Integrating Activities for Advanced Communities

D6.1 - A Preliminary Study of Inquiries and Needs from Station Managers and Researchers

Project No.871120– INTERACT

H2020-INFRAIA-2019-1

Start date of project: 2020/01/01 Duration: 48 months

Due date of deliverable: 2021/02/28 (M14) Actual Submission date: 2021/02/19

Lead partner for deliverable: AFRY
Authors: Maria Erman, Thèrése Skarstedt and Markus Skogsmo

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Table of Contents

Publishable Executive Summary 3
Table of Acronyms 4
1. Introduction 5
1.1. Background 5
1.2. Purpose and Goals 6
1.3. Tasks 8
1.4. Methods 8
1.4.1. Background and Context for the Sampling of Prospective Informants 8
1.4.2. Sampling and Process for the Interviews 9
1.4.3. Questionnaire as the Basis for a Quantitative Study 11
1.5. Outline 11
2. Theoretical Framework 12
2.1. An Introduction to Artificial Intelligence and Machine Learning 12
2.2. An Introduction to Artificial Neural Networks 15
3. A Qualitative Study Based on Interviews 17
3.1. Introduction 17
3.2. Research and Projects 17
3.3. Obstacles 18
3.4. Data Collected 19
3.5. General Opinions and Curiosity about AI 20
3.6. Usage of AI – Current and Future 21
4. A Quantitative Study Based on Questionnaire Responses 23
4.1. Introduction 23
4.2. Results and Visualization of Data 24
5. Analysis and Discussion 29
6. Conclusions 30
6.1. Future Work 30
Appendix A – Template Interview Questions 31
General 31
Input and Data in General 31
Previous Knowledge about AI 32
Finding a Common Ground 32
Questions or Comments to Us 32
Appendix B – Questionnaire 33
Artificial Intelligence in Arctic Research 33
General Information 33
AI Usage 34
AI Potential 35
AI Data 36
Pilot Studies 37
Other 38
Bibliography 39

Publishable Executive Summary

This document presents the results of a combined quantitative and qualitative study on needs, perceptions and uses of AI and ML techniques and methods by station managers and researchers involved in the International Network for Terrestrial Research and Monitoring in the Arctic (INTERACT).

The qualitative part of the study consists of a questionnaire provided to the station managers and researchers, as well as structured interviews. This is collated with an analysis of discussions held during workshops to form an idea of how best to proceed with applying AI and ML techniques in INTERACT III. The quantitative part consists of classifying answers into distinct categories to provide a sense of the most important areas of interest for the researchers.

The results of the study show that several research groups are keenly interested in applying AI and ML in their research but do not know how to do so. The perception appears to be that the main use of AI and ML is to support in decreasing the amount of tedious and error-prone field work and associated analysis and documentation. However, as the researchers are not experts in AI and ML, further work in WP6 is needed to simplify access to ML services, as well as supporting researchers in how to best apply AI and ML to their specific domains and associated data sets.

The contents of this report will serve as input to the upcoming WP6 deliverables D6.3 and D6.4. D6.3 will provide a demonstration and a presentation of a pilot project in which AI and ML techniques are applied on researchers' data. D6.4 will present a report on future strategy and planning for the areas of how AI and ML can be applied on Arctic research.

Table of Acronyms

AGI	Artificial general intelligence
AI	Artificial intelligence
ANI	Artificial narrow intelligence
ANN	Artificial neural networks
COVID-19	Coronavirus Disease 2019
DL	Deep learning
ML	Machine learning
NLP	Natural language processing
OCR	Optical character recognition
SC	Societal challenges
WP	Work package

1. Introduction

1.1.Background

With rising temperatures occurring at least twice as fast as in the rest of the world, the Arctic is an important area to study in order to gain an in depth understanding of climate changes. It is not only a region, but also a physical, biological, chemical and climatological system. No other area is as impacted by global warming as the Arctic. The region may seem remote, but with ice melting, permafrost thawing, retreating glaciers, and consequent rising sea levels, it also affects the climate and the weather elsewhere on earth [1].

Global warming directly and indirectly impacts plants and animals, with for instance species such as polar bears, walruses and Arctic foxes finding it more difficult to hunt with diminishing ice. As an example of cascading effects, it can be mentioned that lichens and mosses, which are food sources to Caribous, are sensitive to warming, hence leading to a declining Caribou population. This in turn also affects scavengers and predators [1][2].

Indigenous people, who have lived in the Arctic for thousands of years, and whose cultures are shaped by the environment, are particularly percipient observers regarding climate changes. Noticeably, indigenous people have observed a reduction of sea ice, weather patterns changing and becoming less predictable, and species not previously seen appearing etcetera [3, pp. 4–5][4].

Environmental change in the Arctic is thus highly complex, and combined with few people living in the Arctic, data is sparse. Hence, to aggregate data, and attempting to model and understand the effects of environmental change on ecosystems, it is of paramount importance to conduct field work in a variety of areas, with the aim of obtaining an understanding of what the parameters are and how they are related. Furthermore, various forms of documentation, be it measurements, pictures, videos, sounds, or satellite imagery, will be required, not only for present, but also future research.

Great progress has been made by Academia in analyzing and documenting environmental change, and with the recent explosion of AI and ML techniques, there are now new possibilities and perspectives to take into account. It has become accepted that "there are many ways of knowing", and in addition to conventional and traditional knowledge, there are also resources such as logbooks, expedition reports, photographs and paintings that can be used to extend existing knowledge.

In essence, AI is about mimicking human intelligence, with ML being a set of tools, techniques and algorithms to achieve AI. Moreover, AI/ML has been recognized not only as powerful research tools in several scientific disciplines, but it also shines when it comes to automating manual and tedious work already performed by humans today. It does this by learning from sources of information and applying this knowledge on new information. As such, application areas are as diverse as image recognition, object detection, self-driving cars, optical character recognition (OCR), syntax analysis, sentiment analysis of text etcetera [5].

1.2. Purpose and Goals

Valuable data are collected by dedicated researchers involved in INTERACT III, and some research stations have even conducted measurements for many decades. Not all data collected thus far may seem to have a purpose. However, in the future, analysis of these data may provide the missing link to understanding key concepts and bringing crucial insights into climate change research. Those additional photographs, recordings or measurements could aid e.g., a future state-of-the-art AI algorithm finding patterns or discover previously unknown relationships between parameters. With this in mind, it is important to take into account that even today there are ways for researchers to prepare and curate their data sets to be as useful and future proof as possible.

With AI/ML in mind, the purpose of D6.1 is to investigate:

1. What inquiries and needs do researchers and station managers have?
2. What research is conducted?
3. What data are collected?
4. What obstacles are encountered by the researchers?

An objective is to look into how AI/ML could be of help in research, in field work or in the documentation and analysis stages. Furthermore, it is also important to consider what kind of data would best be suited for AI/ML, whether that data set is available for other researchers, and how it can be accessed. The results of D6.1 will in turn lay the foundation for future deliverables, such as a demonstration being held where AI/ML is applied on selected data sets in collaboration with selected station managers.

Figure 1 illustrates how D6.1 is connected to the overall work of WP6. This preliminary study both inspired the workshops held by WP6, and conversely, the workshop discussions as focus groups, validated the interviews in D6.2. In addition, master theses have both gained from, and contributed to the results of this preliminary study. D6.1 will further lay the foundation for D6.3, where a demonstration will be held, in which AI/ML will be applied on researchers' data. As illustrated in the figure, these data can consist of both conventional data, in the form of e.g., digital images, text, sound files etcetera, as well as unconventional data, in the form of e.g., logbooks, private photographs, paintings etcetera. The orange circle denotes the final destination of WP6 with D6.4, where a best practice scheme is to be produced that can be used at Arctic research stations.

