkable depth, skip connections, and advanced architecture, which collectively enhance its ability to capture intricate features within the images. The output of the residual block can be represented as:

$$Y = F(x, W_i) + x \tag{2}$$

where x is the input to the residual block, $F(x, W_i)$ is the residual mapping function that is learned by the block, W_i are the learnable parameters of the block, and Y is the output of the block. The ”+ x” term in the equation represents the identity mapping, which allows the output of the block to include the original input x.

C. *VGG*

Visual Geometry Group (VGG) is a CNN architecture, that was proposed in 2014 by Simonyan et al. [19]. It is known for its use of small convolutional filters and deep network architecture, which helps to capture fine-grained details in images. VGG differs from a basic CNN by having more layers and using smaller filter sizes of $3*3$ pixels. This allows VGG to learn more complex features from the input image but also makes it more computationally expensive to train and run. To mitigate this issue, VGG typically uses max pooling layers after every two or three convolutional layers to reduce the spatial dimensions of the output. By downsampling the output, VGG is able to decrease the computational load of the network while still retaining important features.

$$f(x) = W_2 \cdot (W_1 \cdot (W_1 \cdot x + b_1) + b_2) + b_3 \tag{3}$$

where x is the input image, W_1 and W_2 are weight matrices, b_1, b_2, and b_3 are bias vectors represents the max pooling operation.

D. *MobileNet*

MobileNet is a lightweight CNN architecture that was introduced in 2017 by Howard et al. [25]. It was designed for use in mobile and embedded devices and is known for its efficiency and accuracy. MobileNet is a computationally

TABLE I
ARCHITECTURE COMPARISON OF THE LAYER STRUCTURE[1]

Basic CNN	**ResNet**	**VGG**	**MobileNet**	**DenseNet**
conv2d (Conv2D)	resnet50 (Functional)	input_2 (InputLayer)	mobilenetv2_1.00_224 (Functional	densenet201 (Functional)
max_pooling2d (MaxPooling2D)	flatten (Flatten)	block1_conv1 (Conv2D)	global_max_pooling2d_1 (GlowbalMaxPooling2D)	flatten_1 (Flatten)
conv2d_1 (Conv2D)		block1_conv2 (Conv2D)	dense (Dense)	batch_normalization_2 (BatchNormalization)
max_pooling2d_1 (MaxPooling2D)		block1_pool (MaxPooling2D)		dense_3 (Dense)
conv2d_2 (Conv2D)		block1_conv2 (Conv2D)		dropout_2 (Dropout)
flatten (Flatten)		block1_pool (MaxPooling2D)		batch_normalization_3 (BatchNormalization)
dense (Dense)		block2_conv1 (Conv2D)		dense_4 (Dense)
dense_1 (Dense)		block2_conv2 (Conv2D)		dropout_3 (Dropout)
flatten_1 (Flatten)		block2_pool (MaxPooling2D)		dense_5 (Dense)
dense_2 (Dense)		block3_conv1 (Conv2D)		
dense_3 (Dense)		block3_conv2 (Conv2D)		
		block3_conv3 (Conv2D)		
		block3_conv4 (Conv2D)		
		block3_pool (MaxPooling2D)		
		*		
		flatten_1 (Flatten)		
		dense_1 (Dense)		

efficient version of the CNN that uses depthwise separable convolutions to reduce the computational complexity while maintaining accuracy. Its lightweight architecture helps to reduce computation time and energy consumption, while still providing good performance for image classification tasks. MobileNet uses depthwise separable convolutions, and can be represented as:

$$f(x) = (W_1 * x)W_2 + b \tag{4}$$

where $f(x)$: Output of the MobileNet layer, W_1: Depthwise convolutional kernel, W_2: Pointwise convolutional kernel, x: Input to the MobileNet layer (e.g., an image), b: Bias term.

E. DenseNet

DenseNet is a convolutional neural network that was introduced in 2016 by Huang et al. [26]. It is known for its dense connectivity pattern, where each layer is connected to all previous layers, which helps to reduce the number of parameters and mitigate overfitting. The dense blocks allow for a more efficient flow of information and have been shown to improve accuracy and convergence speed for image classification tasks. The output of a single dense block can be represented as:

$$f(x) = [f_1(x), f_2(x), ..., f_k(x)] \tag{5}$$

where $f(x)$: Output of the DenseNet layer, $f_i(x)$: Output of the i-th dense block, x: Input to the DenseNet layer, k: Number of dense blocks in the DenseNet layer.

F. Training Details: Architecture Enhancement

Figure 3 presents the basic layer outline. In Table I, the layers and their implementations employed are explained for the following CNN architectures: Basic CNN, ResNet, VGG, MobileNet, and DenseNet. These architectures are formed by incorporating additional layers and implementations, which act as essential building blocks. They enable efficient feature extraction and effective classification tasks for the given dataset. In the context of Table I, the symbol '$*$' serves as an indicator, denoting the repetition of block 3, which occurs twice. A brief explanation of the purpose and functionality of the layers is provided below [29]:

- Conv2D: this layer performs a convolution operation on the input data to extract relevant features.
- MaxPooling2D: this layer applies max pooling to reduce the spatial dimensions of the input, preserving the most important features.
- Flatten: this layer reshapes the input data into a 1-dimensional vector, preparing it for the fully connected layers.
- Dense: these layers are fully connected layers that perform a linear transformation on the input data, followed by an activation function, to generate class predictions.
- Batch Normalization: this layer normalises the input data, helping with training stability and improving the learning process.
- Dropout: this layer randomly sets a fraction of input units to 0 during training, which helps prevent overfitting.
- GlobalMaxPooling2D: this layer applies max pooling across the entire feature map, reducing the spatial dimensions to a single value for each feature map.

Hyperparameter Optimization: Table II illustrates the hyperparameters employed for training each network after utilising the default parameters. A meticulous selection and fine-tuning process were undertaken to optimise the performance of each network. The hyperparameters were carefully tuned to maximise the performance of each model. Table II presents a concise summary of the hyperparameters utilised for each model and a brief explanation for each parameter is given below:

- Epochs: The number of epochs determines the number of times the training process is repeated. After each epoch's the model's output is compared to the ground truth (actual values), and the loss function calculates the difference between them then it adjusts the weight. This newly created weight is then given to system for next epoch for training and this process goes on until the best possible accuracy is achieved without overfitting. This value is different for each model used.
- Early Stopping: The number of epochs used with the CIFAR-11 dataset differs from CIFAR-2 due to the difference in the number of images. CIFAR-11 has 7954

TABLE II
HYPER-PARAMETER TUNING

Model Name	**Basic CNN**	**ResNet**	**VGG**	**MobileNet**	**DenseNet**
Optimizer	Adam	SGD	Adam	Adam	Adam
Epochs (CIFAR11)	20	10	3	15	5
Epochs (CIFAR2)	3	3	3	5	3
Activation Function	ReLU	SoftMax	SoftMax	SoftMax	SoftMax
Batch Size	16	32	16	32	32

images, while CIFAR-2 has only 1230 images. To prevent overfitting, the dataset with fewer images (CIFAR-2) required early stopping with the number of epochs given in Table II for each network that proved to be effective in attaining optimal performance. By leveraging the early stoppage method [30], overfitting was sucessfully overcome for CIFAR-2.

