Edge Intelligence Taxonomy

Brian Siegel

2416912

Seminararbeit

Lehrstuhl für Wirtschaftsinformatik
und Systementwicklung
Universität Würzburg

Betreuer: Prof. Dr. Frédéric Thiesse

Würzburg, den 30.10.2020

Bearbeitungszeit: 06.08.2020 - 30.10.2020

Contents

Zusammenfassung

Eines der größten Probleme bei Edge Intelligence sind die verschiedenen Definitionen, die es für diesen Begriff gibt. Aufgrund der Verbreitung von IoT, der Ausweitung des 5G-Netzes, des Aufkommens von Edge Computing und der enormen Fortschritte in der künstlichen Intelligenz boomt die Forschung zum Thema EI. Über die Grenzen des Edge-Netzwerks sind sich die Wissenschaftler jedoch immer noch uneins. Offensichtlich ist diese unklare Situation um EI der Hauptauslöser für die Debatte. Während diese Unsicherheit anhält, wächst die Edge-Plattform rasch, und der Bedarf an künstlich bereicherten Anwendungen im Edge-Netzwerk wächst. Die wenigen Literaturrecherchen, die es heute gibt, konzentrieren sich meist auf ein breites Spektrum des Themas. Viele Beiträge beschreiben EI in Koexistenz und Kooperation mit der Cloud. Aufgrund von Sicherheitsgründen und Latenzproblemen im Zusammenhang mit der Cloud sowie von Breitbandproblemen und dem zunehmenden Datenverkehr am Rande des Internets muss die Perspektive, EI vollständig in der Edge-Umgebung zu realisieren, weiter untersucht werden. Aus diesem Grund werden in diesem Dokument Design-Optionen untersucht, welche ermöglichen, dass EI lediglich auf Edge-Knoten läuft. Zunächst wird der theoretische Hintergrund von Edge Intelligence analysiert und eine Arbeitsdefinition für EI bereitgestellt. Basierend auf der Brocke et al. (2009) Methode, wird eine repräsentative Literaturrecherche durchgeführt. Die Literaturanalyse zeigt einen verstärkten Forschungstrend in den Jahren 2019 und insbesondere 2020. Die generierte Datenbasis an Artikeln wird dann genutzt, um eine Taxonomie nach der Nickerson, Varshney, and Muntermann (2013) zu erstellen und zu verfeinern. Die Auswertung zeigt, dass die Edge-Device-Kollaboration und energieeffizientes Rechnen die wichtigsten Forschungsbereiche in dieser Perspektive sind. Der Autor hofft, dass diese Seminararbeit den Forschern eine nützliche Orientierungshilfe bietet und weitere Forschungsinteressen im Bereich der Edge-Intelligenz inspirieren kann.

Abstract

One of the biggest problems with edge intelligence are the various definitions existing for the term. Due to the proliferation of IoT, the expansion of the 5G network, the emerging of edge computing and tremendous progress in artificial intelligence, research on the topic of EI is booming. However, scholars still disagree on the boundaries of the edge network. Obviously, this unclear situation around EI is the main trigger for the debate. While this uncertainty prevails, the edge platform is growing rapidly and the need for artificially enhanced applications on the edge network is growing. The small number of literature researches that exist today, mostly focus on a broad scope of the topic. Many papers describe EI in coexistence and cooperation with the cloud. Due to security reasons and latency issues in the context of the cloud, as well as bandwidth problems and the increasing data traffic on the edge, the perspective of accomplishing EI entirely in the edge environment needs to be further explored. For this reason, in this paper, design options for EI to run merely on edge nodes are studied. At first, the theoretical background of edge intelligence is analysed and a working definition for EI is provided. Based on the Brocke et al. (2009) method, a literature review is carried. A literature analysis shows an increased research trend in the years 2019 and especially 2020. The generated database of papers is then utilized to create and refine a taxonomy adopted by Nickerson, Varshney, and Muntermann (2013). The examination shows, that edge-device collaboration and energy-efficient computation are the most important research areas in this perspective. The author hopes that this paper can provide researchers meaningful guidance and inspire further research interests on edge intelligence.

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence

DNN Deep Neural Network

EI Edge Intelligence

IoT Internet of Things

ML Machine Learning

NIST National Institute of Standards and Technology

V2X Vehicle-to-everything

1 Introduction

The deployment of neural networks on edge computing has several use cases. Although, the challenge of designing novel models, which run on the limited capacities of the edge network, is known, smart devices can empower new business ventures. For instance, connected and autonomous vehicles can make use of EI. The EI-models deployed on the cars need to perform object detection, lane stabilization, decision making and localisation. On-board cameras assist the EI-system with computer vision. “Smart Home” devices are another example. A home leveraging the IoT domain and EI can have two effects. First, by interconnecting sensors, video surveillance and safety systems, EI can guarantee security, privacy and comfort of homes. Secondly the operation, maintenance and installation of new devices can be managed and simplified through EI. The algorithms can monitor and identify risk for home owners. Other EI-application scenarios include “Smart and Connected Health”, “Mobility management” and “Energy efficiency aspects” (X. Zhang et al., 2019; Chang, n.d.).

Since reduced latency, increased security, reduced bandwidth and lower cost are all benefits of the edge network, this studies takes the perspective of accomplishing EI entirely in the edge environment (Lanner, n.d.). The increased data traffic on the edge is another reason for this proposition (Forbes, 2019). The later proposed taxonomy and literature review, highlight the importance of designing EI to run entirely on the edge platform.

The remainder of this paper is organized as follows. Section 2 describes the theoretical background od edge intelligence, such as IoT and edge computing. Here a working definition for edge intelligence is presented. Section 3 describes my literature search process, relevance filtering, literature analysis and a concept matrix. Section 4 uses the generated database to create and refine taxonomy. Section 5 discusses the impacts of the findings and section 6 concludes the study.

2 Theoretical Background

This section covers the basic topics and concepts of this paper. These include edge computing, cloud computing and edge intelligence. Furthermore a working definition for the chapters to follow will be provided.

2.1 The Edge of the Internet

When trying to understand edge intelligence the first step is defining the term “edge”. However, as with many buzzwords in the IT, vendors tend to have different definitions. The edge can be many things and some experts even argue that computer endpoints - where the virtual meets the real world - represent the edge (Dahad, 2020). Most commonly, the edge network is referred to as the counterpart to the cloud, which is usually a centralised data centre. In contrast to the cloud, edge nodes are heterogeneous platforms, such as mobile and Internet of Things (IoT) devices, edge servers, local area networks, base stations or micro data centres among other things (IEC, 2017; Veena et al., 2015; Yu et al., 2017). Others argue, that the distinguishing feature of the edge is its spatial proximity to the origin of the data creation point (Hassan et al., 2018). Though HPE (n.d.) argues, the main characteristic of an edge node is its decentralised aspect.

A key driver for expansion of the edge ecosystem is the emerging of mobile computing and the IoT environment. To date the IoT technology has become fairly mature and its devices are often interconnected. The IoT domain now ranges from “Industry 4.0” to “Smart Cities”, or “Home Automation” and many other areas (Hassija et al., 2019). Billions of IoT devices produce massive data amounts on the edge of the network and are connected to the internet (Zhou et al., 2019). According to the *Cisco Global Cloud Index*[1] by 2019, the projected data produced by people, machines and things amounted to 507 zettabytes per year. By that time, the data traffic to, from and within data centres only accounts for 10.4 ZBs (Cisco, 2015). Consequently, the the internet will not be able to cope with the transfer capacities if directed merely to the cloud (W. Shi, Cao, et al., 2016).

Another relevant trend is, that edge devices with AI-chips are becoming more ubiquitous. A recent study published by Deloitte estimated, that in 2020 over 750 million AI-chips for edge devices will be sold, which will be used for machine learning tasks. At the same time Deloitte also reports, that the consumer market for edge devices is still increasing. Cisco Internet Business Solutions Group predicts that by 2020, over 50 billion devices will be connected to the internet (Bilal et al., 2018). Hence, with the spread of AI-chips in consumer devices, the increasing number of edge device sales and the roll-out of 5G network the authors argue, that not only the consumers will benefit from new AI-features (improved Face-ID smartphone unlocking; quality conversations with voice assistants; high quality photography in adverse environments), but also existing markets, such as manufacturing, construction, logistics, agriculture, and energy will be expanded by smart edge machines (Stewart et al., 2019).

This industry disruption spurs on the importance of the edge platform. However, the edge model lacks key requirements, being real-time decision making, mobile connectivity in dead spot areas and precise big data analysis (IEC, 2017). Considering its disadvantages, this disruptive technology is not a replacement of centralised resources, but rather an extension.

2.2 Edge Computing vs. Cloud Computing

At the heart of the conversation between edge versus cloud computing lie the concepts of centralisation and decentralisation. In the 1960s the alternation began and since then

[1] Link to white paper: `https://www.cisco.com/c/dam/m/en_us/service-provider/ciscoknowledgenetwork/files/547_11_10-15-DocumentsCisco_GCI_Deck_2014-2019_for_CKN__10NOV2015_.pdf`

each approach has had its peaks. Today, cloud computing widely dominates the discussion, but edge computing is the latest technology introduced in favour of decentralisation (Satyanarayanan, 2017). The ongoing miniaturisation of processing units, as well as the increasing storage and computing capacities acted as enablers for more powerful edge devices, which are available today (Garcia Lopez et al., 2015). Since edge computing opens the way for edge intelligence, it is worth comparing the differences to its prominent partner. Table 1 shows a quick summary of the two approaches, highlighting some of the characteristics.

Table 1: Key differences between cloud computing and edge computing model (Sharma and X. Wang, 2017; Naha et al., 2018; Ahmed et al., 2017; Belyi, 2020)

Features	**Cloud Computing**	**Edge Computing**
Computational capacity	high	low to medium
Model	centralised architecture	distributed/decentralised architecture
Node type	data centres	mobile, IoT and end-user devices, sensors, edge servers, base stations
Number of nodes	thousands	billions
Idea	to access software and resources over web, no need for infrastructure locally	scalable infrastructure, where computing happens on node itself
Use Case Szenario	Cloud computing is a perfect match for companies, that require large amounts of data storage and need scalable and cost-effective processing/hosting providers, e.g. data warehousing	Edge computing suits businesses that provide services with ultra low latency and require context awareness, e.g. connected vehicles

Cloud computing Cloud computing has several application fields. Depending on the operating model (e.g. IaaS, PaaS, SaaS) the degrees to which external infrastructure is used vary. Additionally, the cloud is often required for outsourcing storage or performing resource-rich analytical or processing tasks (De Donno, Tange, and Dragoni, 2019). The National Institute of Standards and Technology (NIST) definition states, *"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider* *interaction."* (Mell, Grance, et al., 2011). Furthermore, the cloud model has many benefits. Companies tend to utilize the cloud in order to quicken self-management and ensure data protection. Moreover, large data centres facilitate economies of scale, by lowering the costs of system administration and operation. This allows corporations focus investments away from building on-premise cloud centres. However, cloud computing faces numerous challenges. For instance, cloud computing is confronted with single point of failure, reachability problems and latency issues related to Wide Area Networks (WANs) (IEC, 2017).

Edge computing As explained earlier, the term "edge" is still a widely discussed research topic and various definitions exist for its scope and boundaries. Obviously, a definition for edge computing is dependent on a previously defined "edge" and therefore a working definition for this seminar paper will be introduced. Edge computing uses the processing capabilities of the edge network layer rather than sending all requests directly to the cloud (Varghese et al., 2016). In the IoT-era, billions of intelligent devices are active at the edge of the network and communicate with central systems or with each other in peer-to-peer connections. Due to the significant computation and delivery delay in the cloud, the edge computing paradigm has arisen. Additionally, massive IT-trends like Augmented Realtiy, IoT and Vehicle-to-everything (V2X) communications have demonstrated the need for real-time responses and context sensitivity to the use case. Edge computing aims to tackle these problems by processing data at the edge of network. (Dolui and Datta, 2017). The motivation for edge computing also lies in the need of mobile applications to be bandwidth-sensitive (fe. filtering processing data, which needs to be sent to a distant data centre), to not drain energy consumption (f.e. limited battery of smart phones) and to provide smart distributed systems (fe. analytical pipelines or offloading techniques) (Premsankar, Di Francesco, and Taleb, 2018). Through edge computing, edge devices assume parts of the computation and storage capacities of the cloud and critical latency issues evolving from WANs can be solved. In addition, the edge network permits mobility as a result of rich availability and a distributed environment (Mach and Becvar, 2017; W. Shi and Dustdar, 2016).

It is worth noting, that in numerous scientific papers, researchers further subset the term computing by introducing fog computing. The fog layer is usually situated between the cloud and the edge network (Rahmani et al., 2017). Iorga et al. (2017) define the fog as a computing power, which in contrast to the edge directly works with the cloud. Its main tasks are networking, storage, control and data handling services. An example for a fog node is an IoT-Gateway, which merges the data from several IoT-nodes and transfers the results to the cloud (Höß, 2018). Besides, the fog is responsible for interoperability between the two other layers (Bonomi et al., 2012). The introduction of fog computing restricts the edge network to smart end-devices and their users. For simplicity reasons, in the context of this paper when referring to the edge computing approach in the following sections, it will encompass the characteristics of the fog layer.