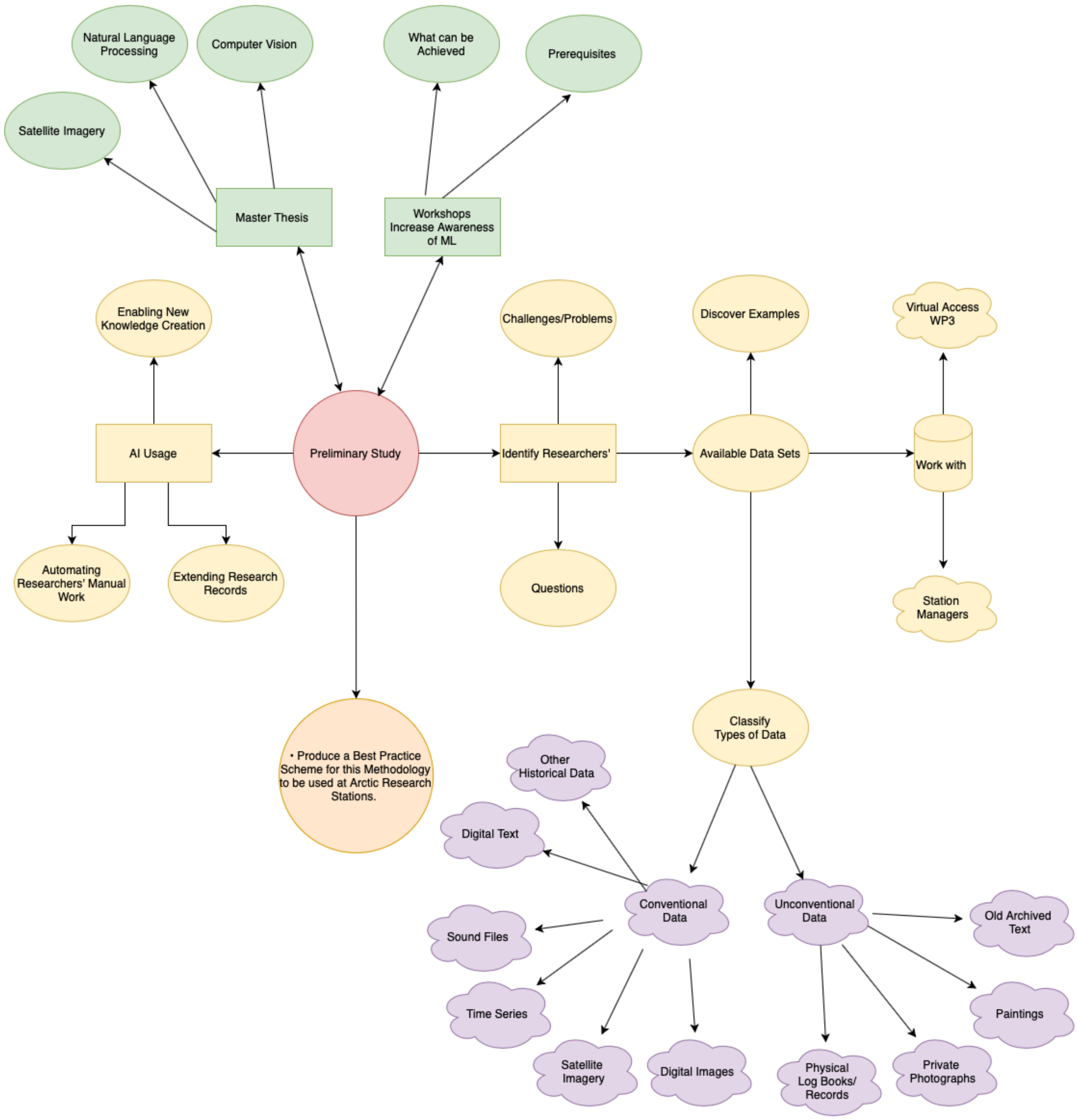

Figure 1: The Role of D6.1 in WP6.

1.3. Tasks

Several tasks form the basis of this preliminary study, which includes qualitative and quantitative analyzes:

- A questionnaire (Appendix A) has been created and subsequently sent out to station managers. The questionnaire deals with whether researchers currently employ AI/ML techniques in their work, a (possible) interest in, and ideas about AI/ML, as well as data sets that can benefit from using AI/ML.
- The compiled responses of the questionnaire constitute the quantitative analysis part.
- Interviews have been conducted. The template for the questions can be found in Appendix B. The purpose of the interviews is to go more in depth compared to the questionnaire, and also to follow up on questions in order to obtain more information from the interviewees.
- Compiled answers from interviews, as well as details from discussions during previously held workshops constituting D6.2, form the qualitive analysis part.

1.4. Methods

1.4.1. Background and Context for the Sampling of Prospective Informants

INTERACT III is a fully functioning infrastructure responding to societal challenges (SC) both at local and worldwide scales. SCs have been chosen based on:

> Discussions with local communities, Indigenous Peoples, decision makers, researchers, policy making communities and reference to the EU Horizon 2020 societal challenges, UN sustainable development goals, the joint statement of Ministers from second Arctic Science Ministerial and the EU-PolarNet white papers. [6, p. 8]

Of particular relevance to WP6 is SC3, which has as its focal challenges to find patterns and distil information from current research data, as well as extending available records by using AI and ML techniques on currently unavailable data [6, p. 8]. SC3 also connects with the UN Sustainable Development Goal 13, “Climate Action” [7].

Research stations of particular relevance to SC3, sorted according to environmental envelopes or ecozones have been collated by WP3 into an infrastructure matrix, see Figure 2. The ecozones A-F refer to:

A. High-Arctic and tundra stations with continuous permafrost.
B. Tundra and boreal forest stations with discontinuous permafrost.
C. Tundra and boreal forest stations with sporadic permafrost.
D. Boreal forest stations with no permafrost.
E. Mountain and alpine stations.

F. North-Atlantic and temperate forest stations. [8]

For more information about the denoted research stations in the infrastructure matrix, see the webpage at [8].

1.4.2.Sampling and Process for the Interviews

Purposive sampling was used, due to there being a restricted number of people with expertise appropriate for the study, and also because of the answers to the questions being deemed the most important [9]. The infrastructure matrix formed the first selection of criterium on what researchers to invite for the interviews, i.e., station managers or researchers from the research stations denoted as particularly relevant to SC3. In addition, informants were drawn from all ecozones, as well as their being represented with geographical spread. The intent was to be able to gain as much variety as possible both with regards to research applications and type of data collected.

In addition, interviewees were also invited according to having previously shown particular interest in AI/ML, such as already having relevant problems with AI/ML in mind, filling out questionnaires (with in-depth interviews being a good opportunity to follow up on the questionnaire answers), as well as participation in WP6 workshops, and/or having previous correspondence with WP6.

Finally, the type of data being available that WP6 find to be the most conducive with regards to AI/ML techniques was taken into consideration. Sampling of prospective informants was further narrowed down according to the number of selection criteria being fulfilled.

In-depth interviews were conducted with station managers and researchers for the purpose of collecting data on:

- Current and previous research projects as a background context.
- Obstacles encountered with regards to monitoring and research.
- Specifying collected research data, and where applicable, how it is structured, sorted, labelled etcetera.
- The general view of AI/ML.
- How AI/ML can be of help in climate monitoring and research.
- Whether station managers or researchers have specific inquiries, needs or ideas with regards of AI/ML, which in turn will influence future work of WP6.

18 station managers and researchers were invited, and out of these, there were four stations that either did not respond or said no to an interview, i.e., a response rate of approximately 78%.

INTERACT III Infrastructure Matrix

Responding to Societal Challenges	Environmental Envelope/Ecozone: A High Arctic and tundra with continuous permafrost	B Tundra and boreal forest with discontinuous permafrost	C Tundra and boreal forest stations with sporadic permafrost	D Boreal forest with no permafrost	E Mountain and alpine stations	F North-Atlantic and temperate forest stations
Unpredictable Arctic: environment, landscapes, icescapes, and ecosystems	56 21.1 55 21.3 44 50 7 21.2 51 32 5.2 37 5.1	10.1 3(1) 29 31 11 10.3 10.2 15.1 38 8 17 14 9.1	12 48 16 18 22 26 13 53 34 3.2 15.3	19 15.2 20.2 4 23	30 24 25	9.2 28 33 54
Connecting the Arctic: Transport and Communication	21.1 21.3 50 21.2 32 37	29 17	12 13 34	4 23	24	
Climate Action: Making data widely available (VA)	56 50 7 32 5.2 37 5.1	10.1 3(1) 29 31 10.3 38 14	12 16 18 22 26 13 53 34 15.3	15.2 20.2 4 23	30 24 25	
Preparing for a future world: improving education and governance	56 21.1 21.3 44 50 7 21.2 51 32 5.2 37 5.1	10.1 3(1) 29 31 11 10.3 15.1 38 17 14 9.1	12 48 16 18 22 26 13 53 15.3	19 15.2 4 23	30 25	9.2 28 33 54
Cleaner Arctic, cleaner world: documenting and reducing pollution	56 21.1 55 21.3 44 21.2 32 5.2 37 5.1	10.1 3(1) 29 31 11 10.3 10.2 15.1 38 8 17 14 9.1	12 48 16 18 22 13 53 34 3.2 15.3	19 15.2 20.2 4	24 25	9.2 33 54
The Arctic Resort: increasing benefits and reducing impact from developing tourism	21.1 21.3 44 21.2 32 37	10.1 3(1) 29 11 10.3 10.2 15.1 38 17 14	12 48 16 18 22 26 13 3.2 15.3	19 15.2 4	30 24	9.2 28 54

Figure 2: Research stations offering TA, RA and VA, particularly relevant for societal challenges 1-6, in accordance with the environmental envelopes/ecozones A-F. [8]

All interviews were conducted virtually either via Microsoft Teams or Zoom, and recorded, in order to facilitate coding the interviews, i.e., finding recurrent themes and patterns, further delineated in Section 3. The recordings were treated according to GDPR regulations. Furthermore, established interview techniques were used to minimize contamination, with the questions being sorted according to themes. Also, open-ended questions were asked in order for the participants to elaborate and use their own words.