- Batch Size: this was adjusted according to the availability of computational units, aiming to efficiently utilise the computational resources to benefit from parallel processing and improve training efficiency.
- Activation Function: the Softmax activation function is commonly employed in multi-class classification tasks as it converts the model's output into probability distributions across different classes used for ResNet,VGG, MobileNet and DenseNet. The Basic CNN utilised the ReLU activation function, which is a widely used activation function known for its ability to mitigate the vanishing gradient problem and introduce non-linearity to the model. Allowing the model to capture complex patterns and features essential for classification tasks.
- Optimizer: The optimizer used in the training process varied across the models. Basic CNN, VGG, MobileNet, and DenseNet were optimised using the Adam optimizer, which is an adaptive learning rate optimisation algorithm known for its efficiency in handling sparse gradients and noisy data. It computes adaptive learning rates for each parameter by taking into account the exponential decay rates of past gradients and their squared gradients. This helps the model converge faster and achieve better performance. Conversely, the ResNet utilised Stochastic Gradient Descent (SGD) as the optimizer. SGD is a widely used optimisation algorithm that iteratively updates the model's parameters based on the gradients computed on randomly selected mini batches of data [29].

In addition to the hyperparameters listed in Table II, specific measures were taken to address overfitting in the ResNet model. Initially, the model was trained using a learning rate of 0.001, and further fine-tuning was performed by employing SGD with a momentum value of 0.9, potentially improving its performance and convergence. Through careful selection and fine-tuning all networks hyperparameters, the models were trained and optimised to achieve the best possible performance on the CIFAR-11 and CIFAR-2 datasets, considering the differences in the number of images and potential overfitting issues.

IV. RESULT ANALYSIS

The following section presents evaluation results for billboard classification. Tables III and IV present the results of evaluating 5 different deep learning architectures (CNN, ResNet, VGG, MobileNet, and DenseNet) on different classes of images using the CIFAR11 and CIFAR2 datasets. The performance metrics used are precision, recall, and F1-score [31]. The F1-score is a commonly used performance metric for evaluating multi-class classifiers. It is the harmonic mean of precision and recall. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. Recall (or Sensitivity) is the ratio of correctly predicted positive observations to all actual positive observations in the dataset:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

In terms of performance for the 'Billboard' class in Table III, different network architectures showed varying levels of accuracy. Among the architectures evaluated, the basic CNN, DenseNet, and VGG exhibited relatively high F1-scores of 0.97, 0.92, and 0.88, respectively. These scores suggest that these architectures achieved a high level of precision and recall in classifying images belonging to the 'Billboard' class, resulting in a balanced overall performance. On the other hand, the ResNet architecture attained an F1-score of 0.86, indicating a slightly lower but good level of accuracy for this class. However, the MobileNet architecture exhibited the lowest F1-score of 0.10 for the 'Billboard' class, implying that it struggled to accurately classify images within this specific class. The comparatively diminished performance of MobileNet might stem from its lightweight architecture optimised for efficiency, which could compromise its ability to capture intricate dataset features, unlike more complex counterparts such as ResNet, DenseNet, and VGG, which have shown consistently good results across all classes.

The 'Billboard' class consistently performs well across architectures, the performance of other classes varies. Some classes, such as 'Airplane' and 'Truck' tend to achieve relatively high scores, while others, like 'Cat' and 'Deer' show lower scores. The 'Ship' class in the CIFAR-11 dataset demonstrates relatively consistent performance across the network architectures. The F1-scores for this class range from 0.48 to 0.61 suggesting a balance between precision and recall, indicating that the models achieve a reasonable trade-off between correctly identifying 'Ship' images and minimising mis-classifications. Therefore, this specific 'Ship' class with 'Billboard' has been chosen for subsequent CIFAR2 class analysis.

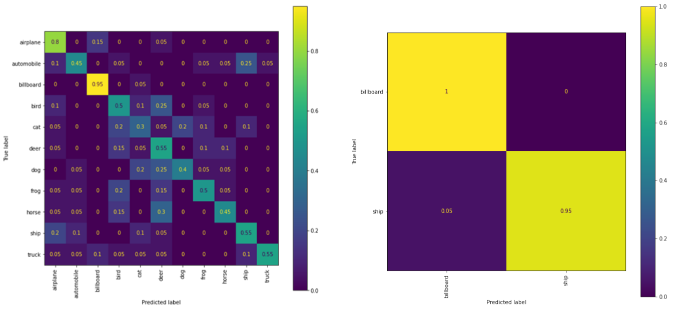
Using CIFAR11, the testing accuracy results for the basic CNN, ResNet, VGG, MobileNet, and DenseNet are 51.36%, 54.55%, 49.55%, 30.45%, and 54.09% respectively. It is evident that ResNet and DenseNet performed well compared to other models. These results highlight the effectiveness of the CIFAR11 model in accurately predicting the target class based on the testing data set. The highest testing accuracy achieved