2.3 Edge Intelligence

Besides the proliferation of IoT and the expansion of the 5G network, in recent years tremendous progress in artificial intelligence has been made. AI-based services and applications are booming and with the advances in edge computing and Artificial Intelligence (AI), the combination of both concepts has become a focal research topic. *As established in section 2.1, nowadays edge devices near the users - such as surveillance, monitoring or control systems - generate and consume large parts of the data traffic. At the same time Machine Learning (ML) models - such as TinyML - that require little memory usage, are gaining popularity (Peltonen et al., 2020).

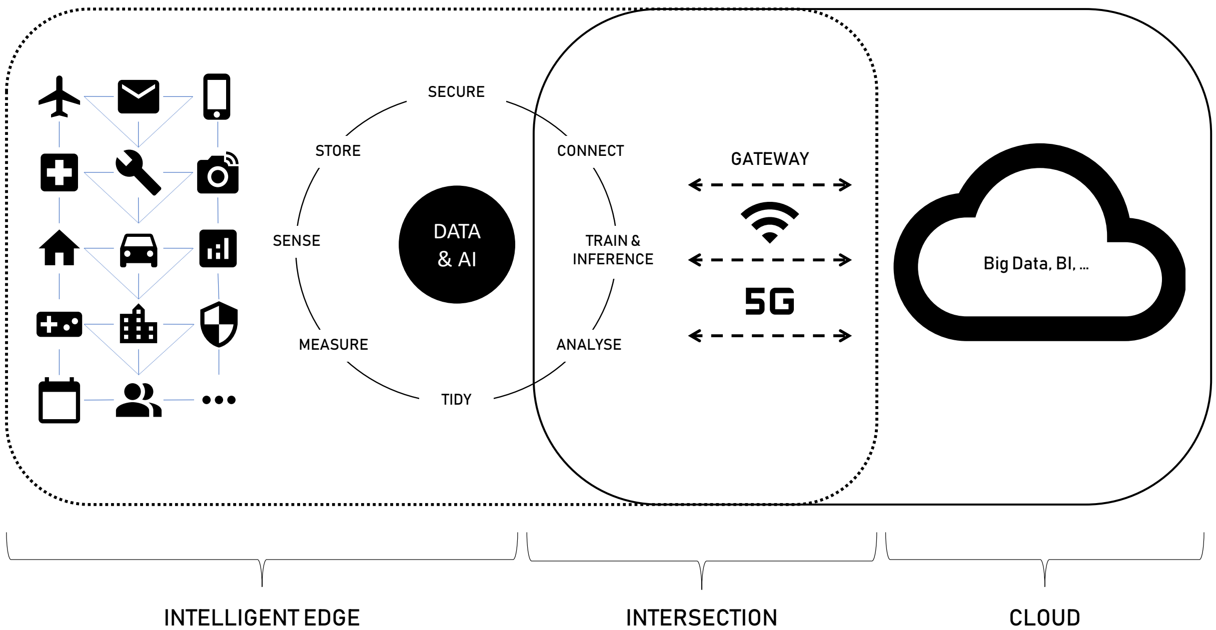

Figure 1: Visualisation of the working defintion for edge intelligence (Beavers, 2017; X. Zhang et al., 2019)

The objective of edge intelligence is to use the proximity of the edge layer for data storage, data analysis, caching, training and inference of deep learning models. Hereby, Edge Intelligence (EI) tries to solve bandwidth, power, latency and privacy issues (D. Xu et al., 2020). To achieve this EI requires deep learning algorithmic design, AI-software, distributed systems and architectures, sensors, networks and many other aspects to provide practical results for the end user (Deng et al., 2020). Figure 1 visualises the working definition of EI. Based on the previously cited papers, this study defines edge intelligence as "*an approach to increase the data analysis and processing capacities of the edge. By sensing, storing, measuring, transforming, connecting and securing data produced on the edge, this model approach filters information close to the data source. Additionally, edge intelligence involves utilising the increasing amount of artificial intelligence on edge nodes. By training ma-*

chine learning models on the cloud, as well as on edge nodes or edge clusters and inferencing close to the origin, new added value services emerge. Concurrently latency and context-sensitivity issues are solved. X. Zhang et al. (2019) attempt to provide a standardised framework for EI, but since the boundaries of the edge and the cloud are vague, an industry-accepted definition is yet to be formulated. Due to obvious limitations of EI, such as computation power, storage deficits and limited battery life, the connection and interoperability between the edge and the cloud is still very important (Plastiras et al., 2018).

3 Documentation of the EI Literature Search Process

As explained earlier, this paper aims to provide a representative literature review of the edge intelligence concept. On this basis, a literature review using a combination of the rigorous literature search frameworks introduced by Brocke et al. (2009) and Cooper (1988), is performed. The process includes [1] defining the scope of the review, [2] the conceptualisation of the topic, [3] scanning databases for topic-related articles, [4] analysing and synthesising the literature and [5] deriving a future research agenda (see section 5).

3.1 Research Scope

The first step in the review process is to define the scope of the literature review. Since edge intelligence is relatively new research topic, the author of this paper hypothesises, that most of the breakthrough research and discussion takes place in journal articles. Therefore, this search is restricted to journal articles available in online databases. The criteria for the literature search will be described in section 3.4. In addition, the search will only query articles written in the English language.

A full overview of the goals and categories of this literature review is provided in figure 2. The *focus* (1) of this paper is to provide insight in to the enabling theories and technologies, research outcomes as well as current and future applications for edge intelligence. Due to the various definitions for edge intelligence and also the changing nature of the computing and the AI paradigm, this study attempts to *integrate* (2) and map the previously peer-reviewed scientific literature, that has dealt with the concept of an intelligent edge. In doing so, the author hopes to better describe and provide a profound understanding of the edge intelligence domain. According to Strike and Posner (1983), the goal to integrate previous studies can be split up into three sub goals: generalisation, resolution of conflicts and linguistic bridges. The literature review at hand, concentrates on generalisation of this new topic. With respect to section 3.2, where the underlying concepts of EI are identified, the *organisation* (3) of this review is built-up in a conceptual structure. The *perspective* (4) chosen is neutral, since no point of view promoted in research is defended. As this study is written in regard to the E-Business Strategies seminar for Information Systems students at the Julius-Maximilians-Universität Würzburg, the main audience (5) this paper addresses are experts in the field of IoT, edge comput-

	CHARACTERISTIC	CATEGORIES			
(1)	focus	research outcomes	research methods	theories	applications
(2)	goal	integration	criticism	central issues	
(3)	organisation	historical	conceptual	methodological	
(4)	perspective	neutral representation	espousal of position		
(5)	audience	specialised scholars	general scholars	practitioners/ politicians	general public
(6)	coverage	exhaustive	exhaustive and selective	representative	central/pivotal

Figure 2: Taxonomy of literature reviews adapted from Cooper (1988) in Brocke et al. (2009)

ing and artificial intelligence. Last but not least, the review includes three of the largest online databases to discover academic literature. However, the author of this paper recognises, that this approach results in a representative *coverage* (6) of the topic.

3.2 Conceptualisation of Edge Intelligence

Based on the findings in section 2.3 and the working definition provided, the next step in the Brocke et al. (2009) framework involves conceptualising the review topic. Particularly for the search strings entered into the academic databases, the development of meta-characteristics is of immense importance. In this study, the author choses to create a concept-map to show the relationships between the subjects. The recorded branches of the concept-map will later be combined and concatenated to generate the search terms. For this reason, figure 3 depicts this paper's conceptualisation of EI. It highlights the aspects of the term, which are, in regard to the working definition, deemed important.

According to the research scope and questions defined in section 3.1, six unique relations to EI are developed. As explained earlier, these heavily rely on the working definition and are enhanced by findings from Zhou et al. (2019). The leaves of the concept-map displayed in light-blue represent terms, which will later be used for the keyword and string search. Due to the expected large number of publications, which contain the search words IoT, deep learning, edge computing and KPI these terms are disregarded in this first step.

Firstly, the research for the theoretical background showed, that the term edge intelligence is often interchangeable with various*synonyms*. Most scientific studies in this area contained the keywords "Intelligent Edge", "Edge Machine Learning" or "Edge Artificial

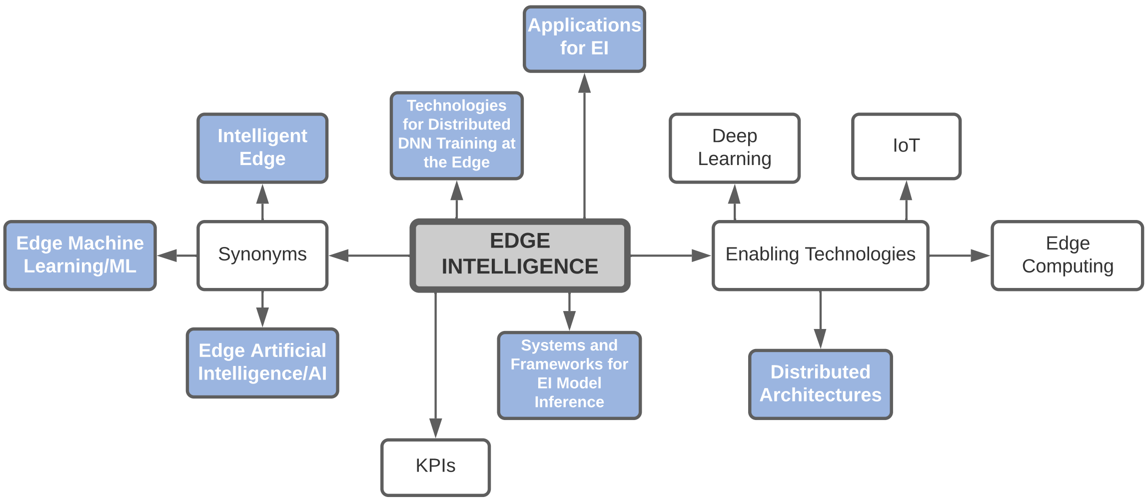

Figure 3: Conceptualisation of the term "edge intelligence" in form of a concept-map, based on findings from Zhou et al. (2019) and this paper's working definition.

Intelligence". Secondly, the working definition states, that *inferencing* and *training* are key concepts in this domain. Hence, individual relations for technologies, frameworks, and systems, which enable these two concepts are included in the conceptualisation. In addition, the scope of the review shows, that current *applications* for EI should be included. Finally, enabling technologies are of interest to the literature review. As established in section 2, this includes IoT, deep learning and edge computing. As the architectures and the interplay between edge nodes, edge clusters and the cloud is also a primary component of edge intelligence, an additional arc is added to the *enabling technologies*.

3.3 Literature Search Process

In this step, the search process will be explained. Firstly, the coding guide and key criteria for the search are defined. As stated before, this paper will scan scientific papers in three large online databases: *Google Scholar*[2], *Web of Science*[3] and *EBSCOhost*[4]. Also, only journal publications will be considered in the process. Especially in Google Scholar patents and citations are by default included in the output. Due to the recent proliferation of edge computing and artificial intelligence, the author hypothesises, that a fixed period of 2010 - 2020, will include all related work on the topic of edge intelligence. To increase the likelihood of finding relevant hits, these options are excluded. The target language is English. This process will comprise six steps:

[2] http://scholar.google.de
[3] http://www.webofknowledge.com
[4] http://search.ebscohost.com

1. Search string development
2. Search string application
3. Preprocessing literature
4. Duplicate cleansing
5. Reading title, abstract and keywords
6. Final selection after complete article reading

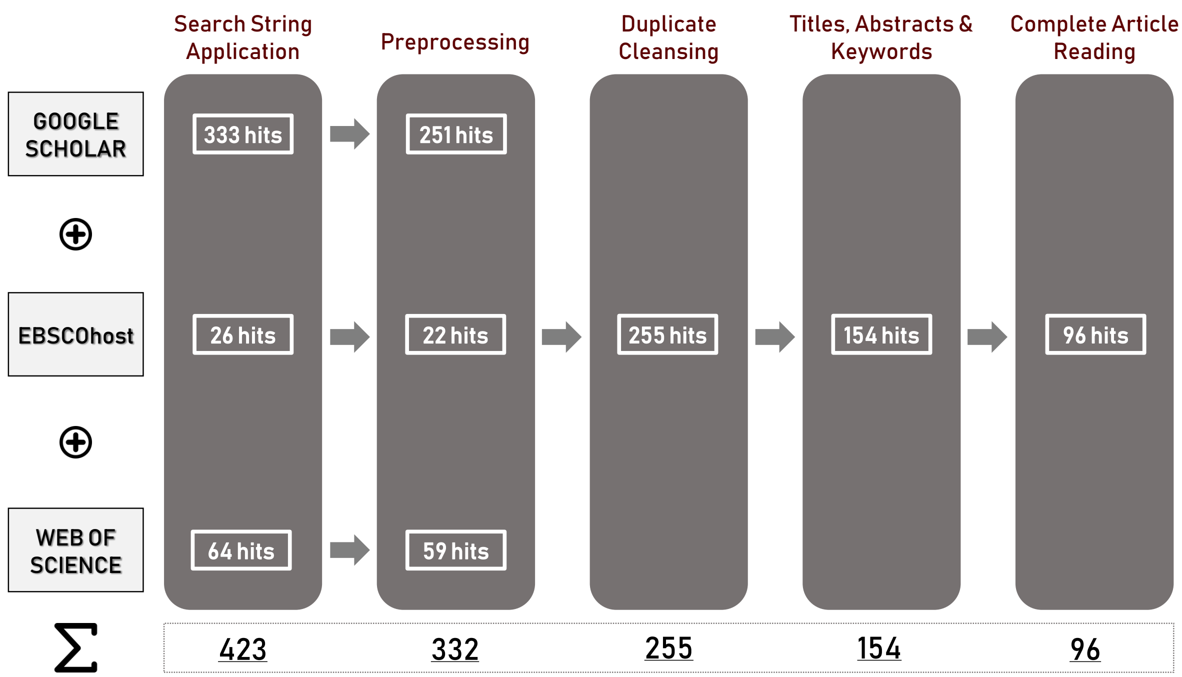

Figure 4: Literature review and relevance filtering process with hit-lists in each step

This process was chosen in order to reduce the bias towards certain scholars. A search string allows the process of finding literature to begin without favouritism. At first the search strings for the databases are developed. Next, the search strings are applied and the results are recorded. With the resulting publications at hand, an initial preprocessing step is included to delete books, presentations or chapters from the list. To guarantee consistency, the results from each database query are then combined and checked for duplicates. Following this step, the papers will be skimmed and the titles, abstracts and keywords will checked for relevancy. Finally, a last selection will be analysed by reading the complete articles and deciding which papers fit the scope. By working through these steps, the author achieves to reduce the hits to the output depicted in figure 4.

