Artificial intelligence and privacy in Sub-Saharan Africa — A systematic literature review

Group 7
Charlotte Langhout, Daryl Zandvliet, David Huser, Kelsey Doolittle, Saskia Woortman
Interdisciplinary Research Methodology for Information Sciences (X_405085)
Vrije Universiteit Amsterdam

Abstract

Background: Artificial intelligence (AI) has been disrupting the developed world. In low- and middle-income countries (LMICs) such as those located in Sub-Saharan Africa (SSA), the literature promises a wide range of possible use cases. However, less is known about negative impacts specifically on citizens with regards to data protection and privacy of citizens in SSA.
Aims: This systematic literature review aims to provide a high-level view of the current research on Big Data collection and the negative implications of using Big Data and AI in SSA. Also, the current practices and policies to protect the privacy and data of citizens together with the concerning challenges are reviewed.
Method: The authors used search queries related to AI, SSA, and privacy to find relevant papers via VU LibSearch, Web of Science and Scopus. Additionally, articles referenced by experts were included.
Results: 21 Studies resulted from the search. The most mentioned type of data collected was data originating from mobile phones. Only around half of the included studies actively discussed negative impacts, and only three studies mentioned more than one principle or practice helping to protect the privacy and/or data protection of citizens in SSA. The majority of the discussion was around Big Data and the effects, rather than AI.
Conclusion: AI implementations could have negative impacts related to the privacy concerns of civilians partially due to data colonization of Western corporations. Guidelines such as FAIR Principles or Principles for Digital Development are well-supported by researchers but not yet widely adopted in SSA. More research into privacy- and context-aware AI is required in the future.

Keywords: Artificial Intelligence, Sub-Saharan Africa, privacy, data protection, Big Data

Introduction

Motivation

Artificial Intelligence (AI) has gained traction around the developed world but is still in its infancy in low- and middle-income countries (LMIC) including those within Sub-Saharan Africa (SSA). The issue of AI is of particular interest because many view it as a powerful resource and those who possess it will hold the greatest power (Miailhe, 2018). AI within Africa is of interest because 7 out of 15 of the fastest-growing economies are in SSA (FocusEconomics, 2019).

A lot of data is being collected already by telecommunication companies and government ministries as well as various research and international development projects. The aim of this study is to explore which types of data are being collected on a large scale within SSA, which are, or could be input for AI applications. In regards to data protection and privacy of citizens of SSA countries, major risks and challenges arise due to the nature and scale of collection. This study also aims to scope out the extent of practices and policies used to constrain and regulate data collection and analysis within AI.

AI, LMICs and privacy

AI, as an academic discipline, was founded over 50 years ago and has experienced a variety of definitions over this period. AI, in the scope of this research, is defined as "the study of devices that perceive their environment and define a course of action that will maximize its chance of achieving a given goal." (Poole et al., 1998). Some examples of AI include advanced algorithms and artificial neural networks which can make predictions alike to human logical decision making on things such as creditworthiness (Nalubega, 2019) or housing prices (Abidoye, 2017).

Another example is the use of drones or other kinds of intelligent robots (Nalubega, 2019), or applications which utilize deep learning, machine learning, natural language processing, and speech recognition (Reis & Melao, 2019). The term AI is often used to describe machine learning and vice versa (Andersen, 2019).

Privacy is a multifaceted concept. A possible definition is "...the collection and handling of personal data, i.e. information identifying a natural living person… This includes how such personal data is gathered, registered, stored, exploited, and disseminated" (Makulilo, 2016).

LMICs, defined using the 2019 thresholds from the World Bank, are classified as having a gross national income per capita below US$3,995 (The World Bank, 2019).

SSA can be defined as the area lying south of the Sahara desert (United Nations, 2003). In regard to this research however, the decision was made to not include South Africa, as the country has experienced substantial growth in AI research (Ferrein & Meyer, 2012) and AI implementations.

Opportunities for AI in Sub-Saharan Africa

Applications of AI systems in SSA have the ability to solve real-life problems and can be applied in multiple sectors of society. One example of its use within healthcare is the utilization of advanced classification techniques — a paper of Hao (2019) reports about the application of a machine-learning model that classifies 42 different types of cancer automatically. The successful prototype dramatically reduced the amount of work for human experts, who needed to classify the diseases manually before. Another application is in the agricultural domain, for example, a diagnostic tool for smartphones that helps farmers diagnosing viral crop diseases in Cassava plants (Dogo et al., 2019; Kao, 2019; Web Foundation, 2017).
Also, economical benefits may arise with the emergence of AI. Multiple recent studies examined how natural language processing and other AI-based technology could provide mobile banking services for those who have difficulties accessing or interacting with browser-based online banking systems (Dogo et al., 2019; Web Foundation, 2017).
Lastly, AI has the potential to address the problem of illiteracy in LMICs in Africa. Using automated translation and voice recognition systems, AI allows people to engage with governmental or public service provision interfaces (Web Foundation, 2017).