TABLE III
COMPARATIVE MODEL SUMMARY OF 5 NETWORK ARCHITECTURES OF CIFAR-11

Class Name	**CNN**			**ResNet**			**VGG**			**MobileNet**			**DenseNet**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Airplane	0.41	0.35	0.38	0.55	0.80	0.65	0.57	0.65	0.6	0.60	0.45	0.51	0.58	0.75	0.65
Automobile	0.67	0.40	0.50	0.60	0.45	0.51	0.40	0.70	0.51	0.78	0.35	0.48	0.60	0.45	0.51
Billboard	**1.00**	**0.95**	**0.97**	**0.79**	**0.95**	**0.86**	**0.86**	**0.90**	**0.88**	**1.00**	**0.05**	**0.10**	**1.00**	**0.85**	**0.92**
Bird	0.37	0.55	0.44	0.38	0.50	0.43	0.29	0.35	0.32	0.27	0.60	0.37	0.41	0.45	0.43
Cat	0.56	0.45	0.50	0.35	0.30	0.32	0.28	0.25	0.26	0.16	0.55	0.25	0.40	0.40	0.40
Deer	0.31	0.25	0.28	0.32	0.55	0.41	0.43	0.50	0.47	0.08	0.15	0.11	0.44	0.60	0.51
Dog	0.38	0.30	0.33	0.67	0.40	0.50	0.50	0.35	0.41	0.50	0.35	0.41	0.63	0.60	0.62
Frog	0.50	0.35	0.41	0.59	0.50	0.54	0.44	0.35	0.39	0.43	0.30	0.35	0.45	0.25	0.32
Horse	0.44	0.55	0.49	0.64	0.45	0.53	0.44	0.40	0.42	0.67	0.10	0.17	0.56	0.50	0.53
Ship	0.54	0.70	0.61	0.55	0.55	0.55	0.60	0.45	0.51	0.50	0.25	0.33	0.48	0.50	0.49
Truck	0.55	0.80	0.65	0.92	0.55	0.69	0.85	0.55	0.67	0.57	0.20	0.30	0.50	0.60	0.55

TABLE IV
COMPARATIVE RESULT OF 5 NETWORK ARCHITECTURES OF CIFAR-2

Class Name	**CNN**			**ResNet**			**VGG**			**MobileNet**			**DenseNet**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Billboard	0.94	0.85	**0.89**	0.51	1.00	**0.68**	0.86	0.95	**0.90**	0.83	1.00	**0.91**	0.95	1.00	**0.98**
Ship	0.86	0.95	0.90	1.00	0.05	0.10	0.94	0.85	0.89	1.00	0.80	0.89	1.00	0.95	0.97



True label \ Predicted label	airplane	automobile	billboard	bird	cat	deer	dog	frog	horse	ship	truck
airplane	0.8	0	0.15	0	0	0.05	0	0	0	0	0
automobile	0.1	0.45	0	0.05	0	0	0	0.05	0.05	0.25	0.05
billboard	0	0	0.95	0	0.05	0	0	0	0	0	0
bird	0.1	0	0	0.5	0.1	0.25	0	0.05	0	0	0
cat	0.05	0	0	0.2	0.3	0.05	0.2	0.1	0	0.1	0
deer	0.05	0	0	0.15	0.05	0.55	0	0.1	0.1	0	0
dog	0	0.05	0	0	0.2	0.25	0.4	0.05	0.05	0	0
frog	0.05	0.05	0	0.2	0	0.15	0	0.5	0.05	0	0
horse	0.05	0.05	0	0.15	0	0.3	0	0	0.45	0	0
ship	0.2	0.1	0	0	0.1	0.05	0	0	0	0.55	0
truck	0.05	0.05	0.1	0.05	0.05	0.05	0	0	0	0.1	0.55

True label \ Predicted label	billboard	ship
billboard	1	0
ship	0.05	0.95

(a) CIFAR11 - ResNet.

(b) CIFAR2 - DenseNet.

Fig. 4. Confusion Matrix.

by ResNet among all the models makes it a prime candidate for further analysis using the confusion matrix shown in Figure 4(a), which provides a visual representation of the correct and incorrect predictions for each class.

Considering CIFAR2, observing Table IV, we can see that the highest F1-score across all architectures and classes is achieved by the DenseNet with a value of 0.98 for billboard class. For the class ’Ship’, as well the highest F1-score is achieved by DenseNet with a value of 0.97. It can be concluded that the DenseNet architecture outperforms other architectures in terms of the F1-Score for both the ’Billboard’ and ’Ship’ classes which is also reflected in Figure 4(b) using a confusion matrix for detailed model performance of each class. The precision and recall values for the DenseNet architecture are close to 1.0 which indicates good performance.

V. CONCLUSION

This paper presents deep learning based approaches for billboard classification. Based on the results presented, we can determine that ResNet (CIFAR11) and DenseNet (CIFAR2) are strong candidates compared with the other 3 CNN, particularly for the dataset for automating the classification of billboards and provide a promising solution for this application domain. However, it is important to note that the choice of architecture depends on the specific requirements of the image classification task at hand, and other factors such as computational complexity and deployment environment may also influence the final choice of architecture. Overall, the field of object classification continues to evolve and advance, with many exciting developments and opportunities for further progress.

ACKNOWLEDGMENT

We would like to express our gratitude to Digital Natives for their generosity in providing the dataset and funding the International PhD Scholarship at Ulster University.

REFERENCES

[1] T. Evans, J. Molly, S. O. Eva, and A. Miles, “The Importance of Billboard Advertising,” *International Digital Organization for Scientific Research*, vol. 5, no. 1, pp. 59–65, 2020.

[2] S. Chavan, D. Kerr, S. Coleman, and H. Khader, “Billboard detection in the wild,” pp. 57–64, Sep. 2021, irish Machine Vision and Image Processing Conference 2021, IMVIP2021 ; Conference date: 01-09-2021 Through 03-09-2021. [Online]. Available: https://iprcs.github.io/IMVIP.html [retrieved: Aug, 2023]

[3] C. R. Taylor, G. R. Franke, and H. K. Bang, “Use and effectiveness of billboards perspectives from selective-perception theory and retail-gravity models,” *Journal of Advertising*, vol. 35, pp. 21–34, 2006.

[4] T. Ho-Phuoc, “Cifar10 to compare visual recognition performance between deep neural networks and humans,” *arXiv preprint arXiv:1811.07270*, 2018.

[5] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, “Do CIFAR-10 Classifiers Generalize to CIFAR-10?” pp. 1–25, 2018. [Online]. Available: http://arxiv.org/abs/1806.00451 [retrieved: Aug, 2023]

[6] D. Park, D. Papailiopoulos, and K. Lee, “Active Learning is a Strong Baseline for Data Subset Selection,” *Conference on Neural Information Processing Systems (NeurIPS 2022)*, no. NeurIPS, pp. 1–9, 2022.

[7] J. Chai, H. Zeng, A. Li, and E. W. Ngai, “Deep learning in computer vision: A critical review of emerging techniques and application scenarios,” *Machine Learning with Applications*, vol. 6, p. 100134, 2021.

[8] N. Sünderhauf et al., “The limits and potentials of deep learning for robotics,” *International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 405–420, 2018.

[9] Q. Rao and J. Frtunikj, “Deep learning for self-driving cars: Chances and challenges: Extended Abstract,” *Proceedings - International Conference on Software Engineering*, pp. 35–38, 2018.