3.3.1 Search String Development & Application

In this study, the author chooses to create a search string that involves scanning the databases based on the title. Due to the different search syntax and interfaces, a unique search string has to be created for each database. In accordance with the conceptualisation in figure 3 the literature search will be based on synonyms associated with the term "edge intelligence". In an earlier stage it was considered to perform a search based on titles, abstracts and keywords. However, the search settings on *Google Scholar* are limited in comparison to the other two databases. For example, *Google Scholar* searching is bounded to 256 characters. Another dissimilarity is, that it has no support for truncation and wildcards in the search. In addition, in an early attempt (using the search string depicted in table 2 and allowing complete-search) the hit rate (*11,000 entries*) exceeded the scope of this literature review. For these reasons, a plain title search is recommended. To guarantee the consistency and reproducibility of the process, the keyword search - based on titles - is applied to all databases.

Table 2: Unique search strings for the databases *Google Scholar*, *Web of Science* and *EBSCOhost*

Google Scholar	*allintitle: "Intelligent Edge" OR "Edge Intelligence" OR "Edge AI" OR "Edge Artificial Intelligence" OR "Edge ML" OR "Edge Machine Learning" OR "Artificial Intelligence on the edge" OR "Artificial Intelligence at the edge" OR "AI at the edge" OR "AI on the edge" OR "Edge Learning"*
EBSCOhost	*(TI "Intelligent Edge" OR TI "Edge Intelligence" OR TI "Edge AI" OR TI "Edge Artificial Intelligence" OR TI "Edge ML" OR TI(Edge AND Machine AND Learning) OR TI "Artificial Intelligence on the edge" OR TI "Artificial Intelligenceat the edge" OR TI "AI at the edge" OR TI "AI on the edge" OR TI "Edge Learning") AND (DT > 20100101 AND DT < 20201231)*
Web of Science	*(TI=("Intelligent Edge" OR "Edge Intelligence" OR "Edge AI" OR "Edge Artificial Intelligence" OR "Edge ML" OR "Edge Machine Learning" OR "Artificial Intelligence on the edge" OR "Artificial Intelligence at the edge" OR "AI at the edge" OR "AI on the edge" OR "Edge Learning")) AND LANGUAGE: (English) Indexes=SCI-EXPANDED, SSCI Timespan=2010-2020*

Table 2 contains the final search strings used for this literature review. The search string for *EBSCOhost* was created using the meta-search engine *LitSonar*[5]. This tool provides a platform to create consistent search queries among different databases. However, the search strings for *Google Scholar* and *Web of Science* need to be created manually, as LitSonar does not provide options for transferring the string to these paper banks. After the creation of the string, users with institutional university logins can access and apply the advanced string searches in the named databases. When applying the search string at *EBSCOhost*, the user additionally has to select the libraries, which he wants to query. For this study the six relevant online libraries were chosen - "Business Source Premier",

[5] `http://litsonar.com`

"eBook Collection (EBSCOhost)", "EconLit", "ERIC", "GreenFILE" and "Library, Information Science & Technology Abstracts". The application of the search strings results in a list of 423 hits - 333 from *Google Scholar*, 26 from *EBSCOhost* and 64 from *Web of Science* - as depicted on the first pillar in figure 4. The produced list of found papers is provided in a the public *GitHub* repository[6].

3.3.2 Inclusion & Exclusion Criteria

Before the steps to reduce the list of papers can be performed criteria for exclusion and inclusion of the papers has to be formulated. As established in section 2.1, the boundaries of the edge and the cloud are highly disputed. Obviously, this discussion has an impact on the definition for EI. For example, Zhou et al. (2019) provide a comprehensive survey of the current efforts in EI. The authors focus on the background of AI for EI, as well as edge computing as its enabling technology. They therefore classify edge intelligence into six levels based on where the models are trained and run. The six levels are "*1. Cloud–Edge Coinference and Cloud Training*", "*2. In-Edge Coinference and Cloud Training*", "*3. On-Device Inference and Cloud Training*", "*4. Cloud–Edge Cotraining and Inference*", "*5. All In-Edge*" and "*6. All On-Device*". The fifth and sixth level focus on training and inferencing the Deep Neural Network (DNN) models on edge nodes.

Table 3: Inclusion and exclusion criteria for the relevance filtering process

Inclusion Criteria	**Exclusion Criteria**
1. Academic Journal Papers & Conference Papers 2. Papers in English 3. Papers published between 2010-2020 4. Enabling theories & technologies for training and inferencing on the edge nodes 5. DNN models and architectures for training & inferencing on the edge network 6. Applications of EI based on training & inferencing on the edge network	1. No citations, patents & presentations (Google Scholar) 2. No prefaces or guest editorials 3. Published papers without access 4. Applications, architectures and research outcomes focused on edge-cloud synergy 5. Studies describing computation and storage offloading to the cloud 6. Papers focusing on differences between edge computing and cloud computing

Taking into consideration that several scholars suggest the edge network is decoupled from cloud, in this literature review the author intends to limit the paper selection to studies, which present content exclusively related to training and inferencing DNN

[6]See appendix A for a link and and explanation of the repository

models on the edge network. Furthermore, existing theories, enabling technologies and EI applications, which support this perspective will be reviewed. As edge intelligence is only recently become a research topic, the author expects pertinent results to be published in academic journals and conference papers. Mere comparisons between edge computing and cloud computing will be excluded, as the differences have already been described in section 2.2. A complete list of formal criteria for exclusion and inclusion is listed in table 3.

3.3.3 Relvance Filtering

This section describes the procedure of limiting the papers to the inclusion and exclusion criteria. It follows the application and gathering of the results found in the different databases. As explained in chaper 3.3, the relevance filtering process is composed of *preprocessing literature*, *duplicate cleansing*, *filtering the papers after reading the titles, abstracts and keywords* and *a final selection based on a complete article scan*. In the following the different phases and each process will be explained in more detail. At the end of the filtering process the final list of papers in appendix B remains.

1. Preprocessing Literature: After the search strings are applied to the databases a list of findings is generated. Naturally, at the beginning this list of 423 hits, does not meet the criteria formulated in 3.3.2. As a first step, the preprocessing is performed. The goal of this step is to discard books, presentations, theses, non-english studies, prefaces, guest editorials and other unrelated documents from the initial search results. This measure is mainly required for *Google Scholar* results, because the search settings of this engine do not permit the exclusion of such documents. Additionally, available papers with no access possibility are excluded. To indicate which papers are carried over to the next filtering measure, the tables present in appendix A hold the value "**1**" in the preprocessing column. Excluded rows contain the value "**0**". For verifiability reasons, each discarded row contains an explanation. Excluded rows are marked in the colour red. Altogether, *332* papers remain after this step - 251 from *Google Scholar*, 22 from *EBSCOhost* and 59 from *Web of Science*. The results are depicted in the second pillar in figure 4.

2. Duplicate Cleansing: Next, the results from the paragraph before are checked for duplicates. Further, the available papers are cross-checked for the existing of earlier versions or preproof studies. These papers are then excluded. In this step, independent views for each database are removed and the the sum of all papers is calculated. Filtering for duplicates results in a total of 255 remaining papers. As with step one, the table column for duplicates holds the values explained previously. Again, a reasoning for elimination is given. The third pillar in figure 4 shows the outcome of this phase.

3. Reading Titles, Abstracts & Keywords: Subsequently, the literature review process continues with filtering papers based on reading the titles, abstracts and keywords. For this part the synonyms formulated in the search strings displayed in table 2 are used. Whilst reading, the text is checked to contain these words. Moreover, the analysis incorporates studying these parts for information on. The papers are especially evaluated based on the the inclusion criteria 4-6 as well as the exclusion 4-6 in table 3. If after reading title, abstract and keywords a clear decision whether to keep or discard a paper cannot happen, the conclusion is taken into consideration. Only then can it be guaranteed that relevant papers are not excluded. As with the steps before, the literature review file in appendix A indicates reasons for elimination or keeping of a paper. The result of this phase is displayed in the fourth pillar in figure 4.

4. Complete Article Scan: Last but not least, the remaining papers are subject to a complete scan. Like in step 3. the papers are assessed based on the content related criteria in 2. Special attention is paid to applications and architectures that endorse a edge-cloud synergy. If this is the case, these papers are excluded due to intention of the author to study EI exclusively on the edge network. Irrelevant studies, that mention EI only marginally, are also excluded. If papers are in accordance with the relevant criteria, a short description is added. Again, reasons for inclusion and exclusion are given. This last step leads to a final count of 96 papers. In figure 4, the fifth pillar depicts the result from the last filtering phase.

3.4 Literature Analysis

Now that the relevance filtering process has been concluded and sufficient literature is collected, an analysis will be carried out. For this representative literature review, a rigorously documented search process, based on the Brocke et al. (2009) method, resulted in 96 papers. A consistent search string was created and applied for three major online databases. The process included preprocessing the 423 hits, filtering papers based on duplicates, titles, abstracts, keywords and scanning the remaining articles for relevancy. Through this, a list of fundamental and topic-related contributions is created.

At the beginning, the author hypothesises that relevant papers have mainly been published between 2010 and 2020, as EI is a new realm. Figure 5 shows the count of papers versus their publication year for the 96 papers. It is noteworthy, that there were no contributions between the years 2010-2015 for the final hit-list. This outcome supports the earlier stated hypothesis and suggests the time period even could have been reduced. Another staggering result, is that the vast majority of research has been published over the last two years. Together the count of publications in the years 2019 and 2020 sums up to 85, which represents approximately 88% of the reviewed literature. This implies, that research on edge intelligence is kicking off and a currently a hot topic. For one printed work the publication year cannot be ascertained.

The next step was to examine the final selection for the published journal. When group-

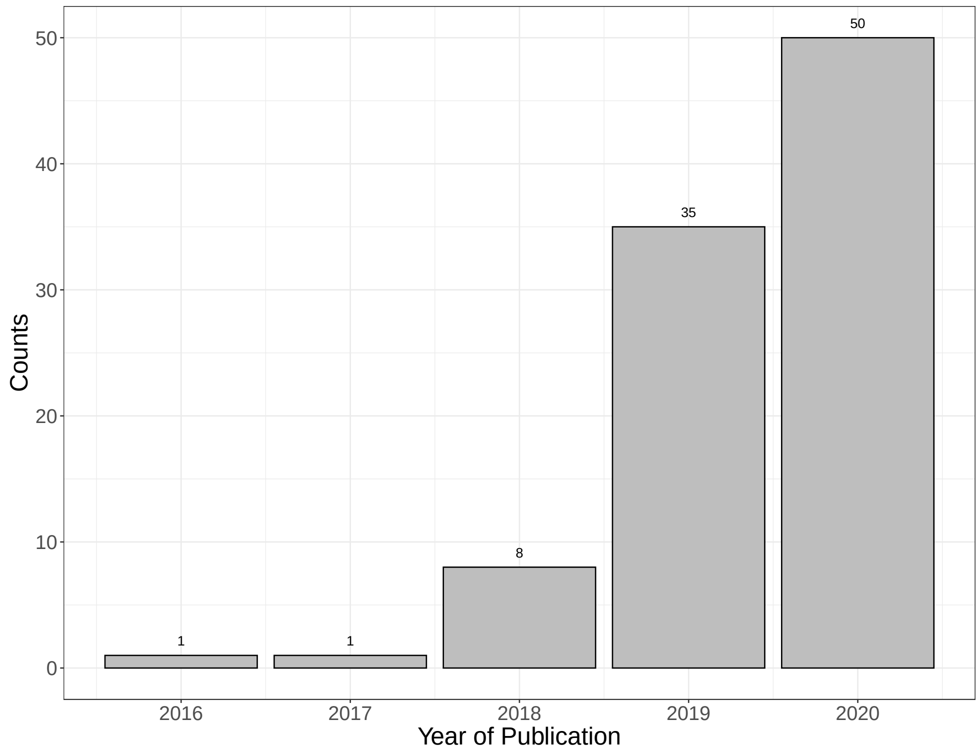

Figure 5: Histogram showing the number of publications per year based on the list of remaining papers after the relevance filtering process

ing the results into their affiliated journals, altogether 59 publishers were found. In first place is the journal *IEEE Network* with 11 publications. Closely followed in second place is the *IEEE Internet of Things Journal* with a total of 8 contributions. The rest of the journals published a maximum of three or less papers. Approximately 46 % of the papers were printed in conference proceedings - mostly IEEE related too. This finding indicates an *IEEE* head start on the research of EI.