The interview format used was "The general interview guide approach" [10], as introduced by Michael Patton. In this approach, there is an outline of topics, with the interview following a semi-structured manner, in which the wording and order of the questions are not pre-determined. The tone of the conversation remains informal and conversational.

In addition to the interviews, the following sources have been considered for the qualitative part of this preliminary study:

- Two mini workshops held by WP6 on June 10th, 2020, accommodating different time zones.
- The main workshop that took place on September 23rd, 2020.
- Where applicable, the open-ended questions in the questionnaire announced during the workshops, and the INTERACT III annual meeting on September 24th, 2020.

The workshop discussions were de-facto focus groups, where participants of similar backgrounds discussed how AI/ML can be used in their respective research fields, and as such, are part of this preliminary study to validate key findings from the interviews.

1.4.3. Questionnaire as the Basis for a Quantitative Study

The questionnaire that can be found in Appendix B, was announced during both the WP6 workshops as well as the INTERACT III annual meeting. 25 responses were received, with data points being extrapolated from relevant passages of conducted interviews, where applicable, to gain more data.

In summary, after extrapolating data from interviews, there are 30 responses that form the basis of the quantitative study in this report. Section 4 will delve further into analyzing and visualizing these data. Unfortunately, the amount of data is too small for more advanced analysis, using for instance data mining.

1.5. Outline

Section 1 presents a background, the purpose and goals, as well as the methods used in this preliminary study. Section 2 provides a theoretical framework, in which AI/ML concepts are introduced, such as computer vision (CV), OCR, natural language processing (NLP) etcetera. Further, Section 3 is a compilation on conducted interviews sorted according to identified topics. Section 4 delves into a quantitative analysis of the questionnaire results, and visualizations of the distilled data. Thereafter, Section 5 goes into discussing, as well as analyzing the results. Finally, Section 6 provides a conclusion and suggestions for future work.

2. Theoretical Framework

2.1. An Introduction to Artificial Intelligence and Machine Learning

The foundation of AI as it is currently known, was already laid in the 1950, even with advanced techniques such as artificial neural networks (ANN) being put forth. At the time, a generation of scientists, philosophers and mathematicians were occupied with AI. Alan Turing devised the Turing Test, arguing that a thinking machine is plausible. In addition, checkers and chess programs, challenging amateurs were created [11]. Previous to that, AI was intimidated in stories, such as with Tin Man, who lacks a heart in the "Wizard of Oz", or the human robot impersonating Maria in "Metropolis".

The hype and interest in AI, however, did not explode until recently. The reason for this is the large amount of data needed to be processed. For example, for a neural network to be able to classify images, thousands of images might be needed to train the network. In the case of text, the equivalent of a thousand books might be needed for a satisfactory result. In other words, a limited amount of computational power and memory needed for processing huge amounts of data prevented the breakthrough of AI until recently.

Today, AI is everywhere, and everyone seems to have an opinion. Companies and institutions are making adjustments for an AI transition, worried voices have become increasingly louder on how AI will affect the job market, there are complex ethical questions being discussed, and on the more dramatic side, AI will be the doom of mankind. Essentially, AI has become a buzz word, and buzz words tend to become unfashionable at a certain point, which is why it is important to shed light on the areas and applications, at which AI excel, and highlight why it is here to stay.

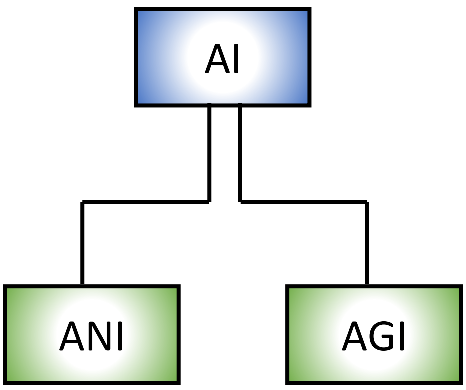

Figure 3: AI divided into two categories, artificial narrow intelligence and artificial general intelligence.
Source: Adapted from: [12]

Two important questions are: “how can we start benefitting from these powerful techniques today?”, and “what are the pre-requisites for using these techniques, and in what applications do they shine?”. Before answering these questions, a good place to start is to clarify what AI is, what it can and cannot do, and to cut through the hype.

Artificial intelligence has several connotations, and one of many ways to approach understanding AI, is to divide it into artificial narrow intelligence (ANI) and artificial general intelligence (AGI), see Figure 3. There is a great divide between these concepts. Whereas ANI does one thing incredibly well such as object detection, image classification, text analysis, translation etcetera; in contrast, AGI is about emulating humans as a whole, including human intelligence, and even surpassing humans. Most notably, there has been very little progress with regards to AGI, while ANI is thriving and creating immense value [12].

AI encompasses both ANI and AGI, with the former concept being what is in actuality used professionally with tangible results. On the contrary, in pop culture and literature, it is AGI that inspires popular myths and stories - after all, AGI does render itself very well into a dramatic and immersive format. The progress of ANI is thus mixed up with AGI, leading to irrational fears concerning the rise of the machines, robots taking over both existentially and/or rendering humans useless.

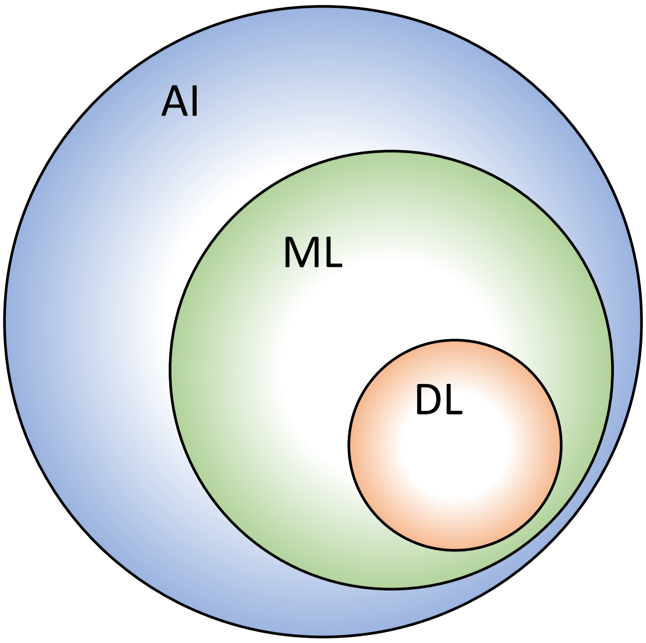

Figure 4: Artificial intelligence, of which machine learning and deep learning are conceptual subsets.

Furthermore, terms such as ML and DL are often used in connection with AI. There is confusion with regards to how to differentiate these concepts, and while they are used interchangeably, they do not mean the same thing. The obvious questions are, “what do these terms mean”, and “how do they relate to each other?” In short, AI is a superset of ML, which in turn is a superset of DL, see Figure 4.

More specifically, AI is the concept of emulating and mimicking human intelligence, behavioural patterns, interactions, sensations etcetera, and is in essence the theory of mind. AI is probably the hardest of the three terms to define since it has taken on different meanings depending on the domain in which it acts.

ML is one approach of reaching AI, by using methods to automatically learn from data without being explicitly programmed to do so. Moreover, ML is able to predict outcomes from similar data, and subsequently, improving from experience. To gain a clearer understanding, ML algorithms are commonly classified into three categories, supervised learning, unsupervised learning and reinforcement learning [13]:

- In the case of supervised learning, algorithms learn by example. With regards to training data, inputs are mapped to the correct output. Subsequently, the network adjusts itself based on training. After proper training, the network is employed, whereby new, unknown data are being mapped to the correct output in accordance with previous training.
- With unsupervised learning, the algorithm finds structures on its own, without having an algorithm being trained on pre-specified labels beforehand. Effectively, the aim of the unsupervised learning algorithm is to find hidden information and patterns in the input data, with the algorithm grouping unsorted data according to similarities. The algorithm thus categorizes data based on their similarities, differences and patterns.
- Reinforcement learning can be likened to teaching a dog. A dog explores, and with curiosity takes actions. These actions lead to consequences. If the dog eats healthy dog food, it is a good dog. On the other hand, if the dog eats the family lasagne, it is a bad dog. The dog learns by rewards and corrections, which is essentially a metaphor for how reinforcement learning works.

DL is the notion of employing ML techniques by using ANNs [14]. It can be said that DL are ML techniques that teach machines to learn by example. In DL, a computer model learns to make classification tasks from a large amount of data in the form of images/videos, text, numbers or sounds [15].