[10] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] D. G. Lowe, “Object recognition from local scale-invariant features,” *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157, 1999.

[12] I. Iqbal, G. A. Odesanmi, J. Wang, and L. Liu, “Comparative Investigation of Learning Algorithms for Image Classification with Small Dataset,” *Applied Artificial Intelligence*, vol. 35, no. 10, pp. 697–716, 2021. [Online]. Available: https://doi.org/10.1080/08839514.2021.1922841 [retrieved: Aug, 2023]

[13] S. Khan et al., “Comparison of multiclass classification techniques using dry bean dataset,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 6–20, 2023.

[14] H. Jiang et al., “A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation,” *Computers in Biology and Medicine*, vol. 157, p. 106726, 2023. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2023.106726 [retrieved: Aug, 2023]

[15] F. Sultana, A. Sufian, and P. Dutta, “Advancements in image classification using convolutional neural network,” *Proceedings - 2018 4th IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2018*, pp. 122–129, 2018.

[16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, 2016.

[17] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11263-015-0816-y [retrieved: Aug, 2023]

[18] A. Shrikumar, P. Greenside, and A. Kundaje, “Learning important features through propagating activation differences,” *34th International Conference on Machine Learning, ICML 2017*, vol. 7, pp. 4844–4866, 2017.

[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–14, 2015.

[20] J. Fan, J. H. Lee, and Y. Lee, “Application of transfer learning for image classification on dataset with not mutually exclusive classes,” in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2021, pp. 1–4.

[21] A. Krizhevsky, “Cifar-10 and cifar-100 datasets,” 2009. [Online]. Available: https://www.cs.toronto.edu/ kriz/cifar.html [retrieved: Aug, 2023]

[22] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” 2009.

[23] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *Journal of Big Data*, vol. 6, no. 1, 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0192-5 [retrieved: Aug, 2023]

[24] V. S. Spelmen and R. Porkodi, “A review on handling imbalanced data,” in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 2018, pp. 1–11.

[25] G. A. Howard et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017. [Online]. Available: http://arxiv.org/abs/1704.04861 [retrieved: Aug, 2023]

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, 2017.

[27] M. Iman, H. R. Arabnia, and K. Rasheed, “A Review of Deep Transfer Learning and Recent Advancements,” *Technologies*, vol. 11, no. 2, pp. 1–14, 2023.

[28] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” *Evolutionary Intelligence*, vol. 15, no. 1, pp. 1–22, 2022. [Online]. Available: https://doi.org/10.1007/s12065-020-00540-3 [retrieved: Aug, 2023]

[29] L. Alzubaidi et al., “Review of deep learning: concepts, cnn architectures, challenges, applications, future directions,” *Journal of Big Data 2021 8:1*, vol. 8, pp. 1–74, 3 2021. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8 [retrieved: Aug, 2023]

[30] Y. Wei, F. Yang, M. J. Wainwright, and S. Member, “Early stopping for kernel boosting algorithms: A general analysis with localized complexities,” *IEEE Transactions on Information Theory*, vol. 65, p. 6685, 2019.

[31] N. W. S. Wardhani, M. Y. Rochayani, A. Iriany, A. D. Sulistyono, and P. Lestantyo, “Cross-validation Metrics for Evaluating Classification Performance on Imbalanced Data,” *2019 International Conference on Computer, Control, Informatics and its Applications: Emerging Trends in Big Data and Artificial Intelligence, IC3INA 2019*, pp. 14–18, 2019.

Smart Contracts for Privacy-Preserving Identity Management: Ethics, Regulatory and Technical Challenges

Carmela Occhipinti
R&D Department
CyberEthics Lab.
Rome, Italy
email:
c.occhipinti@cyberethicslab.com

Tetiana Vasylieva
R&D Department
CyberEthics Lab.
Rome, Italy
email:
t.vasylieva@cyberethicslab.com

Luigi Briguglio
R&D Department
CyberEthics Lab.
Rome, Italy
email:
l.briguglio@cyberethicslab.com

Alessio Bianchini
R&D Department
CyberEthics Lab.
Rome, Italy
email: a.bianchini@cyberethicslab.com

Sayonara Crestani
R&D Department
CyberEthics Lab.
Rome, Italy
email: s.crestani@cyberethicslab.com

***Abstract*— Access to online data and service is a fundamental human right recognized by the United Nations. The current digital era, characterized by a continuously evolving network infrastructure which allows broadband and ubiquitous communication, can ensure this fundamental right to a broader number of citizens. However, if not properly managed, this fundamental right may be threatened by cyber threats and related identity frauds. For this reason, Identity Management (IdM) systems lay the foundation to enable this fundamental right. In the context of the IMPULSE project, the research team is adopting an "ethics-by-conception" approach to embed ethics, regulatory, and technical perspectives into the development process of the IdM system. This paper describes the ethics and regulatory framework which identifies principles and regulations to be applied during the development of an IdM system and, based on this framework, the research team identifies potential concerns and measures. This approach allows to design, implement and validate an IdM by integrating blockchain technology and smart contracts mechanism. Moreover, this IdM proposes an iconic representation to simplify the comprehension of policies in the informed consent and, therefore, it empowers users to take better decisions on access and management of their own personal data.**

Keywords — Privacy-Preserving Technology; Identity Management; Ethics and Regulatory Framework; Smart Contract; Blockchain; Regulation; GDPR; eIDAS.

I. INTRODUCTION

Internet access is recognised as a human fundamental right by the United Nations [1] and the implementation of broadband connection networks, including wireless and mobile technologies, is enabling this fundamental right for a continuously growing global population. Certainly, Internet access and online activities have become integral to human life for millions of citizens.

Reliable Identity Management (IdM) and verification are crucial for various online services, ranging from blog and social media login to online banking and public administration services. Therefore, even though online IdM enables ubiquitous access to data and services to a broader number of users, however, online identity fraud and cyber threats pose significant risks, affecting millions of users and making their identities vulnerable to breaches.

Privacy concerns also arise due to the exchange of personal information and for this reason, when accessing online services, users have to deal with the provision of grant to manage their own personal data (i.e., informed consent), in compliance with the General Data Protection Regulation (GDPR) [2]. Therefore, in theory, for each service, users have to manage consents during their lifecycle. In practice, user often grant consent to personal data management and forget which data they are making available and to whom.

In response to these risks, threats and concerns, the European Commission, as well as academics, corporations, and public opinion are actively focusing on resolving the issues of IdM.

In this scenario, the IMPULSE project [3] and its multidisciplinary team are focusing on the multidimensional and user-centric analysis of the transformative impact of two disruptive technologies (i.e., blockchain and artificial intelligence) on electronic IDentities (eID) for the improvement of digital public services.