Many literature reviews and taxonomies on the topic of edge intelligence exist (Zhou et al., 2019; Y. Shi et al., 2020; D. Xu et al., 2020; Hannan Lodhi, Akgün, and Özkasap, 2020). However, most of them focus on a edge-cloud synergy when describing EI and do not emphasize the potentials of implementing EI strictly on the edge network. What is more, since the data traffic on the edge network is expected to surpass that of the cloud (see section 2.1) a need for studying this view in depth is necessary. For this reason the guiding research question of this paper was "*How are models trained and run on the edge network*". A complete list of the relevance criteria can be found in table 3. For the synthesising of the articles from the search process the author chooses to create a concept-map. This idea was brought forward by Webster and Watson (2002) and is adapted in this study. Due to the keywording of the selected papers in the last step of the relevance filtering, the conceptualisation of the term EI and the definition of the research scope, the four following concepts are formulated for the guiding research question:

1. **Algorithms & DNN design for EI**
2. **Distributed EI architectures, frameworks and systems**
3. **Other EI enabling technologies & theories**
4. **EI application scenarios**

In order to arrange and provide a basis for discussing, the created database of literature is now individually assigned to the concepts. Hereby it is possible that papers can be allocated to multiple groups. *Concept 1* includes those papers which describe for example how to train and inference DNN models on the edge network. Studies describing algorithms for distributed edge architectures are also linked here (e.g. distributed CNNs, DRL, tiny YOLOv3). *Concept 2* contains papers, which deal with EI architectures, frameworks and systems. For instance, hardware and edge computing approaches for EI are included in this concept. *Concept 3* characterises papers, which present other theories and enabling technologies for the realm of EI, such as data security, energy efficiency, data compression and edge service management. *Concept 4* showcases printed work, that deal with real-world applications. This could be V2X systems, healthcare, smart cities, activity recognition, smoke detection and many others. The only condition that unites all concepts is that they focus on a strict edge perspective. The resulting table is attached in appendix B. Due to the grouping of the papers it becomes obvious, that based on the findings, the concept *Distributed EI architectures, frameworks and systems* with 26 entries, is the group with the least linked papers. On the other hand, with 42 links *concept 3* contains the most papers. In section 5 we briefly touch the phase of developing a research agenda presented in the Brocke et al. (2009) cycle and further discuss the implications. Yet, the main focus of the literature review process was to provide data basis for the taxonomy. An in-depth classification will be performed in section 4.

4 Taxonomy Development for EI Research

To further classify all articles, the author chooses to develop a taxonomy according to Nickerson, Varshney, and Muntermann (2013). Based on the results of the previous literature review, the entries in the relevant categories of the concept matrix in appendix B are iterated over to build and refine the taxonomy. As the concept matrix provides a high-level arrangement of papers, the descriptive character defined by Nickerson, Varshney, and Muntermann (2013) is not achieved. However, the goal of this paper is to provide researchers with meaningful guidance and inspire further research interests on the topic of edge intelligence. For this reason an explanatory taxonomy is proposed. To reproduce the taxonomy, the iterative process is saved in the GitHub repository in appendix A.

The taxonomy development process consists of 7 steps and is shown in figure 6. The iterative process starts by (1) determining the meta-characteristics and (2) secondly for-

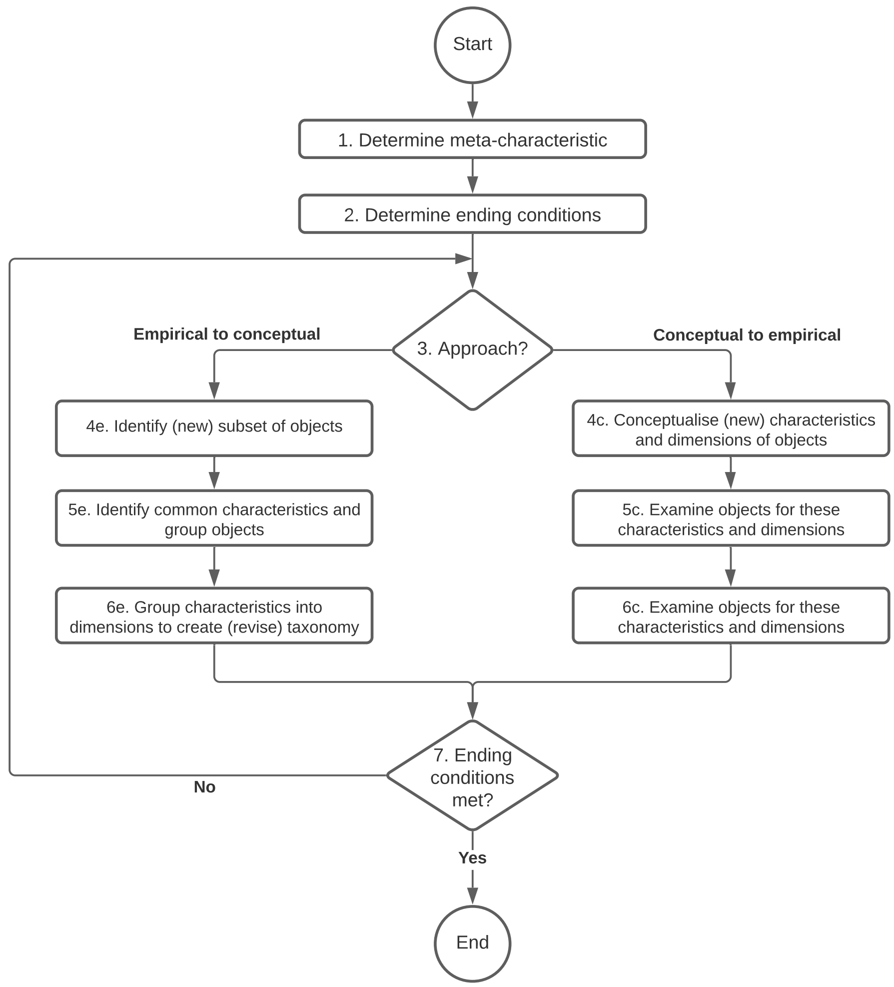

Figure 6: The taxonomy development and refinement method by Nickerson, Varshney, and Muntermann (2013)

mulating ending conditions for the process. The next step is to decide which approach (3) to take. The iteration can be carried out in an *empirical-to-conceptual* or *conceptual-to-empirical* way. In case of the empirical-to-conceptual approach a new subset of objects is identified (4e) and common characteristics to group the objects are established (5e). In the following step mutually exclusive dimensions for the characteristics are derived and an initial or revised taxonomy is created. If the creator chooses the conceptual-to-empirical approach, the characteristics and domains are formulated based on his expertise rather than a subset of objects (4c). This classification is then counterchecked on the basis of the existing objects (5c). In the next step the initial or revised taxonomy is generated (6c). In either case, if the outcome of the approach chosen meets the ending conditions (7), then the taxonomy developing process terminates. Since the dataset available contains 96 papers a large enough database is present and therefore the *empirical-to-conceptual* approach is favoured. Upon identifying the characteristics for each iteration, the author adopts a manual grouping process.

Meta-Characteristics The author defines the meta-characteristic (1) as *design options for EI to run merely on edge nodes*.

Ending Conditions Moreover, the author formulates ending conditions as depicted in figure 6 (2). The iterative taxonomy development process finishes, when the taxonomy provides a meaningful direction for future work on the topic of edge intelligence. This is achieved when subjective and objective prerequisites according to Nickerson, Varshney, and Muntermann (2013) are reached. There are seven objective conditions in total. (I) All objects need to be analysed. (II) At least one object/paper is allocated to each characteristics of each dimensions. (III) There are no new dimensions added in the final iteration. (IV) Dimensions and characteristics stay the same in the final iteration. (V) Dimensions are distinct and not replicated. (VI) The characteristics of the dimensions are distinct. (VII) No cells are duplicates. In addition, the subjective conditions are: explanatory, comprehensive, robust, concise, and extendable (Nickerson, Varshney, and Muntermann, 2013). In each subsection the steps 3-7 will now be iteratively checked until the ending conditions are found.

4.1 Iteration 1 - EI Operating Edge Nodes (EIOEN)

Step 3: Approach: The author decides to utilize the the empirical-to-conceptual approach as no papers have been added yet.

Step 4e: The author selects a sample of papers linked to the category *EI application scenarios* in the concept matrix B. The total amount of papers is 29. For convenience and space reasons, the list of papers is added to the initial taxonomy in appendix C.

Step 5e: Based on the author's generated understanding of what an edge node is and the examination of a convenience sample (Hoshino et al., 2019; Jun Zhang and Letaief, 2019; Keshavarzi and Hoek, 2019; Tang et al., 2020), the following EI operating nodes are identified:

- EI models are run on cameras (C)
- EI models are deployed on Connected Vehicle Systems (V)
- EI models run on other IoT devices (OID)
- EI models are operated by edge servers or gateways (ES/G)
- EI models are distributed across an edge-device synergy platform (EDS)

These characteristics are in line with the meta-characteristic, because each "operator" represents a possibility of running AI models on what is considered an edge platform without a connection to a centralised cloud server.

Step 6e : The number of characteristics is reasonably small, and the author decides to manually group the characteristics into the dimension:

- EI Operating Edge Nodes (EIOEN)

The resulting initial taxonomy is present in appendix C.

Step 7: Ending conditions: As one dimension has been added in this cycle, another iteration has to be carried out. Additionally, 67 unclassified papers remain. Also, the limited number of dimensions leaves room for improvement to satisfy the subjective ending conditions.

4.2 Iteration 2 - EI Training Scheme (EITS)

Step 3: Approach: The author decides to utilize the the empirical-to-conceptual approach as not enough papers have been added to the taxonomy to make it robust.

Step 4e: The author selects a sample size of papers linked to the categories *Algorithms & DNN design for EI* and *Distributed EI architectures, frameworks and systems* in the concept matrix B. The sample size amounts to 44. For convenience and space reasons, the list of papers is added to the refined taxonomy in appendix D.

Step 5e: Based on the author's generated understanding of how AI model can be trained on the edge and the examination of a convenience sample (Rausch, Hummer, et al., 2019; Mills, J. Hu, and Min, 2019; Hung et al., 2020; Yazici, Basurra, and Gaber, 2018), the following EI training techniques are identified:

- On-Device training of AI models (OD)
- Federated learning machine learning across multiple edge nodes (FL)
- Device and edge network computation offloading for training (DEPC)
- Centralised or pre-trained modeling, where training is carried out before deployment on the edge (C/PT)
- Edge pipeline training in sequence (PL)

These characteristics are in line with the meta-characteristic, because each training technique provides an option for AI modeling on the edge.

Step 6e : The number of characteristics is reasonably small, and the author decides to manually group the characteristics into the dimension:

- EI Training Scheme (EITS)

The resulting refined taxonomy is present in appendix D.

Step 7: Ending conditions: As one dimension has been added in this cycle, another iteration has to be carried out. Additionally, 23 unclassified papers remain. Also, the limited number of dimensions leaves room for improvement to satisfy the subjective ending conditions.

4.3 Iteration 3 - Predominant EI Enabling Theory or Technology (PEETT)

Step 3: Approach: The author decides to utilize the the empirical-to-conceptual approach.

Step 4e: The author selects a sample size of papers linked to the categories *Other EI enabling technologies & theories* in the concept matrix B. The sample size amounts to 23. For convenience and space reasons, the list of papers is added to the refined taxonomy in appendix E.

Step 5e: Based on the author's gathered knowledge of the building blocks of EI and the examination of a convenience sample (Doku and Rawat, 2020; C. Dong et al., 2020; Chakraborty et al., 2020; L. Li, Ota, and M. Dong, 2018; Xia et al., 2020; Wang et al., 2019), the following EI enabling theories & technologies are identified:

- Data & cyber security related to EI (D&CS)
- Data filtering & model compression to avoid communication bottlenecks on the edge (DF&C)
- Energy-efficient computation in regard to the limited energy and computing capacities of the edge network (EEC)
- Edge Caching to store content closer to end users (EC)
- (Mobile) Edge computing as the main enabler for EI (MEC)
- Virtualisation and containerisation of software relevant to EI (V)
- Efficient Resource Allocation for edge service management and optimisation (ERA)

These characteristics are in line with the meta-characteristic, because edge intelligence would not function without these technologies and underlying theories.

Step 6e : The number of characteristics is reasonably small, and the author decides to manually group the characteristics into the dimension:

- Predominant EI Enabling Theory or Technology (PEETT)

The resulting refined taxonomy is present in appendix E.

Step 7: Ending conditions: As one dimension has been added in this cycle, another iteration has to be carried out. No unclassified papers remain.

4.4 Iteration 4 - The End

Step 3: Approach: The author decides to use the the empirical-to-conceptual approach.

Step 4e: The author cannot chose a new sample size, as all papers from the concept-matrix are already classified. The sample size is therefore 0.

Steps 5e and 6e: The author cannot identify any new characteristics and dimensions from the sample size. Since there are no papers present no new entries can be added to the taxonomy. Dimensions and characteristics of the taxonomy in appendix E therefore stay the same.

Step 7: Ending conditions: In this iteration there have been no new dimensions added. Also, the characteristics stay the same in the final iteration. Dimensions and characteristics are unique. A meaningful amount of papers (96) has been analysed regarding the topic of edge intelligence. At least one paper is added to each characteristic in each dimension. Therefore, the objective ending conditions are fulfilled. The taxonomy is concise, extendable, robust, explanatory and comprehensive. The taxonomy developing process terminates at this point. In the following discussion we will discuss the implications.