2.2. An Introduction to Artificial Neural Networks

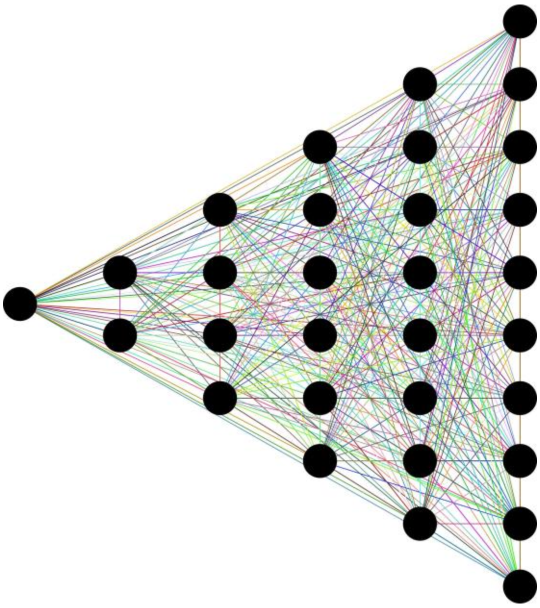

Figure 5: Artificial neural network with interconnected neurons.

Deep learning is generally concerned with ANNs, which are a technology inspired by how a brain operates with its biological neural network, and as such ANNs aim to mimic how a brain finds patterns or relationships among vast amounts of data [16].

An ANN consists of interconnected neurons or nodes structured in layers: An input layer, an output layer and what is denoted as hidden layers that exist between the input and the output layers, see Figure 5. Each layer extracts a different set of features from the input data, where an input, for instance, might be an image, and a set of features being edges. With every layer, the feature detection becomes more and more refined [17].

For a basic overview of how ANNs work, the example of inputs being images and outputs being classes, for instance bird and polar bear, can be used. Firstly, the network is trained, or e.g., a few thousand images are fed into the network. The network then adjusts itself, or the weights of the nodes to be more specific, in accordance with optimization techniques, i.e., we know what images we feed into the network and we know the output classes. The network, after being properly trained, is now ready for unknown images to be fed into it and then output the correct class, see Figure 6 [18].

Figure 6: An image recognition neural network model detecting animals and classifying them into correct species. (a) Bird. (b) Polar bear. [18]

In a similar way, a neural network can be trained on text. For the network to be accurate, usually text of the order of a few thousand books might be needed. Words, according to the algorithm used, are often mapped to vectors, and after training, patterns can be discerned with similar words being closer to each other, see Figure 7. The results will be highly dependent on the data it was trained on, and in this case the data was trained on a BBC News data set[1] [19].

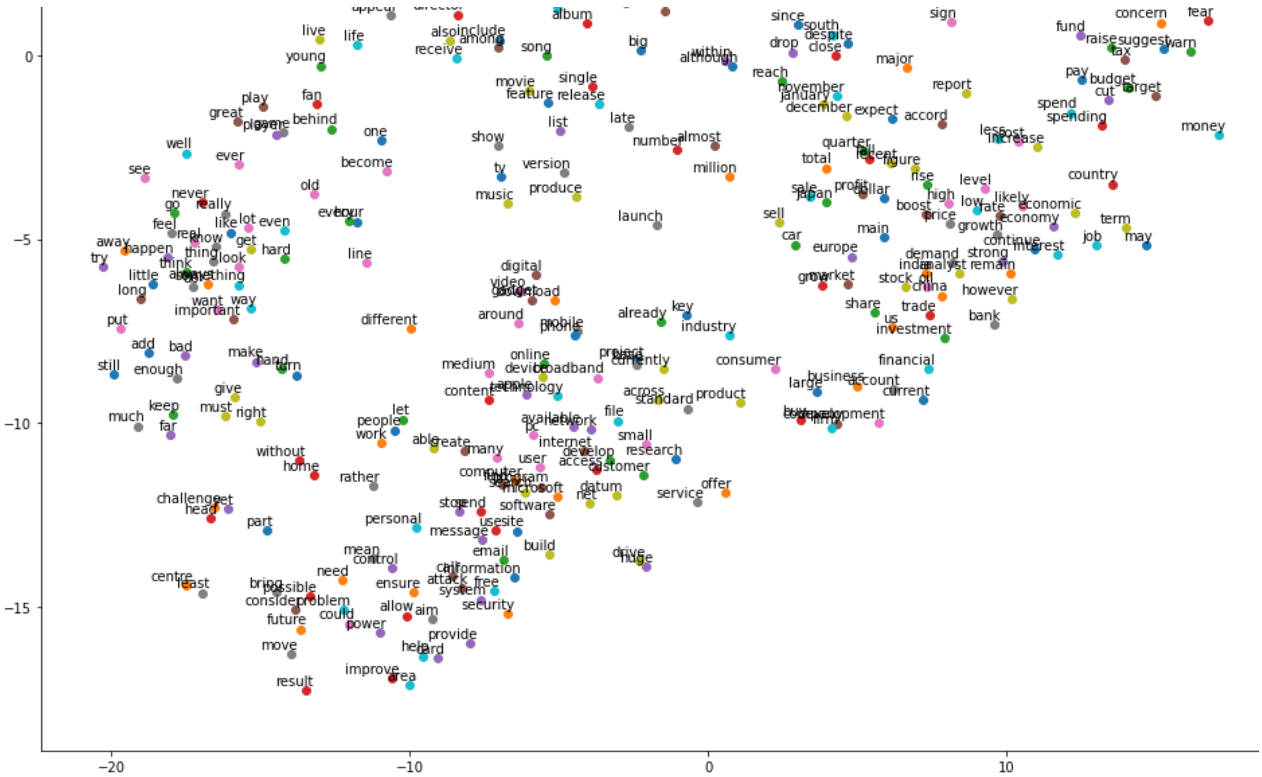

Figure 7: The result of a neural network being trained on text, using a BBC News data set.

Training on huge amounts of data can be problematic, since it is difficult to oversee whether the original data is inherently biased, or worse, there being scathing controversies sparked due to outputs, no matter how rare, being racist, sexist, ageist, etcetera, as was recently the case with the South Korean chatbot, Lee Luda, trained on 10 billion real-life conversations, and consequently suspended from Facebook for hate speech [20].

ANNs are used in a wide variety of applications, for instance, in time series forecasts, image recognition, classification, self-driving cars, generation of incredibly realistic CGI faces, recommender systems, automatic translation, text to speech, sentiment analysis, etcetera.

[1] The Natural language processing algorithm Word2Vec [19] was implemented, using the Gensim Python library [21].

3. A Qualitative Study Based on Interviews

3.1. Introduction

Interviews were held in order to dig deeper into how researchers and station managers regard AI, what kind of data they collect and whether these data could benefit from AI/ML algorithms. 14 researchers and station managers out of 18 agreed to be interviewed, which is a response frequency of 78%. The informants were drawn from various ecozones, and with a geographical spread. 64% of the informants were female.

Henceforth, if the interviewees are quoted, they will be referred to as P1 – P14.

3.2. Research and Projects

Environmental changes in the Arctic are highly complex. In order to gain a deeper understanding of these changes, their various parameters, how they are related, and thus attempting to model them, it is imperative to conduct research and monitoring in a wide variety of areas.

An outline of the interviewees' research and projects, as brought up during the interviews, has been compiled for context, as well as yielding a background on various perspectives and ideas conducive to this study. Specifically, the following research areas were mentioned during the interviews:

- Carbon balance, concerning how plant species interact with carbon and nitrogen cycling in soil, as well as permafrost greenhouse gas emissions.
- Research on glaciers.
- Studies on grazing effects on moose and reindeer.
- Ornithology, i.e., studying birds and their territories, which includes bird marking and tracking.
- Studying wildlife populations with genetic methods and what happens with species due to anthropogenic influences.
- Phenology in plants, which is the study of life cycles of plants.
- Environmental ecological measurements embedded in a social ecological context.
- Bioeconomy, as exemplified by improving crops and other agricultural practices.
- Microbiology, i.e., the study of microorganisms.
- Ecotoxicology, which is the study of how toxic chemicals affect organisms.
- Hydrology, ocean monitoring and marine biology.

- Wind directions and speeds.
- Geology research.

Interviewees mentioned that they were involved in various ways, popularizing science. Crowdfunding actions were brought up in connection to following and naming GPS equipped birds with the intent of gaining traction among artists and politicians. In one instance, there were already existing collaborations, bringing together artists mainly from the digital sciences, utilizing visualizations pertaining to environmental change as part of their performances. As one researcher put it:

> […] let's face it, [people] don't give a **** about our Excel sheets and graphs, because they are boring. So, we need to find other ways to get people interested. (P6)

Furthermore, informants mentioned following GPS equipped birds in the context of educational initiatives, e.g., children learning about birds, and additionally being able to track them. In that way children could study the whereabouts and the patterns of the birds, as well as being able to take an active part in environmental research.

3.3.Obstacles

COVID-19 has affected everyone in various ways. In general, this subject was broached as soon as the interviews commenced, while building rapport. Many informants were working from home, and several informants had not been able to go to their stations due to travel restrictions.