This paper aims at presenting an IdM, based on blockchain and smart contracts, and its impacts from ethics, regulatory and technical perspectives. The paper is structured in four main sections representing the steps during the research process. After the introduction, Section 2 deals with the analysis of "Identity Management" and related ethics and regulatory framework. This section is relevant to better comprehend the underlying framework composed by ethics principles and regulations (i.e., General Data Protection Regulation – GDPR, electronic Identification, Authentication and Trust Services - eIDAS) which impact IdM. This framework establishes "privacy-preserving" requirements and guidelines to be embedded into the development process and outputs. Section 3 describes the "Design Approach" which is built on top of the framework and benefits from a decentralisation model enabled by distributed ledgers (i.e., blockchain technology) and the smart contracts. This section provides an analysis from the legal perspective and describes the innovative approach proposed for an easier comprehension and management of informed consents, which are fundamental elements to legally binding data access to personal data according to the GDPR. Finally, Section 4 provides conclusions on the work in progress of the IMPULSE project and the next steps.

II. IDENTITY MANAGEMENT

A. Literature Review

In literature, there are multiple definitions of Identity Management, each stressing one or more facets of IdM.

"Identity Management system provides the tools for managing all partial identities of an individual in digital world. A partial identity may or may not uniquely identify an individual." [4]

"Identity Management is a set of functions and capabilities, for administration, management, maintenance, discovery, information exchange, policy enforcement and authentication. This is used to ensure identity information and security. It provides tools for managing individual identities in a digital environment." [5]

"Identity Management seeks to solve the problem of remembering different user names and passwords for accessing organizations. It includes fair and lawful processing, purpose specification, data participation and control, disclosure and information security." [6]

"Identity Management systems are used to manage user identities across multiple systems and providing a way to user access in the organization. This is done for the whole life cycle of a user in the organization by single sign-on and keeping a check on user's credentials." [7]

These definitions are collectively summarising various aspects of IdM.

In order to better understand the concept of IdM, it is necessary to turn to its main component, namely the concept of identity. Examining it at the philosophical and legal levels becomes clear its complex nature. Throughout history, identity has been deeply connected to state control over individuals and people's rights and the construction of the self through a set of relational structures.

A unique identity is needed to recognise individuals' rights and responsibilities. However, being associated with a specific identity can grant privileges or lead to exclusion and discrimination. Additionally, the question arises whether identity can remain constant as a person evolves and whether individuals have the right to an identity that aligns with their self-perception.

Manders-Huits [8] points out the risk of reducing an individual's identity to a simplified "administrative" notion of it, which fails to capture its true complexity. Ishmahev and Stokkink [9] emphasise the difficulty of overcoming the complexities of identity in a workable definition that can serve as the ethical foundation for an identity management system, whether digital or otherwise.

Attempts to simplify the approach to "identity" inevitably impact key moral aspects like autonomy, self-determination and self-identification (Manders-Huits and Hoven) [10]. Ishmahev and Stokkink's analysis reveals a tension between an approach to identity management focused on the individual and his/her rights and the society.

Ishmahev and Stokkink point to the tension between individual rights and societal interests in Identity Management, using the example of the Chinese "Social Credit System" (SCS) where social scoring aims to identify and isolate "bad elements," potentially favouring society over individual rights and well-being.

The above is true of any Identity Management System but Zwitter et al. [11] highlight the new set of problems brought by Digital Identity, including the fragmentation of identity. In the digital space, individuals possess multiple Digital Identities issued by different providers, each with distinct attributes. This raises the question of whether we should strive for a single persistent identity and impose it.

Digital identity has become a crucial infrastructure service with different rules and obligations for accessing various services. It is not neutral in its shaping and management, as each provider sets its own rules, leading to fragmentation.

This raises ethical considerations about how much an individual can be considered accountable over his/her action, and how much anonymity and freedom can be favoured over public responsibility and liability. Lessig [12] emphasises that these choices are political and moral decisions when designing network capabilities and participation rules.

There are three primary models of Identity Management: Centralised, Federated, and Decentralised. The IMPULSE solution aligns with the Decentralised model, following the Self-Sovereign Identity (SSI) model and is based on Blockchain (Blockchain-Based Identity Management System (B-Based IdM)). It is important to remark that the SSI model allows the users to fully control their own data, satisfying a fundamental right defined in the GDPR regulation [2], and for this reason the European Commission selected the SSI model to for its European Digital Wallet initiative.

B. Ethics and Regulatory Framework

Within the EU framework, there is no specific regulation for B-Based IdM, but the eIDAS regulation (910/2014) [13] addresses Identity Management in general. While only one standard directly focuses on B-Based IdM, other standards cover general IdM or distributed ledgers. Consequently, there is an urgent need for EU regulation and standardisation for B-Based IdM Systems. In terms of relevant regulation law framework, the GDPR's [2] applicability to IdM systems based on Blockchain is a subject of debate. Limited scientific literature systematically addresses this issue, necessitating further understanding. Many scholars have identified several challenges to blockchain's GDPR compliance. Among them are:

The immutability of Blockchain poses difficulties in fulfilling the "right to erasure" (Article 17 [2]) and the "right of rectification" (Article 16 [2]) since blocks cannot be deleted or modified without compromising the chain's integrity. Similarly, "withdrawing consent" (Article 7 [2]) and "defining data controllers" (Article 4 [2]) become complex due to Blockchain's replication and peer-to-peer nature.

Additionally, the classification of hashed identifiers on the Blockchain as anonymised or pseudonymised data raises questions. While some argue that hashed personal data is pseudonymised, others consider it personal data subject to GDPR. The issue of pseudonymity or anonymity of hashing remains unresolved, creating a grey area.

Moreover, Self-Sovereign IdM, Kondova & Erbguth [14] state that:

"Self-Sovereign Identity (SSI) involves personal data. A detailed analysis of the system used and the use-case is required to determine what data components of the SSI constitute personal data, how the GDPR applies and who is considered to be a controller and what justifications exist. When storing some data on an immutable blockchain, it has to be ensured, that either the data stored on a blockchain will not or no longer constitute personal data, that the data subject is considered to be the controller, that the household exemption applies or a permanent justification for continuous storage on the blockchain exists. In many cases, according to Art. 35 GDPR [2], a data protection impact analysis (DPIA) will be required.”