5 Discussion

In this paper the author discusses the theoretical background of edge intelligence and provides a working definition. Until today no universally accepted definition on the scope of the edge exists. However, the data traffic produced on the edge has surpassed that of the cloud. The author hypothesizes, that this situation is part of the reason why edge intelligence is becoming a trending topic. Research shows, that surveys deal with EI in a broad context (Zhou et al., 2019; Y. Shi et al., 2020; D. Xu et al., 2020; Hannan Lodhi, Akgün, and Özkasap, 2020). This includes describing EI as an edge-cloud synergy. The perspective of focusing entirely on the edge environment to train, operate and process AI-models is often neglected. Nonetheless, this perspective is important, because of latency issues, security reasons, increasing data traffic amounts and bandwidth problems in rural areas.

The present paper tries to shed light on this perspective. A representative literature review based on the Brocke et al. (2009) is performed. The research scope involves applications, theories and research outcomes for this restricted view. First, the topic is conceptualised. Then, the generated synonyms from the conceptualisation are used to create the title-based search string. Three consistent search strings are applied to the online databases *Google Scholar*, *EBSCOhost* and *Web of Science*. This results in 423 hits. Following this, the relevance filtering process is carried out - (1) The hits are preprocessed, (2) duplicate papers are deleted, (3) filtering is executed based on reading the titles, abstract and keywords and (5) a complete text scan of the remaining papers is carried out. This process results in 96 papers. 88% of this reviewed literature is published in the years 2019 and 2020. This confirms the earlier stated EI research trend. Based on these papers a concept matrix is created and the studies are grouped into four concepts. The concept *Other EI enabling technologies & theories* has the most links, with 42 papers classified to this category. On the other hand, the concept *Distributed EI architectures, frameworks and systems* contains the least links, with 26 entries. This suggests, that although the technologies are available, implementation is still stalling.

The concept matrix is also the starting point for the taxonomy developing process (Nickerson, Varshney, and Muntermann, 2013). In four iterations a refined taxonomy of 96 object, 3 dimensions and 17 characteristics are created. The dimensions are (1) *EI Training Scheme (EITS)*, *EI Operating Edge Nodes (EIOEN)* and *Predominant EI Enabling Theory or Technology (PEETT)*. All of the dimensions are created in connection to the meta-characteristic - *design options for EI to run merely on edge nodes*. The analysis shows, that operating EI models is most frequently accomplished in an edge-device synergy (42 papers). In addition, most AI models are trained edge-device collaboration scheme (45 papers). If a paper presents an edge-device training technique, then the operating of the AI-models is also accomplished in an edge-device synergy. Technologies related to *data filtering & compression*, as well as *energy-efficient computation* are deemed the most important. This may seem unsurprising, because of the limited computation capacities of the edge network. For convenience and space related reasons the concept matrix and taxonomy developing stages are located in the appendix. All files necessary for the literature research and the taxonomy developing process are outsourced to a GitHub repository.

Future work can focus on on-device related EI to achieve further transparency on all levels of edge intelligence. The created search strings can be extended to include further synonyms and be applied to more databases. Furthermore, this paper can be used to research current EI applications for the edge network and generate insights for future EI-specific frameworks and systems. As the taxonomy clearly states, that edge-device synergy, energy-efficiency and data compression techniques are the main building blocks of EI, these topics have to be researched more. The taxonomy can be extended to include further dimensions by analysing additional objects.

6 Conclusion

Edge intelligence is a booming research topic. However, the standardisation of this realm is far from being finalised. Taken into consideration, that edge intelligence can have different levels, in this paper the author focuses on design options for EI to run merely on the edge environment. For this, the theoretical background of edge intelligence is analysed and a comprehensive literature review is carried out. Based on the created study database a concept map and a taxonomy are created. The resulting analysis shows, that for this EI perspective, research focuses on edge-device synergy, energy-efficiency and data compression techniques. Future work should study on-device EI training and operation, as well as developing frameworks and systems for EI for the edge network.

References

Ahmed, E., A. Ahmed, Ibrar Yaqoob, Junaid Shuja, A. Gani, Muhammad Imran, and Muhammad Shoaib (2017). "Bringing Computation Closer toward the User Network: Is Edge Computing the Solution?" In: *IEEE Communications Magazine* 55, pp. 138–144.

Alanezi, Khaled and Shivakant Mishra (n.d.). "An edge-based architecture to support the execution of ambience intelligence tasks using the IoP paradigm". In: *Future Generation Computer Systems* 114 (), pp. 349–357.

Alaslani, Maha and Basem Shihada (2018). "Intelligent edge: An instantaneous detection of iot traffic load". In: *2018 IEEE International Conference on Communications (ICC)*. IEEE, pp. 1–6.

Alonso, Ricardo S, Inés Sittón-Candanedo, Óscar Garcia, Javier Prieto, and Sara Rodriguez-González (2020). "An intelligent Edge-IoT platform for monitoring livestock and crops in a dairy farming scenario". In: *Ad Hoc Networks* 98, p. 102047.

Azar, Joseph, Abdallah Makhoul, Mahmoud Barhamgi, and Raphaël Couturier (2019). "An energy efficient IoT data compression approach for edge machine learning". In: *Future Generation Computer Systems* 96, pp. 168–175.

Beavers, Ian (Aug. 2017). "Intelligence at the Edge Part 1: The Edge Node". In: visited on 10.10.2020. URL: `https://www.analog.com/media/en/technical-documentation/tech-articles/Intelligence-at-the-Edge-Part-1-The-Edge-Node.pdf`.

Belyi, Vladimir (Feb. 2020). *How Efficient is Edge Computing Compared to Cloud Computing?* visited on 09.10.2020. URL: `https://www.dataversity.net/how-efficient-is-edge-computing-compared-to-cloud-computing/`.

Bilal, Kashif, Osman Khalid, Aiman Erbad, and Samee U Khan (2018). "Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers". In: *Computer Networks* 130, pp. 94–120.

Bonomi, Flavio, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli (2012). "Fog computing and its role in the internet of things". In: *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pp. 13–16.

Brennan, Robert L (2019). "AI, IoT hardware and Algorithmic Considerations for Hearing aid and Extreme Edge Applications". In: *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, pp. 841–844.

Brocke, Jan vom, Alexander Simons, Bjoern Niehaves, Bjorn Niehaves, Kai Reimer, Ralf Plattfaut, and Anne Cleven (2009). "Reconstructing the giant: On the importance of rigour in documenting the literature search process". In:

Chakraborty, Indranil, Deboleena Roy, Isha Garg, Aayush Ankit, and Kaushik Roy (2020). "Constructing energy-efficient mixed-precision neural networks through principal component analysis for edge intelligence". In: *Nature Machine Intelligence* 2.1, pp. 43–55.

Chang, Zheng (n.d.). *Edge Intelligence for Immersive Communications*. visited on 09.09.2020. URL: https://www.comsoc.org/publications/journals/ieee-ojcoms/cfp/edge-intelligence-immersive-communications.

Chen, Sifan, Peng Gong, Bin Wang, Alagan Anpalagan, Mohsen Guizani, and Chungang Yang (2019). "EDGE AI for Heterogeneous and Massive IoT Networks". In: *2019 IEEE 19th International Conference on Communication Technology (ICCT)*. IEEE, pp. 350–355.

Chen, Tung-Chien, Wei-Ting Wang, Kloze Kao, Chia-Lin Yu, Code Lin, Shu-Hsin Chang, and Pei-Kuei Tsung (2019). "NeuroPilot: A Cross-Platform Framework for Edge-AI". In: *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, pp. 167–170.

Chen, Zhuang, Qian He, Lei Liu, Dapeng Lan, Hwei-Ming Chung, and Zhifei Mao (2019). "An artificial intelligence perspective on mobile edge computing". In: *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*. IEEE, pp. 100–106.

Cisco (2015). *Cisco Global Cloud Index 2014–2019: Forecast and methodology, 2015 Update, white paper*. visited on 25.09.2020. URL: https://www.cisco.com/c/dam/m/en_us/service-provider/ciscoknowledgenetwork/files/547_11_10-15-DocumentsCisco_GCI_Deck_2014-2019_for_CKN__10NOV2015_.pdf.

Cooper, Harris M (1988). "Organizing knowledge syntheses: A taxonomy of literature reviews". In: *Knowledge in society* 1.1, p. 104.

Corcoran, Peter, Joseph Lemley, Claudia Costache, and Viktor Varkarakis (2019). "Deep Learning for Consumer Devices and Services 2—AI Gets Embedded at the Edge". In: *IEEE Consumer Electronics Magazine* 8.5, pp. 10–19.

Dahad, Nitin (Feb. 19, 2020). *Let's Talk Edge Intelligence*. URL: https://www.eetimes.com/lets-talk-edge-intelligence/.

De Donno, Michele, Koen Tange, and Nicola Dragoni (2019). "Foundations and evolution of modern computing paradigms: Cloud, iot, edge, and fog". In: *Ieee Access* 7, pp. 150936–150948.

Deng, Shuiguang, Hailiang Zhao, Weijia Fang, Jianwei Yin, Schahram Dustdar, and Albert Y Zomaya (2020). "Edge intelligence: the confluence of edge computing and artificial intelligence". In: *IEEE Internet of Things Journal*.

Doku, Ronald and Danda B Rawat (2020). "IFLBC: On the Edge Intelligence Using Federated Learning Blockchain Network". In: *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. IEEE, pp. 221–226.

Dolui, Koustabh and Soumya Kanti Datta (2017). "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing". In: *2017 Global Internet of Things Summit (GIoTS)*. IEEE, pp. 1–6.

Dong, Chao, Yun Shen, Yuben Qu, Qihui Wu, Fan Wu, and Guihai Chen (2020). "UAVs as a Service: Boosting Edge Intelligence for Air-Ground Integrated Networks". In: *arXiv preprint arXiv:2003.10737*.

Forbes (2019). *Predictions 2020: Edge Computing Makes The Leap*. https://www.forbes.com/sites/forrester/2019/12/02/predictions-2020-edge-computing-makes-the-leap/?sh=10bc29404201. (Accessed on 10/30/2020).

Gamanayake, Chinthaka Madhushan, Lahiru Aruna Jayasinghe, Benny Ng, and Chau Yuen (2020). "Cluster Pruning: An Efficient Filter Pruning Method for Edge AI Vision Applications". In: *IEEE Journal of Selected Topics in Signal Processing*.

Garcia Lopez, Pedro, Alberto Montresor, Dick Epema, Anwitaman Datta, Teruo Higashino, Adriana Iamnitchi, Marinho Barcellos, Pascal Felber, and Etienne Riviere (2015). *Edge-centric computing: Vision and challenges*.

Gong, Chao, Fuhong Lin, Xiaowen Gong, and Yueming Lu (2020). "Intelligent Cooperative Edge Computing in the Internet of Things". In: *IEEE Internet of Things Journal*.

Guleng, Siri, Celimuge Wu, Zhi Liu, and Xianfu Chen (2020). "Edge-Based V2X Communications With Big Data Intelligence". In: *IEEE Access* 8, pp. 8603–8613.

Hannan Lodhi, Ahnaf, Barış Akgün, and Öznur Özkasap (2020). "State-of-the-art Techniques in Deep Edge Intelligence". In: *arXiv e-prints*, arXiv–2008.

Hao, Cong, Yao Chen, Xiaofan Zhang, Yuhong Li, Jinjun Xiong, Wen-mei Hwu, and Deming Chen (2020). "Effective Algorithm-Accelerator Co-design for AI Solutions on Edge Devices". In: *Proceedings of the 2020 on Great Lakes Symposium on VLSI*, pp. 283–290.

Hassan, Najmul, Saira Gillani, Ejaz Ahmed, Ibrar Yaqoob, and Muhammad Imran (2018). "The role of edge computing in internet of things". In: *IEEE communications magazine* 56.11, pp. 110–115.

Hassija, Vikas, Vinay Chamola, Vikas Saxena, Divyansh Jain, Pranav Goyal, and Biplab Sikdar (2019). "A survey on IoT security: application areas, security threats, and solution architectures". In: *IEEE Access* 7, pp. 82721–82743.

Hoshino, Yasutaka, Obuchi Masaya, Kazuhiro Motegi, and Yoichi Shiraishi (2019). "Implementation of Skeleton Extraction Process from Camera Images on Edge AI Processor and its Applications". In:

Höß, Oliver (2018). *Cloud Computing, Edge Computing und Fog Computing – Unterschiede kurz erklärt*. visited on 09.10.2020. URL: https://innovative-trends.de/2018/01/02/cloud-computing-edge-computing-und-fog-computing-unterschiede-kurz-erklart/.

HPE (n.d.). *Edge Computing Definition*. visited on 02.09.2020. URL: https://www.hpe.com/us/en/what-is/edge-computing.html.

Huang, Yu-De, Kai-Yen Wang, Yun-Lung Ho, Chang-Yuan He, and Wai-Chi Fang (2019). "An edge AI system-on-chip design with customized convolutional-neural-network architecture for real-time EEG-based affective computing system". In: *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, pp. 1–4.

Huang, Zhenqiu, Kwei-Jay Lin, and Chi-Sheng Shih (2016). "Supporting edge intelligence in service-oriented smart iot applications". In: *2016 IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, pp. 492–499.

Huang, Zhenqiu, Kwei-Jay Lin, Bo-Lung Tsai, Surong Yan, and Chi-Sheng Shih (2018). “Building edge intelligence for online activity recognition in service-oriented IoT systems”. In: *Future Generation Computer Systems* 87, pp. 557–567.