A common theme that emerged during the interviews was obstacles encountered with regards to monitoring and research. The most pressing issue was that of funding. No matter the importance, longevity or even prestige of the research station, funding always crept up as a major hinderance. In fact, funding was mentioned as a constant struggle, even in cases when the research station in question would be used as promotion for climate change research. Moreover, in relation to funding, another factor that was perceived as a potential problem by the informants, was that of their work being steered towards the desired outcome of the donor.

Monotonous work, for instance manually classifying animals in photographs, counting birds, plants, or samples is very time-consuming and tedious, and not a very good use of already lacking resources. These resources could more efficiently be utilized by performing analysis and research. As one interviewee put it:

> One of our biologists […], she sits there every single day, by a microscope counting with this little petri dish, samples in it, and she's analyzing just manually. She has a counter and I hear click, click, click, click […]. It's very time-consuming and probably very tiring, and it's a lot for one person to do. (P11)

There was the issue of collecting data or sensitive instruments needing monitoring and calibration, in many cases, in a high-risk environment and usually a large land area. It was also mentioned that there were sometimes issues with power supply.

At the location of one station, there was a ban on the use of Wi-Fi and Bluetooth, due to there being a radio telescope sensitive to such frequencies. This makes it more difficult to conduct science and disables, among other things, the use of drones. It is possible to apply for permission, but that is often a futile process.

The following quote also indicates one particular issue one station faced in connection with tourism, affecting their measurements:

> If we have a group of smokers up there smoking in front of the door, we can easily see the peak in our aerosol concentration because then we have, for example, the particle concentration rise up to 1000 to 10,000 particles more than usual. (P7)

3.4.Data Collected

Having an agnostic approach to data collected with regards to environmental changes, both scientists and engineers can find a common ground, albeit utilize data in various ways and doing different kinds of analyses. It is important to note that not all data collected today may serve a useful role in present climate research. However, data presently collected might provide beneficial for future research. Furthermore, a data set in itself can be very valuable, either monetary or for the sake of research, generally depending on its usefulness, as in being properly curated, labelled and easily accessible.

Most, if not all stations have access to time series of temperature, air pressure and in some cases radiation data. Furthermore, it was also indicated that there are measurements of wind speeds and soil temperatures. In a similar vein, informants also mentioned field measurements of snow depth, as well as permafrost measurements and glacier data, with e.g., temperature sensors in bore holes.

Image data were commonly brought up with regards to both camera traps, i.e., motion sensitive cameras, where both animals and sometimes people appeared, as well as conventional cameras. These data were also used for the purpose of studying phenology of plants, as well as studying the arrival and departures of various bird species. Some stations were using drones, usually with the aim of monitoring glaciers and looking at vegetation indices. There were also some usage of satellite data and synthetic aperture radar to study geological hazards.

Moreover, samples were taken in a variety of areas such as:

- Water (rain and snow), as well as air samples.
- Blood samples from birds.
- Samples of tissue or feces of animals for the purpose of genetic analysis, to find out more about the population structure of wild animals, and how they are distributed across the landscape.

Additionally, the topic of digitizing old data, was an emerging theme, and as one informant put it:

> We have some old data, that would be nice if someone would have time [to digitize]. There are some projects that we know what has been done, we know the sites and the station is old enough [...] so it would be nice to compare this with the present situation. (P14)

All informants were positive about sharing data for the benefit of science. However, it was also brought up that there was fierce competition among researchers with regards to publications and being the primary author. For instance, conducting very hard field work under strenuous circumstances, no less because publications equal more funding, only to end up as one author among many, would make some scientists hesitant to share their hard-earned data. On the other side of the coin, there was the dilemma of the scientist, having conducted a large number of measurements over decades, and stored their data on their laptop. In the worst-case scenario, the hard drive crashed, or the scientist died, and the data would disappear with them.

One interviewee stated that in their domain, there were vast amounts of data collected, albeit not stored for public use in a common repository. For instance, the informant considered it interesting to obtain existing data with regards to walking and scooter trails, both human and animal activity. However, there were both political and commercial interests not to share these data.

3.5.General Opinions and Curiosity about AI

During the discussions of the workshops held by WP6, it became apparent that many researchers were new to AI, and while curious, had little knowledge about it and were unsure how to proceed. Some researchers, however, had realized the potential of AI/ML and were eager to get involved:

> AI is here to stay and is not going away. (P1)

It was interesting to note that for several interviewees, a demonstration of image classification in particular had caught their attention during a WP6 workshop. For more information, see the D6.2. report. Generally, interviewees had a positive view of AI, while being unclear on particulars of what it can and cannot do. For instance, it was expressed that while AI/ML previously seemed like a novelty, it now seemed that it could be practical and helpful. Furthermore, researchers would rather start using AI/ML with a less advanced algorithm that is easier to use than an advanced “killer” algorithm.

Python is currently the most popular programming language used for AI/ML. A main reason for this is that many Python libraries are easy to use, while still being powerful. However, not everyone is able to, or have the time to learn how to program Python. It was stressed that a way to utilize AI/ML without having to program would be preferred. One researcher also mentioned that not everyone had enough experience in mathematics to grasp AI, while others stated that they in general were unsure on how to approach AI practically.

One of the most common themes was how to use AI to facilitate manual work and save valuable hours. As one interviewee put it:

> Humans are so damn expensive! (P1)

As was outlined in Section 3.3, funding is an issue for many stations, and if AI can help save man hours that in turn can be used for more advanced research, many researchers would be eager to incorporate AI/ML in their workflow.

3.6. Usage of AI – Current and Future

Several ideas on how to utilize AI algorithms were put forth, for instance finding patterns and making good use of large amounts of data. However, the most commonly mentioned application was to use automatic classification of image data, e.g., remove empty pictures from the data set, classifying animals as well as drawing conclusions concerning anthropogenic influences of wildlife. Detecting invasive species was mentioned in the same context.

Furthermore, AI as an aid in phenology monitoring, both tracking cycles and e.g., automatically counting flowers was brought up. There was also an interest in identifying bird song and in connection help counting birds. Further, collecting, merging and analyzing data from diverse data sets and drawing inferences from these, e.g., how weather data affect reindeer grazing quality. Similarly, a potential AI application was to flag when monitoring data were suspected of being incorrect.

With regards to social studies, it was expressed that it would be a great help if AI/ML could aid in extracting relevant answers from surveys, also, in several languages, using automatic translations. As one interviewee put it:

> I'm developing a survey questionnaire for the public to understand ecosystem services better. So basically, benefits of natural systems for humans and how they're perceiving certain species […] And let's say I get a lot of answers. It could be interesting to extract answers of the survey, probably with a particular algorithm so I can categorize this in a good way without going through answers manually again. (P10)

In this vein, it was suggested that AI could be used to collect more data for minority languages, as well as in studying how humans organize and communicate, including tacit knowledge.

Drones and how to use them in connection with AI was a popular topic; for instance, using automatic recognition in quantifying riverbank erosion. In one case, a researcher was investigating what technical specifications of drones and camera equipment would be needed to use DL applied on glaciers or looking at hot spots regarding gas emissions. Similarly, SAR data was brought up for detection of geo-events, as well as using AI with remote sensors, for automatic image analysis on site.

4. A Quantitative Study Based on Questionnaire Responses

4.1.Introduction

During the two mini-workshops that were held on June 10th, 2020, participants were given a link and encouraged to fill in a questionnaire. Also, during the main workshop on September the 24th, time was allotted for the attendees to fill in the same questionnaire. The questionnaire was also mailed to all INTERACT III participants. The full questionnaire can be found in Appendix B.

The questionnaire addresses themes similar to the interviews in Section 3, such as:

- The projects and tasks currently being worked on.
- Whether respondents currently use AI in research projects, plans to use AI in the future and whether the respondents know of any projects or ideas (even crazy ones) where AI could be interesting to use.
- What data is collected.
- Data that could benefit from using AI, and whether publicly available.
- If the respondents would be willing to conduct a pilot small research project using AI.

The aim has been to analyze what inquiries and needs the researchers and station managers may have pertaining to AI/ML, and how it can tie into their research. Furthermore, an investigation of what data researchers have access to has been conducted. This will lay the foundation for D6.3 where a demonstration on AI/ML techniques on researchers' data will be given, as well as D6.4, where a future strategy will be outlined for utilization of AI/ML in Arctic research.

25 responses were received, with data points being extrapolated from relevant passages of conducted interviews, where applicable, to gain more data. The total number of responses ended up being 30. Note that the results presented in this section only take into consideration the answers from the respondents and have not been validated against external information (apart from, where applicable, the interviews in Section 3).

4.2.Results and Visualization of Data

Figure 10 gives an overall view of the results from the questionnaire. It illustrates the percentage of respondents as a function of the questions answered in the main topics:

- Do you currently use AI?
- Do you plan to use AI in the future?
- Interesting project applications?
- How can AI help you?
- What kind of data do you have access to?