Concerning eIDAS (electronic Identification, Authentication and Trust Services) Regulation, it aims to establish trustable and reliable Digital Identities, replacing physical devices like smart cards with other authentication methods. eIDAS provides a common framework in Europe for e-signature and e-identity authentication. The Regulation distinguishes three assurance levels for electronic identification means, varying based on the degree of confidence in asserted identities.

Although the literature on Blockchain's compliance with e-IDAS is limited, integrating Blockchain into existing eIDAS standards does not seem contradictory. Blockchain can enhance security and enable the signing of various object types. Immutability remains a significant concern, also in the case of digital seals and signatures, but it does not seem to go against any part of eIDAS. Further analysis is needed to ensure compliance between B-Based IdM systems and eIDAS regulation, considering the crucial role of digital signatures and certificates in identity management.

Aiming to deliver a secure and trusted digital identity for all EU citizens, The European Commission (EC) on 3 June 2021 proposed a framework for a European Digital Identity, which builds on the revision of the current eIDAS Regulation. The EU Proposal [15], which is commonly named eIDAS 2, is currently under trilogue negotiations.

Among many amendments and changes to eIDAS, the Proposal has introduced a novel element, European Digital Identity (EUDI) Wallet which will be issued by every Member State and will be available to all EU citizens, residents, and businesses in the EU. The Wallet promotes social inclusion and fundamental rights, complying with the Charter of Fundamental Rights of the European Union. It emphasises personal data protection, security, reduced risk of ID theft, equality, solidarity, inclusion, engagement, freedom of movement, and residence. The EUDI Wallet aligns with the aims of the European Union, particularly regarding the protection and promotion of individual rights, personal data, access to services, and freedom of movement. Another innovative aspect presented by the European Commission in the Proposal concerns Article 45 (Section 11) regarding Electronic Ledgers which are effective against cyberattacks and they are present in Blockchain and Distributed Ledger Technologies (DLT). However, the European Parliament's vote on the electronic identification Regulation removed electronic ledgers as trust services, potentially impacting blockchain companies' business opportunities in providing e-identity solutions. The European DIGITAL SME Alliance, together with major IT and Blockchain associations, has sent a letter [16] to the members of the European Parliament’s ITRE Committee, expressing concerns and calling for reinstating the provisions to ensure a future-proof eIDAS 2 Regulation supporting innovation and economic development.

III. DESIGN APPROACH

The IMPULSE project design approach to IdM is built on top of the established ethics and regulatory framework. This approach aims at maximising benefits from decentralisation model enabled by distributed ledgers taking into consideration ethics guidelines and principles, regulatory constraints, and the impact of blockchain for electronic IDentities (eIDs) in public services from diverse standpoints, including legal, ethical, socio-economic and socio-political, technical and operational.

Indeed, decentralised model of the blockchain guarantees:

i) **Tamper-resistance and data integrity** - by design, the blockchain is a permanent and immutable storage of data blocks, therefore data added on the ledger cannot be intentionally or accidentally changed, altered or deleted by anyone. This maintains data integrity.

ii) **Data transparency and auditability** - data in the blockchain can be traced and verified by everyone in the network belonging to the blockchain, as well as data blocks constitute themselves an auditable trail of data.

iii) **Data protection** - the use of cryptography to process and store data in the blockchain protects from unauthorised accesses.

iv) **Data sharing and availability** – each node of the distributed network replicates and shares a copy of data. This network of nodes ensures high availability infrastructure.

These benefits lay the foundation for an effective IdM system [17].

Moreover, to ensure compliance with the ethics and regulatory framework, IMPULSE enables data subjects to full control sharing of and access to their own personal data through informed consent. This means that the IMPULSE system has to provide a “Consent Management” feature allowing a user to consult policies, grant consents, show the history of provided consents, and modify (when feasible) provided consents. Moreover, by considering access to public services, the IMPULSE system has to provide a “Policy Management” feature allowing a public administration (PA) to create policies, modify existing policies and retrieve the status of the users’ granted consents. These two fundamental features derive from the consent mechanism defined in the GDPR regulation and for this reason the service is called GDPR Service. The use case diagram representing interactions among user, PA and GDPR Service is shown in Figure 1.

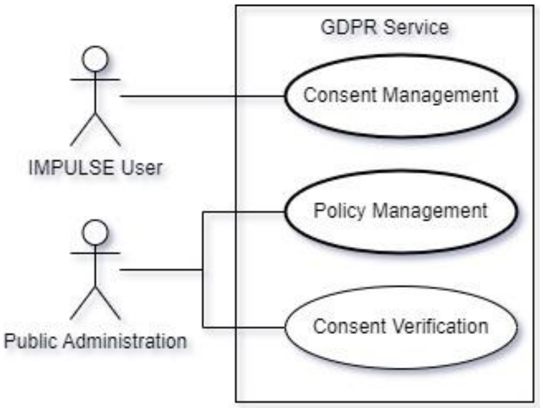


Figure 1. IMPULSE GDPR Service - use case diagram.

The GDPR Service is implemented by combining the usage of blockchain and smart contracts.

For the sake of clarity, usually, informed consents are too complicated to be fully interpreted and managed by users. For this reason, IMPULSE aims at simplifying this mechanism through the implementation of a more understandable and intuitive instance of informed consents, including a set of user-friendly meaningful icons. These icons allow user to understand what kind of personal data is going to be shared in a quick and simple way.

A. *Smart Contracts: Legal Perspectives*

The nature of a smart contract is subject to numerous conflicting viewpoints.

Legally, a "smart contract" refers to a contract represented and executed by software. Programmers instead view smart contracts as algorithmic code that performs tasks when certain conditions are met, often on a distributed ledger.

Nick Szabo was the first who proposed the concept of a smart contract in the late 1990s, defining it as a "computerised transaction protocol that executes the terms of a contract" [18]. Although not particularly innovative, Szabo argues that this idea distinguishes smart contracts by being purely digital and involving complex calculations, multipart deals, rights transfers, and encryption. It is crucial to acknowledge that a smart contract is not equivalent to a legal contract. In fact, it is often claimed that the term "smart contract" is misleading since, in many cases, smart contracts are neither intelligent nor contracts. Scholars aligned with the programmer's viewpoint contend that smart contracts do not meet the legal definition of contracts. Geiregat, instead, describes smart contracts as hardware or software “that initiates, controls, and/or documents legally relevant acts, depending on predetermined and digitally proven events, and by means of which legally binding contracts may be concluded, depending on the circumstances” [19].