Hung, Je-Min, Xueqing Li, Juejian Wu, and Meng-Fan Chang (2020). “Challenges and Trends inDeveloping Nonvolatile Memory-Enabled Computing Chips for Intelligent Edge Devices”. In: *IEEE Transactions on Electron Devices* 67.4, pp. 1444–1453.

IEC (2017). *White Paper Edge Intelligence*. URL: https://www.iec.ch/whitepaper/edgeintelligence.

Iorga, Michaela, Larry Feldman, Robert Barton, Michael Martin, Nedim Goren, and Charif Mahmoudi (2017). *The nist definition of fog computing*. Tech. rep. National Institute of Standards and Technology.

Kaneko, Tatsuya, Kentaro Orimo, Itaru Hida, Shinya Takamaeda-Yamazaki, Masayuki Ikebe, Masato Motomura, and Tetsuya Asai (2019). “A study on a low power optimization algorithm for an edge-AI device”. In: *Nonlinear Theory and Its Applications, IEICE* 10.4, pp. 373–389.

Keshavarzi, Ali and Wilbert van den Hoek (2019). “Edge intelligence—On the challenging road to a trillion smart connected IoT devices”. In: *IEEE Design & Test* 36.2, pp. 41–64.

Kolomvatsos, Kostas and Christos Anagnostopoulos (2020). “An Intelligent Edge-centric Queries Allocation Scheme based on Ensemble Models”. In: *ACM Transactions on Internet Technology (TOIT)* 20.4, pp. 1–25.

Lan, Qiao, Zezhong Zhang, Yuqing Du, Zhenyi Lin, and Kaibin Huang (2019). “An Introduction to Communication Efficient Edge Machine Learning”. In: *arXiv preprint arXiv:1912.01554*.

Lanner (n.d.). *Intelligent Edge | Lanner*. https://www.lanner-america.com/knowledgebase/intelligent-edge/. (Accessed on 10/30/2020).

Larras, Benoit and Antoine Frappé (2020). “On the Distribution of Clique-Based Neural Networks for Edge AI”. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*.

Li, En, Liekang Zeng, Zhi Zhou, and Xu Chen (2019). “Edge AI: On-demand accelerating deep neural network inference via edge computing”. In: *IEEE Transactions on Wireless Communications* 19.1, pp. 447–457.

Li, Gaolei, Guangquan Xu, Arun Kumar Sangaiah, Jun Wu, and Jianhua Li (2019). “EdgeLaaS: Edge learning as a service for knowledge-centric connected healthcare”. In: *IEEE Network* 33.6, pp. 37–43.

Li, Liangzhi, Kaoru Ota, and Mianxiong Dong (2018). “DeepNFV: A lightweight framework for intelligent edge network functions virtualization”. In: *IEEE Network* 33.1, pp. 136–141.

Liang, Qianlin, Prashant Shenoy, and David Irwin (2020). “AI on the Edge: Rethinking AI-based IoT Applications Using Specialized Edge Architectures”. In: *arXiv preprint arXiv:2003.12488*.

Libri, Antonio, Andrea Bartolini, and Luca Benini (2020). “pAElla: Edge-AI based Real-Time Malware Detection in Data Centers”. In: *IEEE Internet of Things Journal*.

Liu, Dongzhu, Guangxu Zhu, Jun Zhang, and Kaibin Huang (2020). "Data-importance aware user scheduling for communication-efficient edge machine learning". In: *IEEE Transactions on Cognitive Communications and Networking*.

Liu, Yaqiong, Mugen Peng, Guochu Shou, Yudong Chen, and Siyu Chen (2020). "Toward Edge Intelligence: Multiaccess Edge Computing for 5G and Internet of Things". In: *IEEE Internet of Things Journal* 7.8, pp. 6722–6747.

Liu, Yi, Chao Yang, Li Jiang, Shengli Xie, and Yan Zhang (2019). "Intelligent edge computing for IoT-based energy management in smart cities". In: *IEEE Network* 33.2, pp. 111–117.

Lv, Zhihan, Dongliang Chen, and Qingjun Wang (2020). "Diversified technologies in internet of vehicles under intelligent edge computing". In: *IEEE Transactions on Intelligent Transportation Systems*.

Mach, Pavel and Zdenek Becvar (2017). "Mobile edge computing: A survey on architecture and computation offloading". In: *IEEE Communications Surveys & Tutorials* 19.3, pp. 1628–1656.

Mazzia, Vittorio, Aleem Khaliq, Francesco Salvetti, and Marcello Chiaberge (2020). "Real-Time Apple Detection System Using Embedded Systems With Hardware Accelerators: An Edge AI Application". In: *IEEE Access* 8, pp. 9102–9114.

Mell, Peter, Tim Grance, et al. (2011). "The NIST definition of cloud computing". In:

Mills, Jed, Jia Hu, and Geyong Min (2019). "Communication-efficient federated learning for wireless edge intelligence in iot". In: *IEEE Internet of Things Journal*.

Mittal, Varnit and Bharat Bhushan (2020). "Accelerated Computer Vision Inference with AI on the Edge". In: *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*. IEEE, pp. 55–60.

Mittal, Varnit, Ashi Tyagi, and Bharat Bhushan (2020). "Smart Surveillance Systems with Edge Intelligence: Convergence of Deep Learning and Edge Computing". In: *Available at SSRN 3599865*.

Muhammad, Khan, Salman Khan, Vasile Palade, Irfan Mehmood, and Victor Hugo C De Albuquerque (2019). "Edge intelligence-assisted smoke detection in foggy surveillance environments". In: *IEEE Transactions on Industrial Informatics* 16.2, pp. 1067–1075.

Mukherjee, Mithun, Mian Guo, Jaime Lloret, and Qi Zhang (2020). "Leveraging Intelligent Computation Offloading with Fog/Edge Computing for Tactile Internet: Advantages and Limitations". In: *Ieee Network* 34.5, pp. 322–329.

Naha, Ranesh, Saurabh Garg, Dimitrios Georgakopoulos, Prem Prakash Jayaraman, Longxiang Gao, Yong Xiang, and R. Ranjan (Aug. 2018). "Fog Computing: Survey of Trends, Architectures, Requirements, and Research Directions". In: *IEEE Access* PP, pp. 1–1. DOI: 10.1109/ACCESS.2018.2866491.

Nickerson, Robert C, Upkar Varshney, and Jan Muntermann (2013). "A method for taxonomy development and its application in information systems". In: *European Journal of Information Systems* 22.3, pp. 336–359.

Oderhohwo, Ogheneuriri, Hawzhin Mohammed, Tolulope Odetola, Terry N Guo, Syed Hasan, and Felix Dogbe (2020). "An Edge Intelligence Framework for Resource Constrained Community Area Network". In: *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, pp. 97–100.

Pan, Jieming, Yuxuan Luo, Yida Li, Chen-Khong Tham, Chun-Huat Heng, and Aaron Voon-Yew Thean (2020). "A Wireless Multi-Channel Capacitive Sensor System for Efficient Glove-Based Gesture Recognition With AI at the Edge". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 67.9, pp. 1624–1628.

Pandit, Mohammad Khalid, Roohie Naaz Mir, and Mohammad Ahsan Chihisti (2017). "Machine learning at the edge of internet of things". In: *CSI Communications* 41.8, pp. 28–30.

Park, Jihong, Sumudu Samarakoon, Mehdi Bennis, and Mérouane Debbah (2019). "Wireless network intelligence at the edge". In: *Proceedings of the IEEE* 107.11, pp. 2204–2239.

Peltonen, Ella, Mehdi Bennis, Michele Capobianco, Merouane Debbah, Aaron Ding, Felipe Gil-Castiñeira, Marko Jurmu, Teemu Karvonen, Markus Kelanti, Adrian Kliks, et al. (2020). "6G White Paper on Edge Intelligence". In: *arXiv preprint arXiv:2004.14850*.

Plastiras, George, Maria Terzi, Christos Kyrkou, and Theocharis Theocharidcs (2018). "Edge intelligence: Challenges and opportunities of near-sensor machine learning applications". In: *2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*. IEEE, pp. 1–7.

Premsankar, Gopika, Mario Di Francesco, and Tarik Taleb (2018). "Edge computing for the Internet of Things: A case study". In: *IEEE Internet of Things Journal* 5.2, pp. 1275–1284.

Qu, Yuben, Chao Dong, Jianchao Zheng, Qihui Wu, Yun Shen, Fan Wu, and Alagan Anpalagan (2020). "Empowering the Edge Intelligence by Air-Ground Integrated Federated Learning in 6G Networks". In: *arXiv preprint arXiv:2007.13054*.

Queralta, J Pena, Tuan Nguyen Gia, Hannu Tenhunen, and Tomi Westerlund (2019). "Edge-AI in LoRa-based health monitoring: fall detection system with fog computing and LSTM recurrent neural networks". In: *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, pp. 601–604.

Queralta, Jorge Peña, Jenni Raitoharju, Tuan Nguyen Gia, Nikolaos Passalis, and Tomi Westerlund (2020). "AutoSOS: Towards Multi-UAV Systems Supporting Maritime Search and Rescue with Lightweight AI and Edge Computing". In: *arXiv preprint arXiv:2005.03409*.

Qureshi, Kashif Naseer, Abeer Iftikhar, Shahid Nazeer Bhatti, Francesco Piccialli, Fabio Giampaolo, and Gwanggil Jeon (2020). "Trust management and evaluation for edge intelligence in the Internet of Things". In: *Engineering Applications of Artificial Intelligence* 94, p. 103756.

Rahman, Mohammad Saidur, Ibrahim Khalil, Mohammed Atiquzzaman, and Xun Yi (2020). "Towards privacy preserving AI based composition framework in edge networks using fully homomorphic encryption". In: *Engineering Applications of Artificial Intelligence* 94, p. 103737.

Rahmani, Amir M, Pasi Liljeberg, Jürgo-Sören Preden, and Axel Jantsch (2017). *Fog computing in the internet of things: Intelligence at the edge*. Springer.

Rausch, Thomas and Schahram Dustdar (2019). "Edge intelligence: The convergence of humans, things, and ai". In: *2019 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, pp. 86–96.

Rausch, Thomas, Waldemar Hummer, Vinod Muthusamy, Alexander Rashed, and Schahram Dustdar (2019). "Towards a serverless platform for edge {AI}". In: *2nd {USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 19)*.

Romaszkan, Wojciech, Tianmu Li, and Puneet Gupta (2020). "3PXNet: Pruned-Permuted-Packed XNOR Networks for Edge Machine Learning". In: *ACM Transactions on Embedded Computing Systems (TECS)* 19.1, pp. 1–23.

Satyanarayanan, Mahadev (2017). "The emergence of edge computing". In: *Computer* 50.1, pp. 30–39.

Sharma, S. K. and X. Wang (2017). "Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks". In: *IEEE Access* 5, pp. 4621–4635.

Shi, Weisong, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu (2016). "Edge computing: Vision and challenges". In: *IEEE internet of things journal* 3.5, pp. 637–646.

Shi, Weisong and Schahram Dustdar (2016). "The promise of edge computing". In: *Computer* 49.5, pp. 78–81.

Shi, Yuanming, Kai Yang, Tao Jiang, Jun Zhang, and Khaled B Letaief (2020). "Communication-efficient edge AI: Algorithms and systems". In: *arXiv preprint arXiv:2002.09668*.

Skatchkovsky, Nicolas, Hyeryung Jang, and Osvaldo Simeone (2020). "Federated Neuromorphic Learning of Spiking Neural Networks for Low-Power Edge Intelligence". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 8524–8528.

Skatchkovsky, Nicolas and Osvaldo Simeone (2019). "Optimizing pipelined computation and communication for latency-constrained edge learning". In: *IEEE Communications Letters* 23.9, pp. 1542–1546.

Stewart, Duncan, Jeff Loucks, Mark Casey, and Craig Wigginton (2019). *Bringing AI to the device: Edge AI chips come into their own, TMT Predictions 2020*. visited on 25.09.2020. URL: `https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2020/ai-chips.html`.

Strike, K and G Posner (1983). *Types of synthesis and their criteria. Knowledge, structure, and use*.

Tang, Xin, Xu Chen, Liekang Zeng, Shuai Yu, and Lin Chen (2020). "Joint Multi-User DNN Partitioning and Computational Resource Allocation for Collaborative Edge Intelligence". In: *IEEE Internet of Things Journal*.

Thangiah, Leny, Chandrashekar Ramanathan, and Lakshmi Sirisha Chodisetty (2019). "Distribution Transformer Condition Monitoring based on Edge Intelligence for Industrial IoT". In: *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*. IEEE, pp. 733–736.

Varghese, Blesson, Nan Wang, Sakil Barbhuiya, Peter Kilpatrick, and Dimitrios S Nikolopoulos (2016). "Challenges and opportunities in edge computing". In: *2016 IEEE International Conference on Smart Cloud (SmartCloud)*. IEEE, pp. 20–26.

Vatti, Rambabu Arjunarao, K Vinoth, and Yerram Sneha (2020). "EDGE INTELLIGENCE FOR PREDICTING AND DETECTING CARDIAC PATHOLOGIES BY ANALYZING STRESS AND ANXIETY". In: *Journal of Critical Reviews*.

Veena, Pureswaran, Sanjay Panikkar, Sumabala Nair, and Paul Brody (2015). "Empowering the edge-practical insights on a decentralized internet of things". In: *IBM Institute for Business Value* 17.