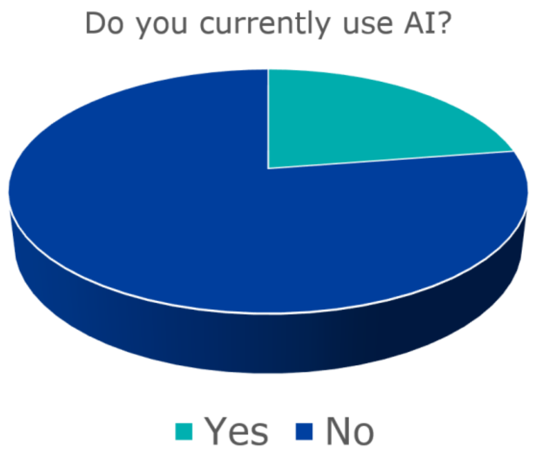

Figure 8: Current usage of AI.

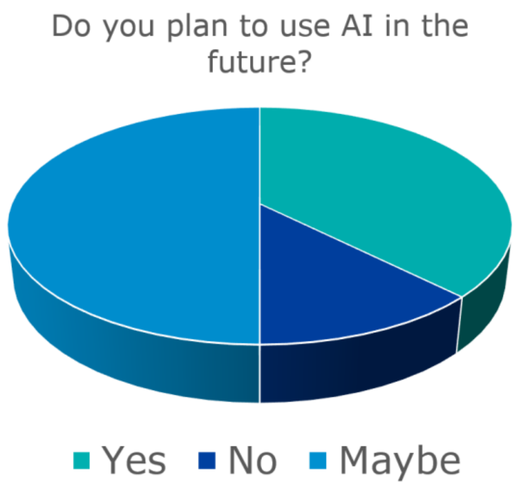

Figure 9: Future usage of AI.

Figure 8 shows that fewer than 25% of respondents currently use AI. Figure 9 on the other hand illustrates that respondents in general are positive towards using AI in the future. Although there are examples of respondents not planning to use AI, it is clear from the answers from subsequent questions, that this could be interpreted as previous to the workshops, the respondents did not plan to use AI, but changed their mind during or after one of the workshops.

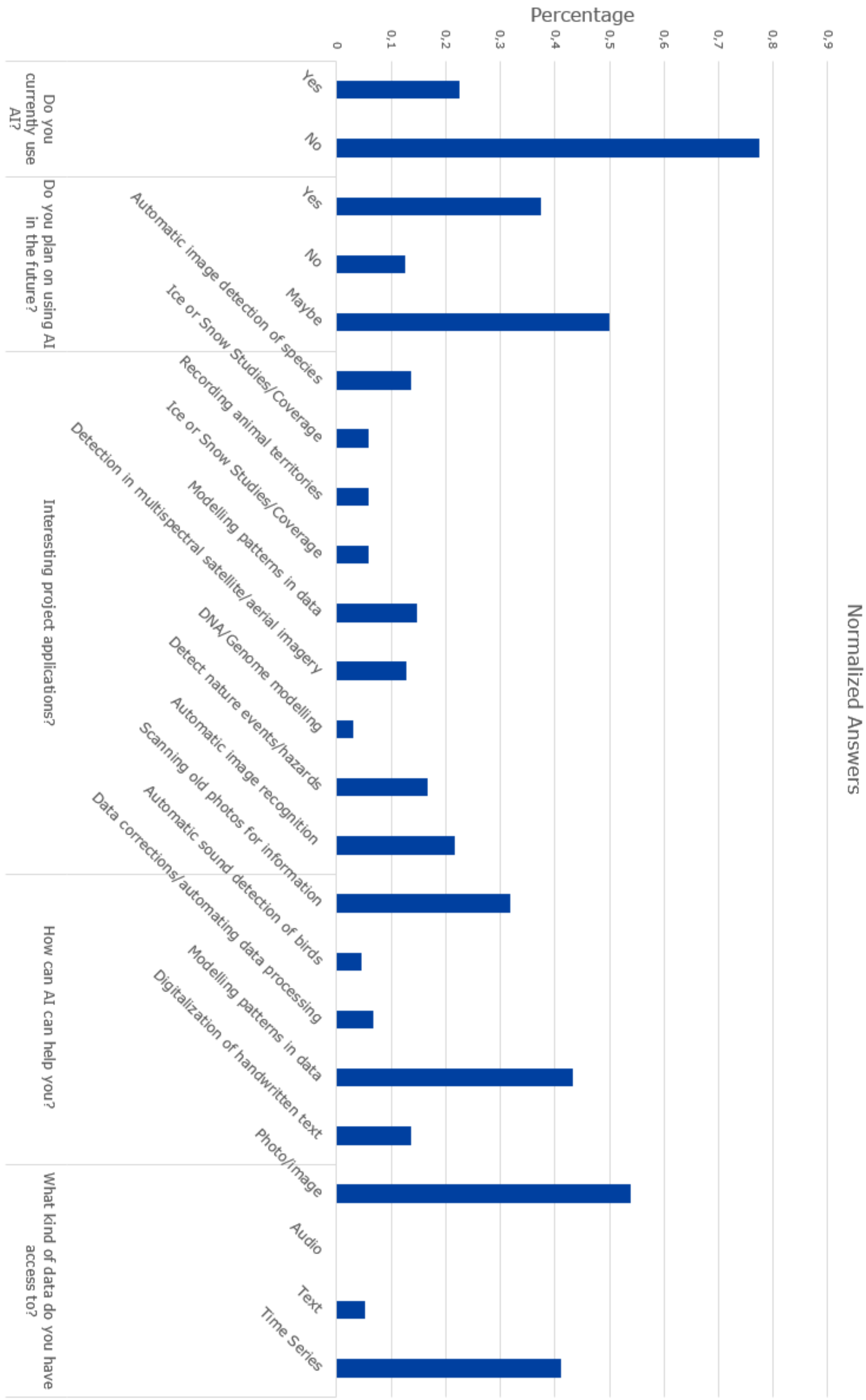

Figure 10: Normalized answers, with respect to number of respondents, for areas pertaining to AI, and usage thereof.

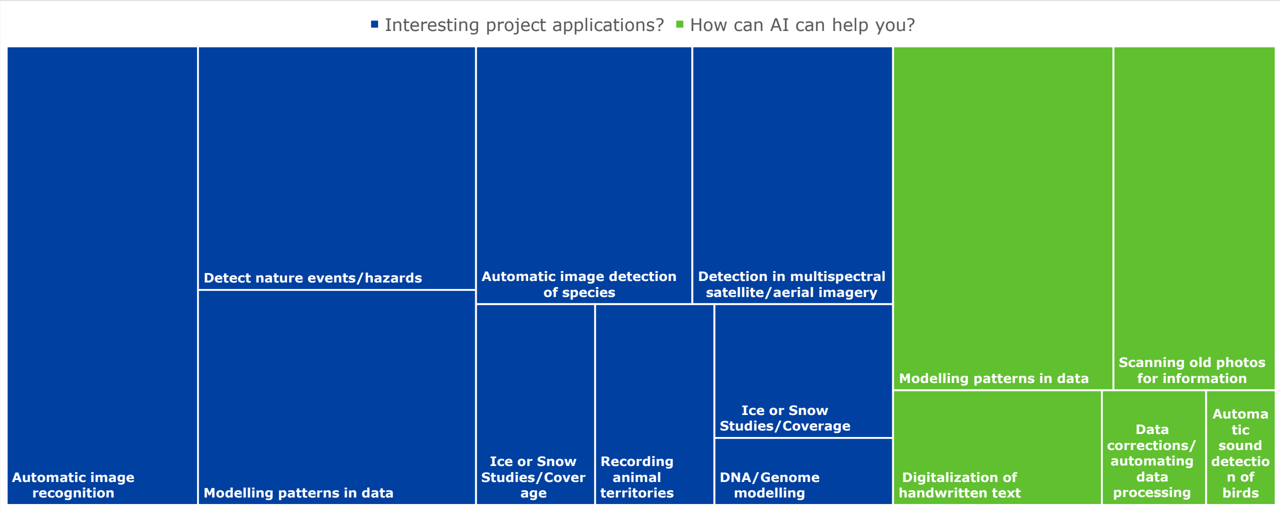

Figure 11: (a) Interesting project applications. (b) How can AI help you?

Figure 11 shows in blue what respondents consider to be interesting areas of applying AI, and in green how AI could be of help. The size of the various rectangles indicates how many respondents have replied, i.e., the larger the rectangle, the higher the number of respondents. Areas with fewer than three responses are not shown. The charts have differentiated between answers from how AI can practically help the researchers and what they consider to be interesting project applications. Several respondents regard automatic image detection and recognition, modelling patterns in data, and detection of natural events or hazards to be interesting project applications. Regarding how AI could be of help, the answers reflect manual tasks, like digitizing handwritten texts, scanning photos and modelling of patterns in data.