Regarding the notion of smart contracts and their legal nature, there are two exact antipodes of opinion in the scientific community. Each perspective offers different solutions to the scientific challenge of establishing the legal force and effect of smart contracts. From the programmer's standpoint, a smart contract is a code designed to perform various tasks when specific conditions are satisfied. In contrast, another group of scholars supporting both traditional and eclectic views see a smart contract as a dual phenomenon that encompasses both technical and legal components. These two aspects do not merge into a unified whole.

B. *Implementation based on Smart Contracts*

Smart contracts are considered and used since the “onboarding” process, i.e., when users grant consent to a third party to access a set of predefined data, and applied during the “usage” process, i.e., when users can decide to or not to grant consent, or when necessary, to modify it according to specific needs and conditions. For these processes, users need to view and accept terms of usage of a given service and the IdM system provides them with a comprehensive view of required consents. To manage the informed consents, the IdM system defines a Consent Object which can assume two main states:

- Denied: when users want to deny access to their data to an entity, that, in the specific case of IMPULSE project, is represented by a Public Administration Service (PAS);
- Allowed: when users agree to a certain policy that enables a PAS to process data according to specific rules.

At the beginning of its lifecycle, a Consent Object is, obviously, not yet created. It is created when a user accepts or rejects a certain policy presented by the requested entity.

The Consent Object can change its own state, according to one of the following events: i) a user accepts the previously rejected policy; ii) a user rejects a previously accepted policy; iii) a user revokes the granted consent; iv) a policy change occurs; and v) a consent expires over time.

Development with smart contracts and blockchain requires to mind the “immutability” constraint which implies the unsuitability to add in the smart contract data model users’ identifiers. To prevent this concern, the GDPR service will replace users’ identifiers with pseudo-identifiers (pseudo-ids): these act as links among users and their consent objects stored on the smart contracts. Pseudo-ids are generated by using hash functions which ensure the following characteristics:

- **Deterministic**: a specific input message returns always the same output hash message. This allows the GDPR Service to process identifiers from input data without managing pseudo-ids;
- **Unique**: for each input message exists always a unique output message. This ensures security against brute force attacks and avoid collisions between pseudo-ids;
- **One-way**: hash functions cannot allow to derive the input message from the hash output message. This ensures that there is no way to directly or indirectly identify users and their data.

The GDPR service plans to make use of secure hash functions (e.g., SHA-256) to generate pseudo-ids. Pseudonymization mechanism can be enriched by combining other attributes (e.g., consent object data). The following schema in Figure 2 shows a possible approach where the pseudo-id is generated using as input the concatenation of user’s public key, pas id and policy id.

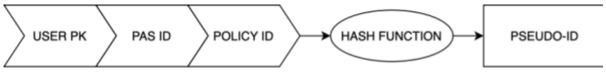


Figure 2. Pseudonymization approach.

By using the pseudo-ids, the GDPR Service works as an intermediary between personal data storages, which rely on traditional databases (such as PostgreSQL – see Figure 3), and third-party components, which need access to personal data to grant services.

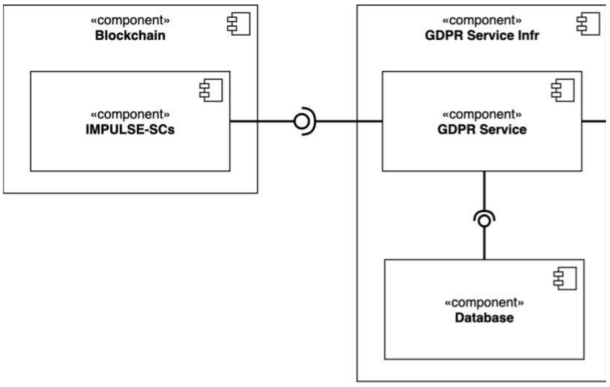


Figure 3. Component Diagram.

This approach ensures compliance with GDPR principles such as "right to modification" and "right to erasure". Indeed, smart contracts are used to store transactions (i.e., consents) with references to data stored in the traditional database and by applying hashing functions. This ensures data integrity and auditability. To better clarify the typical use case scenario, it is considered the user requesting access to a public administration service through an application (e.g., IMPULSE app).

The GDPR Service acts as intermediary between the mobile app and the public administration service, managing the smart contracts to allow i) user registration and ii) request of the user to access. The registration phase allows users to read and understand the terms and conditions to use the system, and if they agree with the policy in place, they accept the consent. Policy acceptance enacts the GDPR Service to register metadata of the consent object, generate the pseudo-identifiers and create the smart contract.

When users want to access a public administration service, the GDPR Service checks the existence and status of the consent object, therefore based on this consent verifies its validity and enables/disables access to the public administration service. In case the consent object is "not valid" (e.g., policy expiration, revocation or modification) the application will have to show users with policy and perform appropriate rectification to the consent by involving the GDPR Service.

All these steps rely on actions performed by users which grant/revoke/modify consents based on policies defined with the Public Administration. Policies are described in the informed consents and therefore these have to be clearly understood to take the right decision.

C. *Icons for representing consent information*

To solve the long-standing problem of understanding often hard to grasp consent forms, the IMPULSE project proposes the use of a visual-based language, based on a selected and adapted set of icons from the Italian Data Protection Authority (see Figure 4) and integrated in the IdM system, with which citizens can make a complete decision about their data (e.g., who will have the permission of process their data, for which purpose, for how much time) in a fully informed, simpler and more comprehensible way.

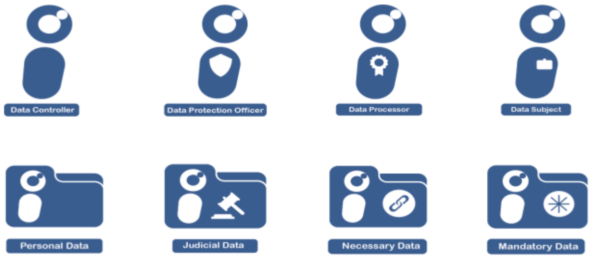
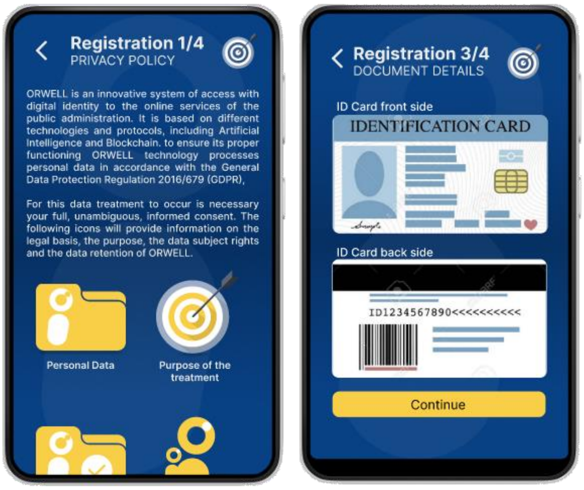


Figure 4. Set of icons to represent consent information.