Wang, Dong, Daniel Zhang, Yang Zhang, Md Tahmid Rashid, Lanyu Shang, and Na Wei (2019). "Social edge intelligence: Integrating human and artificial intelligence at the edge". In: *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, pp. 194–201.

Wang, Fangxin, Cong Zhang, Jiangchuan Liu, Yifei Zhu, Haitian Pang, Lifeng Sun, et al. (2019). "Intelligent edge-assisted crowdcast with deep reinforcement learning for personalized QoE". In: *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, pp. 910–918.

Wang, Yiwen Han, Chenyang Wang, Qiyang Zhao, Xu Chen, and Min Chen (2019). "In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning". In: *IEEE Network* 33.5, pp. 156–165.

Wang, Qianlong, Yifan Guo, Xufei Wang, Tianxi Ji, Lixing Yu, and Pan Li (2020). "AI at the Edge: Blockchain-Empowered Secure Multiparty Learning with Heterogeneous Models". In: *IEEE Internet of Things Journal*.

Wang, Shiqiang, Tiffany Tuor, Theodoros Salonidis, Kin K Leung, Christian Makaya, Ting He, and Kevin Chan (2018). "When edge meets learning: Adaptive control for resource-constrained distributed machine learning". In: *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, pp. 63–71.

Wang, Shuai, Rui Wang, Qi Hao, Yik-Chung Wu, and H Vincent Poor (2020). "Learning Centric Power Allocation for Edge Intelligence". In: *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, pp. 1–6.

Wang, Shuai, Yik-Chung Wu, Minghua Xia, Rui Wang, and H Vincent Poor (2020). "Machine intelligence at the edge with learning centric power allocation". In: *IEEE Transactions on Wireless Communications*.

Wang, Tian, Yuzhu Liang, Yi Yang, Guangquan Xu, Hao Peng, Anfeng Liu, and Weijia Jia (2020). "An Intelligent Edge-Computing-Based Method to Counter Coupling Problems in Cyber-Physical Systems". In: *IEEE Network* 34.3, pp. 16–22.

Webster, Jane and Richard T Watson (2002). "Analyzing the past to prepare for the future: Writing a literature review". In: *MIS quarterly*, pp. xiii–xxiii.

Wen, Dingzhu, Xiaoyang Li, Qunsong Zeng, Jinke Ren, and Kaibin Huang (2019). "An overview of data-importance aware radio resource management for edge machine learning". In: *Journal of Communications and Information Networks* 4.4, pp. 1–14.

Wolf, Marilyn (2019). "Machine Learning+ Distributed IoT= Edge Intelligence". In: *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, pp. 1715–1719.

Xia, Junxu, Geyao Cheng, Deke Guo, and Xiaolei Zhou (2020). "A QoE-Aware Service Enhancement Strategy for Edge Artificial Intelligence Applications". In: *IEEE Internet of Things Journal*.

Xiao, Yong, Guangming Shi, Yingyu Li, Walid Saad, and H Vincent Poor (2020). "Towards Self-learning Edge Intelligence in 6G". In: *arXiv preprint arXiv:2010.00176*.

Xie, Feiyi, Aidong Xu, Yixin Jiang, Songlin Chen, Runfa Liao, and Hong Wen (2019). "Edge Intelligence based Co-training of CNN". In: *2019 14th International Conference on Computer Science & Education (ICCSE)*. IEEE, pp. 830–834.

Xu, Dianlei, Tong Li, Yong Li, Xiang Su, Sasu Tarkoma, and Pan Hui (2020). "A Survey on Edge Intelligence". In: *arXiv preprint arXiv:2003.12172*.

Xu, Shengjie, Yi Qian, and Rose Qingyang Hu (2019). "Data-driven edge intelligence for robust network anomaly detection". In: *IEEE Transactions on Network Science and Engineering*.

— (2020). "Edge Intelligence Assisted Gateway Defense in Cyber Security". In: *IEEE Network* 34.4, pp. 14–19.

YANG, Howard H, ZHAO Zhongyuan, and Tony QS QUEK (2020). "Enabling Intelligence at Network Edge Network Edge: An Overview of Federated Learning". In: *ZTE COMMUNICATIONS* 18.2.

Yang, Kai, Yong Zhou, Zhanpeng Yang, and Yuanming Shi (2020). "Communication-Efficient Edge AI Inference Over Wireless Networks". In: *arXiv preprint arXiv:2004.13351*.

Yang, Lei, Yanyan Lu, Jiannong Cao, Jiaming Huang, and Mingjin Zhang (2020). "E-Tree Learning: A Novel Decentralized Model Learning Framework for Edge AI". In: *arXiv preprint arXiv:2008.01553*.

Yang, Xiangyu, Sheng Hua, Yuanming Shi, Hao Wang, Jun Zhang, and Khaled B Letaief (2020). "Sparse Optimization for Green Edge AI Inference". In: *Journal of Communications and Information Networks* 5.1, pp. 1–15.

Yazici, Mahmut Taha, Shadi Basurra, and Mohamed Medhat Gaber (2018). "Edge machine learning: Enabling smart internet of things applications". In: *Big data and cognitive computing* 2.3, p. 26.

Yu, Wei, Fan Liang, Xiaofei He, William Grant Hatcher, Chao Lu, Jie Lin, and Xinyu Yang (2017). "A survey on the edge computing for the Internet of Things". In: *IEEE access* 6, pp. 6900–6919.

Yuan, Bo, John Panneerselvam, Lu Liu, Nick Antonopoulos, and Yao Lu (2019). "An inductive content-augmented network embedding model for edge artificial intelligence". In: *IEEE Transactions on Industrial Informatics* 15.7, pp. 4295–4305.

Zeng, Liekang, En Li, Zhi Zhou, and Xu Chen (2019). "Boomerang: On-demand cooperative deep neural network inference for edge intelligence on the industrial Internet of Things". In: *IEEE Network* 33.5, pp. 96–103.

Zhang, Jie, Futai Zhang, Xin Huang, and Xin Liu (2020). "Leakage-Resilient Authenticated Key Exchange for Edge Artificial Intelligence". In: *IEEE Transactions on Dependable and Secure Computing*.

Zhang, Jun and Khaled B Letaief (2019). "Mobile edge intelligence and computing for the internet of vehicles". In: *Proceedings of the IEEE* 108.2, pp. 246–261.

Zhang, Shaojun, Wei Li, Yongwei Wu, Paul Watson, and Albert Zomaya (2018). "Enabling edge intelligence for activity recognition in smart homes". In: *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, pp. 228–236.

Zhang, Tiehua, Zhishu Shen, Jiong Jin, Xi Zheng, Atsushi Tagami, and Xianghui Cao (2020). "Achieving Democracy in Edge Intelligence: A Fog-based Collaborative Learning Scheme". In: *IEEE Internet of Things Journal*.

Zhang, Xingzhou, Yifan Wang, Sidi Lu, Liangkai Liu, Weisong Shi, et al. (2019). "OpenEI: An open framework for edge intelligence". In: *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, pp. 1840–1851.

Zhang, Yin, Xiao Ma, Jing Zhang, M Shamim Hossain, Ghulam Muhammad, and Syed Umar Amin (2019). "Edge intelligence in the cognitive Internet of Things: Improving sensitivity and interactivity". In: *IEEE Network* 33.3, pp. 58–64.

Zhang, Yiwen, Haishuai Guo, Zhihui Lu, Lu Zhan, and Patrick CK Hung (2020). "Distributed gas concentration prediction with intelligent edge devices in coal mine". In: *Engineering Applications of Artificial Intelligence* 92, p. 103643.

Zhang, Yushu, Hui Huang, Lu-Xing Yang, Yong Xiang, and Ming Li (2019). "Serious challenges and potential solutions for the industrial Internet of Things with edge intelligence". In: *IEEE Network* 33.5, pp. 41–45.

Zhang, Yongxu Zhu, Sabita Maharjan, and Yan Zhang (2019). "Edge intelligence and blockchain empowered 5G beyond for the industrial Internet of Things". In: *IEEE Network* 33.5, pp. 12–19.

Zhou, Zhi, Xu Chen, En Li, Liekang Zeng, Ke Luo, and Junshan Zhang (2019). "Edge intelligence: Paving the last mile of artificial intelligence with edge computing". In: *Proceedings of the IEEE* 107.8, pp. 1738–1762.

Zhu, Guangxu, Dongzhu Liu, Yuqing Du, Changsheng You, Jun Zhang, and Kaibin Huang (2020). "Toward an intelligent edge: wireless communication meets machine learning". In: *IEEE Communications Magazine* 58.1, pp. 19–25.

Zhu, Guangxu, Yong Wang, and Kaibin Huang (2019). "Broadband analog aggregation for low-latency federated edge learning". In: *IEEE Transactions on Wireless Communications* 19.1, pp. 491–506.

Zhu, Zheng, Yingjie Tian, Fan Li, Hongshan Yang, Zheng Ma, and Guoping Rong (2020). "Research on Edge Intelligence-based Security Analysis Method for Power Operation System". In: *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*. IEEE, pp. 258–263.

A Appendix A - GitHub Repository

The code for this study can be found at:

`https://github.com/sibr1011/edge_intelligence_taxonomy`.

The file *literature_process.xlsx* contains a complete list of hits found after applying the search strings. Besides, it shows the steps taken to reduce the list of entries described in figure 4 and reasons for inclusion or exclusion of papers.

The file *final_summary.xlsx* lists the papers, which remained after the relevance filtering process. The information saved includes "year", "author", "journal", "title" "link to paper" and other. This list of papers was used for the concept matrix and the taxonomy.

The file *iterationen.xlsx* contains the final taxonomy divided into its iterations. Each sample is separated by double strokes.

In addition, these files are saved in CSV-format for further data analysis. The repository also contains *R* code, which was used to generate the graphs for this study.

For any accessing problems please contact *siegel.brian93@gmail.com*.

B Appendix B - High-level Concept Matrix

Article (Author)	CONCEPTS			
	Algorithms & DNN design for EI	Distributed EI architectures, frameworks and systems	Other EI enabling technologies & theories	EI application scenarios
Zhou et al. (2019)	✓	✓	✓	
Wang et al. (2019)	✓	✓		
Yi Liu et al. (2019)	✓			✓
G. Zhu, D. Liu, et al. (2020)		✓		
E. Li et al. (2019)	✓			
Azar et al. (2019)			✓	
Zhang et al. (2019)			✓	
Jun Zhang and Letaief (2019)				✓
Yin Zhang et al. (2019)				✓
Yushu Zhang et al. (2019)			✓	
Muhammad et al. (2019)	✓			✓
Yazici, Basurra, and Gaber (2018)	✓			
Y. Shi et al. (2020)	✓	✓	✓	
L. Li, Ota, and M. Dong (2018)	✓		✓	
Rausch, Hummer, et al. (2019)	✓			
Mills, J. Hu, and Min (2019)		✓	✓	
F. Wang et al. (2019)	✓	✓	✓	
Alonso et al. (2020)				✓
S. Xu, Qian, and R. Q. Hu (2019)				✓
Plastiras et al. (2018)			✓	
Z. Huang, Lin, Tsai, et al. (2018)				✓
D. Liu et al. (2020)	✓		✓	
Yuan et al. (2019)			✓	
Zeng et al. (2019)	✓		✓	
J Pena Queralta et al. (2019)				✓
D. Xu et al. (2020)	✓	✓	✓	
T. Zhang et al. (2020)	✓			

Article (Author)	CONCEPTS			
	Algorithms & DNN design for EI	Distributed EI architectures, frameworks and systems	Other EI enabling technologies & theories	EI application scenarios
S. Xu, Qian, and R. Q. Hu (2020)				✓
Rausch and Dustdar (2019)		✓		✓
Mazzia et al. (2020)	✓			✓
Keshavarzi and Hoek (2019)			✓	
S. Zhang et al. (2018)				✓
Lv, D. Chen, and Qingjun Wang (2020)			✓	✓
Wen et al. (2019)	✓		✓	
Qureshi et al. (2020)				✓
Yiwen Zhang et al. (2020)	✓			✓
Qianlong Wang et al. (2020)		✓		
Yaqiong Liu et al. (2020)			✓	
Shuai Wang, R. Wang, et al. (2020)				✓
Tang et al. (2020)	✓		✓	
Libri, Bartolini, and Benini (2020)		✓		✓
Chakraborty et al. (2020)	✓		✓	
Z. Huang, Lin, and Shih (2016)		✓		
Lan et al. (2019)			✓	
Xiao et al. (2020)				✓
T. Wang et al. (2020)		✓		
X. Yang et al. (2020)			✓	
Skatchkovsky, Jang, and Simeone (2020)	✓	✓		
Gamanayake et al. (2020)	✓		✓	
Xia et al. (2020)				✓
Romaszkan, T. Li, and Gupta (2020)	✓			
Hung et al. (2020)			✓	
Jie Zhang et al. (2020)			✓	
Wolf (2019)		✓	✓	