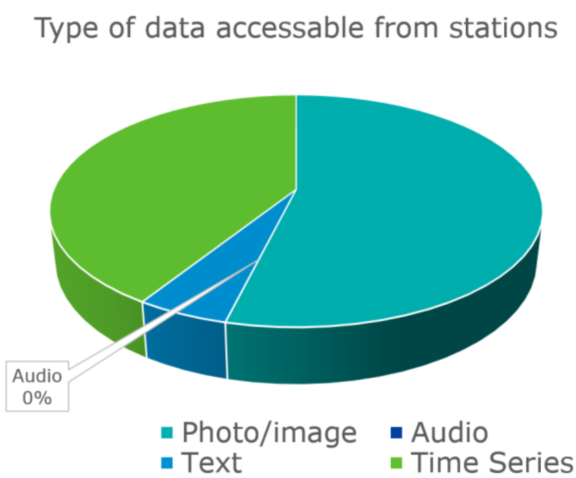

Figure 12: Categories of available data.

Figure 12 shows that the types of data accessible from stations are mostly photo or image data and time series data. None of the respondents or interviewees have answered that they have sound data accessible. As for sharing data, in the case of the data being accessible, a large majority of respondents were willing to share the data with other researchers.

If the data came from external sources, these would mostly include meteorological data, camera trap data and time series from other research projects.

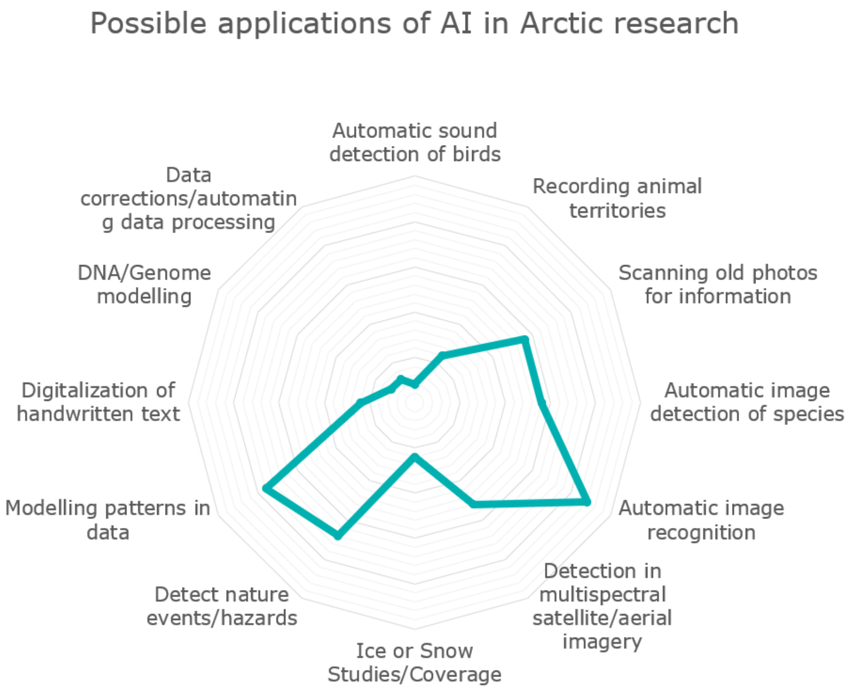

Figure 13: Possible applications of AI in Arctic Research.

Figure 13 illustrates the possible applications of AI in Arctic research, with the general trend of the answers with the largest amount being farthest from the origin. The order of the categories are clustered in similarity of domains, techniques and applications that use similar approaches.

Figure 14 uses a cumulative chart to illustrate the same question, i.e., possible applications of AI in Arctic research. This chart shows the number of responses, as a percentage, in decreasing order, and the cumulative distribution of responses. After “Detection in multispectral satellite/aerial imagery”, there is a clear decrease in interest, indicating that 50% or more of the respondents are interested in topics to the left of this.

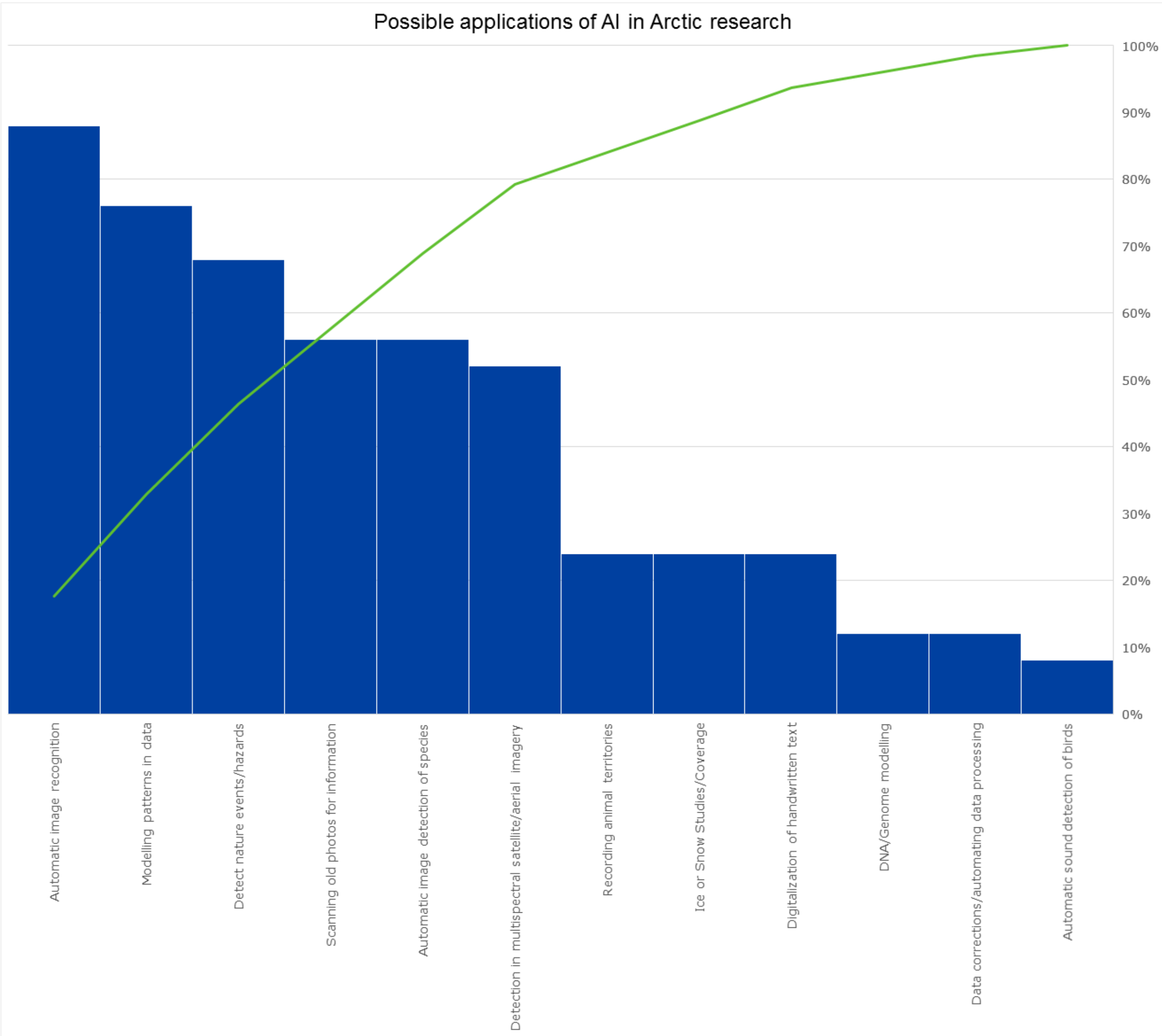

Figure 14: Possible applications of AI in Arctic Research.

5. Analysis and Discussion

As part of D6.1, this report tries to answer the questions posed in Section 1.2, i.e.:

- What inquiries and needs do researchers and station managers have?
- What research is conducted?
- What data are collected?
- What obstacles are encountered by the researchers?

The task was approached by combining a qualitative interview study and a quantitative study based on questionnaire responses. This choice by necessity introduces certain sources of potential measurement bias or error. Some questions in the questionnaire were left unresponded to, the reasons for which are not known. A potential reason might be that respondents were too uncertain regarding the field and were not able to provide concrete enough answers and thus preferred to leave the fields blank. This is corroborated by the fact that several respondents both in the questionnaire and the interviews indicated interest in AI but were unable to specify in what way AI could assist them in their work.

A second source of uncertainty is the misinterpretation of the respondents' answers. As they were from a wide range of cultural and research backgrounds, misinterpretations and misunderstandings in communications are possible. Furthermore, several respondents as well as participants in D6.1 do not have English as their native language, which may compound the uncertainty. This also affects the quantitative analysis, e.g., by the bias in choosing the ordering in Figure 13.