This approach strengthens the comprehension of policies described in the consent, in compliance with the GDPR rules, and it represents a novelty in Identity Management systems.

This approach empowers users by respecting and protecting their fundamental rights (i.e., enabling full control of personal data), is used since the onboarding phase of the user (i.e., registration) as shown in Figure 5.



ORWELL is an innovative system of access with digital identity to the online services of the public administration. It is based on different technologies and protocols, including Artificial Intelligence and Blockchain. to ensure its proper functioning ORWELL technology processes personal data in accordance with the General Data Protection Regulation 2016/679 (GDPR),

For this data treatment to occur is necessary your full, unambiguous, informed consent. The following icons will provide information on the legal basis, the purpose, the data subject rights and the data retention of ORWELL.

Figure 5. Two phases of the user' registration.

IV. CONCLUSION AND FUTURE WORK

This paper described a proposed solution to tackle the identified concerns of Identity Management (IdM) taking into considerations ethics, regulatory and technical perspectives. To achieve this goal, the research team of the IMPULSE project has established an "ethics-by-conception" approach starting from the identification of ethics principles, guidelines and regulation constraints impacting the IdM system.

Specifically, the IdM is expected to i) be compliant with current European regulations (i.e., GDPR and eIDAS); ii) be respectful of fundamental human rights (i.e., privacy and Internet access); iii) simplify user experience when dealing with full control of personal data and informed consents; iv) be built by adopting blockchain technology and smart contracts mechanisms.

While the current implementation in the IMPULSE project is demonstrating the feasibility and validity of design approach, as well as its compliance with the established ethics and regulatory framework, the research team is working for defining a mobile application (app) which, integrated with the GDPR Service, will allow users to full control their own personal data and informed consents.

A preliminary app is currently under development and the research team is working for finalizing and validating it.

The human-centred approach adopted in the IMPULSE project demonstrates that the integration of Social Science and Humanities (SSH) perspectives into technology development improves the comprehension of aspects and details that usually might be hidden and overlooked from the technical team, due to lack of competences. Therefore, an integrated and multidisciplinary team, as the IMPULSE project experienced, allows to understand and identify the multifaceted aspects belonging to a system to be adopted in the society, including inter-alia its impacts, and this contributes to identify potential social barriers and adopt appropriate countermeasures.

Adopting this approach since the beginning of the project and applying a continuous assessment allows a flawlessly development. Indeed, SSH experts and technology developers will be able to identify requirements and implement them since the beginning. Assuming that a single assessment will be able to ensure compliance with ethics and regulatory frameworks is practically unreasonable: new requirements will come up during the assessment, and the implemented system will risk the regulatory compliance, the acceptance from citizens and its business objectives.

ACKNOWLEDGMENT

This work is part of the IMPULSE project that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101004459.

REFERENCES

[1] United Nations, "The promotion, protection and enjoyment of human rights on the Internet," July 2021. [Online] [Retrieved: August, 2023]. Available: https://documents-dds-ny.un.org/doc/UNDOC/LTD/G21/173/56/PDF/G2117356.pdf

[2] European Parliament and the Council, “The Regulation (EU) 2016/679 on on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),” [Online] [Retrieved: August, 2023]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[3] IMPULSE Project website, [Online] [Retrieved: August, 2023]. Available: https://www.impulse-h2020.eu/.

[4] S. Clauß and M. Köhntopp, "Identity management and its support of multilateral security," Computer Networks, vol. 37, no. 2, pp. 205-219, 2001.

[5] D. Chadwick, "Federated Identity Management," Foundations of Security Analysis and Design, Lecture Notes in Computer Science, vol. 5705, pp. 96-120, 2009.

[6] T. Olsen and T. Mahler, "Identity Management and Data Protection Law: Risk, Responsibility and Compliance," Circles of Trust. Computer Law & Security Report, vol. 23, no. 4, pp. 342-351, 2007.

[7] K. Tracy, "Identity management systems," IEEE Potentials, vol. 27, no. 6, pp. 34-37, 2008.

[8] N. Manders-Huits, "Practical versus moral identities in identity management," Ethics and Information Technology, vol. 12, p. 43–55, 2010.

[9] S. Q. Ishmaev Georgy, "Identity Management Systems: Singular Identities and Multiple Moral Issues, Frontiers in Blockchain," vol. 3, 2020. [Online] [Retrieved: August, 2023]. Available: https://doi.org/10.3389/fbloc.2020.00015

[10] N. Manders-Huits and J. van der Hoven, "The Need for a Value-Sensitive Design of Communication Infrastructures," Evaluating New Technologies. The International Library of Ethics, Law and Technology, vol. 3, pp. 51-60, 2009.

[11] A. Zwitter, O.J. Gstrein, and E. Yap, "Digital Identity and the Blockchain: Universal Identity Management and the Concept of the ‘Self-Sovereign’ Individual," 28 May 2020.

[12] L. Lessig, "Code and Other Laws of Cyberspace, Version 2.0," 2006. ISBN 978-0465039142.

[13] European Parliament and the Council, "The Regulation (EU) No 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC," [Online] [Retrieved: August, 2023]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0910.

[14] G. Kondova and J. Erbguth, “Self-Sovereign Identity on Public Blockchains and the GDPR”. Proceedings of ACM SAC Conference, Brno, Czech Republic, March 30- April 3, 2020 (SAC’20), pp. 342 – 345, 2020. DOI: 10.1145/3341105.3374066.

[15] The Commission of the European Union, "Proposal for a Regulation amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity," [Online] [Retrieved: August, 2023]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0281.

[16] "Open Letter for the preservation of the electronic ledger’s provisions in eIDAS 2," [Online] [Retrieved: August, 2023]. Available: https://inatba.org/news/savesection11-eidas2-trusted-electronic-ledgers-open-letter/.

[17] D. Augot, H. Chabanne, T. Chenevier, and W. George, "A user-centric system for verified identities on the bitcoin blockchain," Lecture Notes in Computer Science, vol. 10436 LNCS, p. 390–407, 2017.

[18] N. Szabo, "Smart Contracts: Building Blocks for Digital Markets," 1996.

[19] S. Geiregat, "Cryptocurrencies Are (Smart) Contracts,” Computer Law & Security Review, vol. 34, issue 5, pp. 1144-1149, 2018. ISSN 0267-3649.