Article (Author)	CONCEPTS			
	Algorithms & DNN design for EI	Distributed EI architectures, frameworks and systems	Other EI enabling technologies & theories	EI application scenarios
Alaslani and Shihada (2018)			✓	✓
Qu et al. (2020)		✓		
K. Yang et al. (2020)	✓	✓		
Y.-D. Huang et al. (2019)	✓			
Kaneko et al. (2019)	✓			
Kolomvatsos and Anagnostopoulos (2020)			✓	
Liang, Shenoy, and Irwin (2020)		✓		
T.-C. Chen et al. (2019)		✓		
Hannan Lodhi, Akgün, and Özkasap (2020)			✓	
Doku and Rawat (2020)			✓	
Pan et al. (2020)			✓	✓
S. Chen et al. (2019)				✓
L. Yang et al. (2020)		✓		
Oderhohwo et al. (2020)	✓			
Xie et al. (2019)	✓	✓		
Mittal and Bhushan (2020)			✓	
Z. Zhu et al. (2020)			✓	✓
C. Dong et al. (2020)		✓	✓	✓
Thangiah, Ramanathan, and Chodisetty (2019)		✓		
Mittal, Tyagi, and Bhushan (2020)				✓
Vatti, Vinoth, and Sneha (2020)				✓
Hoshino et al. (2019)			✓	
Z. Chen et al. (2019)			✓	
Mukherjee et al. (2020)			✓	
Corcoran et al. (2019)			✓	✓
Shuai Wang, Wu, et al. (2020)	✓			✓
Shiqiang Wang et al. (2018)	✓	✓		

Article (Author)	CONCEPTS			
	Algorithms & DNN design for EI	Distributed EI architectures, frameworks and systems	Other EI enabling technologies & theories	EI application scenarios
Rahman et al. (2020)			✓	
Park et al. (2019)			✓	
Pandit, Mir, and Chihisti (2017)	✓			
Brennan (2019)	✓			
Jorge Peña Queralta et al. (2020)				✓
YANG, Zhongyuan, and QUEK (2020)	✓		✓	
Gong et al. (2020)			✓	
Hao et al. (2020)	✓			
Guleng et al. (2020)				✓
G. Zhu, Yong Wang, and K. Huang (2019)	✓	✓		
G. Li et al. (2019)	✓	✓		
Skatchkovsky and Simeone (2019)	✓	✓		
Alanezi and Mishra (n.d.)		✓		
Larras and Frappé (2020)			✓	
D. Wang et al. (2019)			✓	

C Appendix C - Taxonomy Iteration 1

Paper	EI OPERATING EDGE NODE				
	Cameras	Vehicles	Other IoT Devices	Edge Server or Gateway	Edge-Device Synergy
Yi Liu et al. (2019)				X	
Jun Zhang and Letaief (2019)		X			
Yin Zhang et al. (2019)					X
Muhammad et al. (2019)	X				
Alonso et al. (2020)			X		
S. Xu, Qian, and R. Q. Hu (2019)	X				
Z. Huang, Lin, Tsai, et al. (2018)			X		
J Pena Queralta et al. (2019)			X		
S. Xu, Qian, and R. Q. Hu (2020)					X
Rausch and Dustdar (2019)				X	
Mazzia et al. (2020)	X				
S. Zhang et al. (2018)				X	
Lv, D. Chen, and Qingjun Wang (2020)		X			
Qureshi et al. (2020)				X	
Yiwen Zhang et al. (2020)				X	X
Shuai Wang, R. Wang, et al. (2020)				X	
Libri, Bartolini, and Benini (2020)				X	
Xiao et al. (2020)			X		
Xia et al. (2020)				X	
Alaslani and Shihada (2018)				X	
Pan et al. (2020)			X		
S. Chen et al. (2019)			X		
Z. Zhu et al. (2020)				X	
C. Dong et al. (2020)			X		
Mittal, Tyagi, and Bhushan (2020)	X				X
Vatti, Vinoth, and Sneha (2020)			X		
Corcoran et al. (2019)	X				
Jorge Peña Queralta et al. (2020)					X
Guleng et al. (2020)	X				

D Appendix D - Taxonomy Iteration 2

Paper	EI OPERATING EDGE NODE					EI TRAINING SCHEME				
	C	**V**	**OID**	**ES/G**	**EDS**	**OD**	**FL**	**DEPC**	**C/PT**	**PL**
Yi Liu et al. (2019)				X				X		
Jun Zhang and Letaief (2019)		X					X			
Yin Zhang et al. (2019)					X			X		
Muhammad et al. (2019)	X								X	
Alonso et al. (2020)			X					X		
S. Xu, Qian, and R. Q. Hu (2019)	X								X	
Z. Huang, Lin, Tsai, et al. (2018)			X							X
J Pena Queralta et al. (2019)			X					X		
S. Xu, Qian, and R. Q. Hu (2020)					X		X			
Rausch and Dustdar (2019)				X			X	X	X	X
Mazzia et al. (2020)	X					X				
S. Zhang et al. (2018)				X				X		
Lv, D. Chen, and Qingjun Wang (2020)		X						X		
Qureshi et al. (2020)				X						
Yiwen Zhang et al. (2020)				X	X		X			
Shuai Wang, R. Wang, et al. (2020)				X						X
Libri, Bartolini, and Benini (2020)				X					X	
Xiao et al. (2020)			X					X		
Xia et al. (2020)				X					X	
Alaslani and Shihada (2018)				X				X		
Pan et al. (2020)			X			X				
S. Chen et al. (2019)			X						X	
Z. Zhu et al. (2020)				X				X	X	
C. Dong et al. (2020)			X				X			
Mittal, Tyagi, and Bhushan (2020)	X				X				X	
Vatti, Vinoth, and Sneha (2020)			X						X	
Corcoran et al. (2019)	X							X		
Jorge Peña Queralta et al. (2020)					X					
Guleng et al. (2020)	X							X		
Zhou et al. (2019)					X		X			
Wang et al. (2019)				X	X		X	X		
G. Zhu, D. Liu, et al. (2020)				X				X		
E. Li et al. (2019)			X					X		
Yazici, Basurra, and Gaber (2018)			X			X				
Y. Shi et al. (2020)					X			X		

L. Li, Ota, and M. Dong (2018)			X						X	
Rausch, Hummer, et al. (2019)			X			X				
Mills, J. Hu, and Min (2019)					X		X			
F. Wang et al. (2019)					X	X				
D. Liu et al. (2020)				X				X		
Zeng et al. (2019)					X			X		
D. Xu et al. (2020)			X	X	X		X	X		
T. Zhang et al. (2020)					X			X		
Wen et al. (2019)				X			X	X		
Qianlong Wang et al. (2020)			X				X			
Tang et al. (2020)					X			X		
Chakraborty et al. (2020)			X					X		
Z. Huang, Lin, and Shih (2016)			X							X
T. Wang et al. (2020)					X			X		
Skatchkovsky, Jang, and Simeone (2020)					X		X			
Gamanayake et al. (2020)					X			X		
Romaszkan, T. Li, and Gupta (2020)										X
Wolf (2019)					X		X			
Qu et al. (2020)					X		X			
K. Yang et al. (2020)			X					X		
Y.-D. Huang et al. (2019)			X			X				
Kaneko et al. (2019)			X			X				
Liang, Shenoy, and Irwin (2020)					X			X		
T.-C. Chen et al. (2019)			X			X				
L. Yang et al. (2020)					X			X		
Oderhohwo et al. (2020)				X				X		
Xie et al. (2019)					X			X		
Thangiah, Ramanathan, and Chodisetty (2019)				X					X	
Shuai Wang, Wu, et al. (2020)					X			X		
Shiqiang Wang et al. (2018)					X			X		
Pandit, Mir, and Chihisti (2017)			X				X			
Brennan (2019)					X			X		
YANG, Zhongyuan, and QUEK (2020)					X		X			
Hao et al. (2020)					X			X		
G. Zhu, Yong Wang, and K. Huang (2019)					X		X			
G. Li et al. (2019)				X					X	
Skatchkovsky and Simeone (2019)			X							X
Alanezi and Mishra (n.d.)					X			X		

E Appendix E - Taxonomy Iteration 3

Paper	EIOEN					EITS					PEETT						
	C	V	OID	ES/G	EDS	OD	FL	DEPC	C/PT	PL	D&CS	DF&C	EEC	EC	MEC	V	ERA
Yi Liu et al. (2019)				X				X				X	X		X		X
Jun Zhang and Letaief (2019)		X					X							X			
Yin Zhang et al. (2019)					X			X			X				X		
Muhammad et al. (2019)	X								X				X				
Alonso et al. (2020)			X					X				X					
S. Xu, Qian, and R. Q. Hu (2019)	X								X		X				X		
Z. Huang, Lin, Tsai, et al. (2018)			X							X							X
J Pena Queralta et al. (2019)			X					X				X					
S. Xu, Qian, and R. Q. Hu (2020)					X		X				X						
Rausch and Dustdar (2019)				X			X	X	X	X		X			X		
Mazzia et al. (2020)	X					X							X				
S. Zhang et al. (2018)				X				X							X		
Lv, D. Chen, and Qingjun Wang (2020)		X						X					X	X			
Qureshi et al. (2020)				X							X						
Yiwen Zhang et al. (2020)				X	X		X					X		X			
Shuai Wang, R. Wang, et al. (2020)				X						X							X
Libri, Bartolini, and Benini (2020)				X					X		X				X		
Xiao et al. (2020)			X					X									X
Xia et al. (2020)				X					X								
Alaslani and Shihada (2018)				X				X			X						
Pan et al. (2020)			X			X							X				
S. Chen et al. (2019)			X						X		X						
Z. Zhu et al. (2020)				X				X	X			X					
C. Dong et al. (2020)			X				X							X	X		
Mittal, Tyagi, and Bhushan (2020)	X				X				X			X					
Vatti, Vinoth, and Sneha (2020)			X						X						X		
Corcoran et al. (2019)	X							X					X				
Jorge Peña Queralta et al. (2020)					X										X		
Guleng et al. (2020)	X							X						X	X		
Zhou et al. (2019)					X		X				X	X		X			

Yazici, Basurra, and Gaber (2018)			X			X							X		X		
Y. Shi et al. (2020)					X			X				X					
L. Li, Ota, and M. Dong (2018)			X						X							X	
Rausch, Hummer, et al. (2019)			X			X									X		X
Mills, J. Hu, and Min (2019)					X		X					X					
F. Wang et al. (2019)					X	X						X					
D. Liu et al. (2020)				X				X									X
Zeng et al. (2019)					X			X				X					
D. Xu et al. (2020)			X	X	X		X	X			X	X	X	X	X	X	X
T. Zhang et al. (2020)					X			X					X				
Wen et al. (2019)				X			X	X							X		
Qianlong Wang et al. (2020)			X				X				X						
Tang et al. (2020)					x			X									x
Chakraborty et al. (2020)			X					X					X				
Z. Huang, Lin, and Shih (2016)			X							X			X		X		
T. Wang et al. (2020)					X			X						X			
Skatchkovsky, Jang, and Simeone (2020)					X		X						X				
Gamanayake et al. (2020)					X			X				X					
Romaszkan, T. Li, and Gupta (2020)										X		X					
Wolf (2019)					X		X									X	
Qu et al. (2020)					X		X								X		
K. Yang et al. (2020)			X					X					X				
Y.-D. Huang et al. (2019)			X			X							X				
Kaneko et al. (2019)			X			X							X				
Liang, Shenoy, and Irwin (2020)					X			X				X					
T.-C. Chen et al. (2019)			X			X							X				
L. Yang et al. (2020)					X			X							X		
Oderhohwo et al. (2020)				X				X									X
Xie et al. (2019)					X			X					X				
Thangiah, Ramanathan, and Chodisetty (2019)				X					X						X		
Shuai Wang, Wu, et al. (2020)					X			X					X				
Shiqiang Wang et al. (2018)					X			X									X
Pandit, Mir, and Chihisti (2017)			X				X						X				
Brennan (2019)					X			X									X
YANG, Zhongyuan, and QUEK (2020)					X		X					X					
Hao et al. (2020)					X			X					X				
G. Zhu, Yong Wang, and K. Huang (2019)					X		X						X				
G. Li et al. (2019)				X					X					X			

Wang et al. (2019)				X	X		X	X			X			X			X
G. Zhu, D. Liu, et al. (2020)				X				X				X					
E. Li et al. (2019)			X					X					X				
Skatchkovsky and Simeone (2019)			X							X			X				
Alanezi and Mishra (n.d.)					X			X							X		
Azar et al. (2019)		X						X				X	X				
Zhang et al. (2019)					X			X			X			X			X
Yushu Zhang et al. (2019)					X			X			X		X				
Plastiras et al. (2018)					X			X									X
Yuan et al. (2019)				X					X			X					
Keshavarzi and Hoek (2019)			X										X				
Yaqiong Liu et al. (2020)			X					X							X		
Lan et al. (2019)				X			X						X				
X. Yang et al. (2020)					X		X						X				
Hung et al. (2020)			X			X							X				
Jie Zhang et al. (2020)			X			X					X						
Kolomvatsos and Anagnostopoulos (2020)					X			X				X					
Hannan Lodhi, Akgün, and Özkasap (2020)					X		X						X				
Doku and Rawat (2020)					X		X				X						
Mittal and Bhushan (2020)	X				X			X							X		
Hoshino et al. (2019)					X			X					X				
Z. Chen et al. (2019)				X					X						X		
Mukherjee et al. (2020)				X					X								X
Rahman et al. (2020)				X					X		X						
Park et al. (2019)			X	X	X		X	X	X			X	X	X	X		
Gong et al. (2020)					X			X			X						
Larras and Frappé (2020)			X						X			X					
D. Wang et al. (2019)					X		X					X					X

Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate deutlich kenntlich gemacht zu haben.
Ich erkläre weiterhin, dass die vorliegende Arbeit in gleicher oder ähnlicher Form noch nicht im Rahmen eines anderen Prüfungsverfahrens eingereicht wurde.

Würzburg, den July 25, 2022 Brian Siegel