The results of the quantitative study clearly shows that the majority of the respondents have an interest in image recognition and detection, with applications ranging from satellite imagery to scanning old photos for salient information. A second clear application of interest is modeling of patterns in various kinds of data, e.g. seismographic data and DNA and genome modeling. Most of the suggested applications can be mapped into these two categories. The primary outliers are the recording of animal territories, sound detection of birds, and digitalization of hand written text. The latter is particularly suited for NLP.

However, to be able to further assist the station managers, additional work is needed to identify their specific needs and to obtain more specific information about the kinds of data they can make available to the project. To be able to more efficiently use the resources of WP6, generalizable methods of, e.g., image recognition and detection are to be preferred to application specific methods. As many of the respondents clearly indicate that they would benefit from automating certain manual tasks, this may prove to be quite challenging. An ameliorating circumstance in this work is the enthusiasm of the respondents, which will make further dialogues easier.

6. Conclusions

Field work and subsequent analysis and documentation of said work, performed by Arctic researchers, are in many cases time-consuming, with tasks such as:

- Looking at photographs and identifying and classifying plants in various stages.
- Identifying and classifying animals and their species, as well as counting them.
- Listening to bird song and identifying the bird species.

From examining questionnaire responses, analyzing discussions held during workshops, as well as gaining insights from conducted interviews, a way forward has emerged. That is for WP6 to employ AI/ML techniques by helping to reduce manual work for researchers.

This can be done in two main ways:

- Simplifying application of ML by providing “ML as a service”, i.e., providing a simple method for accessing, using and managing ML functionality, perhaps as a cloud enabled service.
- Providing support to researchers in applying ML to their data.

6.1.Future Work

D6.1 will lay the foundation for D6.3, where a demonstration and a presentation of a pilot project in which AI and ML techniques are applied on researchers’ data. A pilot project is a unique yet informative way to introduce the utilization of new technology to the research community. The results of this preliminary study will form the basis of what applications of AI and ML to be utilized for the demonstration. D6.4 will present a report on future strategy and planning for the areas of how AI and ML can be applied on Arctic research.

Appendix A – Template Interview Questions

General

- Can you tell us a little bit about yourself and the research you have been involved in?
- Can you tell us about your work at […]?
- What projects/tasks are you currently working on?
- What does a typical workday look like for you?
- How long have you worked on the various projects?
- What are your goals or objectives with the projects?
- Who are the stakeholders?
- What challenges are you are facing?

Input and Data in General

- You said that […]. Can you tell us a little bit about your work process with regards this?
    - What is the task to be accomplished?
    - Can you give a description on how the data are analyzed currently?
    - Are there any particular challenges, or other factors to be taken into consideration to perform this task?
    - How do you structure your data?
        - Are the data labelled?
    - Do you have criteria that you use for inclusion or exclusion of data into the data set?
        - If so, what are those?
    - How do you store your data?
- Do you collect other kinds of data?
- Is there any kind of data you would like to collect that you do not collect currently?

Previous Knowledge about AI

- What are your general thoughts about AI?
- If applicable: You said that you had started using AI and machine learning. Can you elaborate a little bit about this and how it's progressing?

Finding a Common Ground

- How do you think AI can help you?
- What are your thoughts about the possible benefits of AI in your projects/tasks/research?
- Are there any areas of ML, or algorithms that you find particularly interesting that you are looking into?

Questions or Comments to Us

- Do you have any inquiries or questions to us?
- Any other comments?

Appendix B – Questionnaire

Artificial Intelligence in Arctic Research

As part of our deliverables, WP6 would be grateful if you could fill in this form on your ideas of how artificial intelligence (AI) and machine learning (ML) can be used in Arctic research.

*Required

1. Email address *

General Information

2. What is your name and affiliation?

3. What is your current primary position (e.g., station manager, researcher at a university, researcher on site, student etcetera)?

4. What are you working on currently?

AI Usage

5. Do you currently use AI in your research projects?

- ☐ Yes
- ☐ No

6. If affirmative, what do you use AI for more exactly? Also, what kind of algorithms do you use (e.g., support vector machines, convolutional neural networks etcetera)?

AI Potential

7. Do you plan on using AI in your projects in the future?

- ☐ Yes
- ☐ No
- ☐ Maybe

8. Do you know of any projects where AI could be interesting to use, either within your group or in other groups?

9. Do you have any ideas that can be pursued using AI, even if they are pure fantasy? Even crazy answers are welcome!

10. How do you think AI can help you? Some examples could be automating data processing or finding hidden patterns in the data.

AI Data

11. Do you have access to data that can benefit from using AI, even if you are not using AI now or in the near future?

- ☐ Yes
- ☐ No

12. If yes, can you specify what kind of data you are using (e.g., field measurements, photographs etcetera)?

13. If yes, have your group collected the data or did it come from an external source?

☐ Collected
☐ External
Other: ☐ ______________________________

14. If yes, have these data been published by either your group or some other group (e.g., Copernicus)? If so, can you specify the dataset, preferably with a URL? Would you be willing to share your data within the Interact network if it is not public?

____

____

____

____

____

Pilot Studies

15. Would you be willing to conduct a pilot small research project using AI in your area of expertise?

☐ Yes
☐ No
☐ Maybe

16. If affirmative, can you be more specific?

Other

17. Do you have any other comments?

Bibliography

[1] National Snow and Ice Data Center, “All About Arctic Climatology and Meteorology,” 2020. https://nsidc.org/cryosphere/arctic-meteorology/index.html (accessed Feb. 05, 2021).

[2] IPCC, “IPCC Fifth Assessment Synthesis Report-Climate Change 2014 Synthesis Report,” *IPCC Fifth Assess. Synth. Report-Climate Chang. 2014 Synth. Rep.*, 2014.

[3] S. J. Hassol, *Impacts of a Warming Arctic*. Cambridge University Press, 2004.

[4] J. D. Ford, “Indigenous health and climate change,” *Am. J. Public Health*, vol. 102, no. 7, pp. 1260–1266, 2012, doi: 10.2105/AJPH.2012.300752.

[5] M. Chui, “Artificial intelligence the next digital frontier,” *McKinsey Co. Glob. Inst.*, vol. 47, pp. 3–6, 2017.

[6] International Network for Terrestrial Research and Monitoring in the Arctic, “Research infrastructures for terrestrial research in the Arctic, Technical Annex 1-3,” *H2020-INFRAIA-2018-2020*, 2020.

[7] G. Assembly, “Resolution adopted by the General Assembly on 19 September 2016,” A/RES/71/1, 3 October 2016 (The New York Declaration), 2015.

[8] “Societal Challenges and TA - INTERACT.” https://eu-interact.org/accessing-the-arctic/societal_challenges_and_ta/ (accessed Feb. 11, 2021).

[9] M. D. C. Tongco, “Purposive sampling as a tool for informant selection,” *Ethnobot. Res. Appl.*, vol. 5, pp. 147–158, 2007.

[10] M. Q. Patton and R. E. C. M. Fund, *Qualitative Research & Evaluation Methods*. SAGE Publications, 2002.

[11] M. Wooldridge, *A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going*. Flatiron Books, 2021.

[12] A. Ng, “AI for everyone. Coursera.,” 2019. https://www.coursera.org/lecture/ai-for-everyone/week-1-introduction-SRwLN (accessed Feb. 07, 2020).

[13] “Types of Machine Learning Algorithms You Should Know | by David Fumo | Towards Data Science.” https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861 (accessed Feb. 13, 2021).

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

[15] “What Is Deep Learning? | How It Works, Techniques & Applications - MATLAB & Simulink.” https://se.mathworks.com/discovery/deep-learning.html (accessed Feb. 13, 2021).

[16] R. Rojas, *Neural networks: a systematic introduction*. Springer Science & Business Media, 2013.

[17] D. C. C. Cires¸an, U. Meier, J. Masci, L. M. Gambardella, and J. ¨ Urgen Schmidhuber, “Flexible, High Performance Convolutional Neural Networks for Image Classification.”

[18] T. Melcherson, “Image Augmentation to Create Lower Quality Images for Training a YOLOv4 Object Detection Model,” Signals and Systems, Department of Electrical Engineering, Technology, Disciplinary Domain of Science and Technology, Uppsala University, 2020.

[19] S. Li, “Understanding Word2vec Embedding in Practice |Towards Data Science,” 2019. https://towardsdatascience.com/understanding-word2vec-embedding-in-practice-3e9b8985953 (accessed Feb. 08, 2021).

[20] J. McCurry, “South Korean AI chatbot pulled from Facebook after hate speech towards minorities,” *The Guardian*, 2021. https://www.theguardian.com/world/2021/jan/14/time-to-

properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook (accessed Feb. 08, 2021).
[21] "Gensim: Topic modelling for humans." https://radimrehurek.com/gensim/ (accessed Feb. 08, 2021).