Method

The methodology of this study is based on the guidelines for systematic literature reviews (Kitchenham, 2004).

Research question

RQ: How could AI (AI) potentially impact Sub-Saharan Africa (SSA) with regards to data protection and privacy of citizens?

Sub-questions

RQ1: Which type of data is collected or expected to be collected on a massive scale in SSA?
RQ2: What are the actual and potential negative impacts of Big Data in SSA?
RQ3: What are the current and proposed practices and policies used in SSA to protect the data and privacy of citizens?
RQ4: What are the challenges of current practices and policies?

Search process

Multiple sources were used for the search process. For this study, only the databases of VU LibSearch, Scopus and Web of Science plus an expert's mentioned references have been used as described in Table 1.

Data source	Name of source	Date of search	Search property
Electronic database	VU LibSearch	14-11-2019	Abstract
Electronic database	Web of Science	14-11-2019	Topic: Searches title, abstract, author keywords and keywords plus
Electronic database	Scopus	14-11-2019	Article title, abstract, keywords
Expert's references	Dr. Anna Bon	Continuous	-

Table 1. Data sources

The study only includes three databases—a study by Buchanan and Bryman (2009) has shown that reducing the number of databases causes a higher level of transparency and makes it easier to reproduce the results. The authors additionally searched four Information Communication Technology for Development (ICT4D) journals: Information Technology and Development, Information Technology and International Development, Electronic Journal of Information Systems in Developing Countries, and the Journal of Community Informatics. However, no results were returned, thus the journals were not included in the data sources.

Search query

The aim of the search query is to filter through the articles available in the databases. Table 2 describes how the specific keywords were selected according to the concepts identified as highly important to answer the research questions. To compose the search terms and limit the scope of this data collection, the subsequent steps were used:

1. Identify concepts from the research question
2. Identify keywords related to the concept, as well as synonyms
3. Look up the validity of keywords in a few relevant papers
4. Group concepts with the boolean operator 'AND'
5. Link each concepts' keywords with 'OR'.

Category	Concept	Keywords
ICT	Artificial Intelligence	Artificial Intelligence Data science Machine learning Big data
Location	Sub-Saharan Africa	Africa Developing country Low- and middle-income country
Subject	Data protection and privacy	Privacy Data protection Regulation Policy Policies

Table 2. Search category, concepts, and keywords

All keywords have been merged with an 'OR' operator to include them in the search query. The operator 'AND' has been used to combine the keywords from different categories. This process resulted in the following search query:

("Artificial Intelligence" OR "data science" OR "Big Data" OR "machine learning")
AND
("Africa" OR "developing country" OR "low- and middle-income country")
AND
("data protection" OR "privacy" OR "regulation" OR "policy" OR "policies")

Paper selection process

The quality of the papers was assessed based on a paper selection process, as described in the steps below:

1. *Identification:* The search query was entered into three data sources, according to the previous definition.
2. *Deduplication*: A deduplication was necessary due to duplicates in the Web of Science and Scopus. VU LibSearch has been the master data source to compare duplicates against.
3. *Screening:* The papers were divided among the authors and screened:
    a. Each author read the title and abstract assigned papers and reviewed if the papers fitted the inclusion and exclusion criteria.
    b. The papers were exported using Endnote and organized within a spreadsheet.
    c. Another author confirmed or challenged the decision to avoid selection bias.

    d. Updated (exclusion) criteria to decrease the number of articles.
4. *Eligibility:* Evaluated the quality and relevance of the papers according to the quality assessment.
5. *Inclusion:* The remaining papers (n=21) which have met the inclusion criteria and quality assessment were analyzed in the data collection and analysis process.

A graphical overview of the search process is depicted in Figure 1.

Inclusion and exclusion criteria

Inclusion area

- Publication date between 01-01-2000 and 14-11-2019 (due to little or no development of ICT facilities in Africa before the year 2000 which could enable medium to large computations required for this study's topic)
- Literature must be a paper or book
- Papers about developing countries in general, or LMICs, when applicable to SSA countries as well

Exclusion area

- Papers about specific countries not in Sub-Saharan Africa, such as China, South Africa or Algeria
- Papers not written in English
- Papers not findable or openable
- Conference papers

Quality assessment

Each paper was evaluated based on the following quality assessment questions related to the relevancy to the posed research questions:

1. Does the article cover collection on the scale of Big Data, and/or does it mention AI technologies or methodologies?
2. Does the article mention any actual or potential negative impacts of using Big Data or AI?
3. Does the article mention any practices or policies related to privacy and/or data protection of citizens?

Based on the above questions, we scored each paper on the number of matching topics:
irrelevant (0), low (1), medium (2), high (3). We included papers that achieved a score of at least 2 (medium or high).

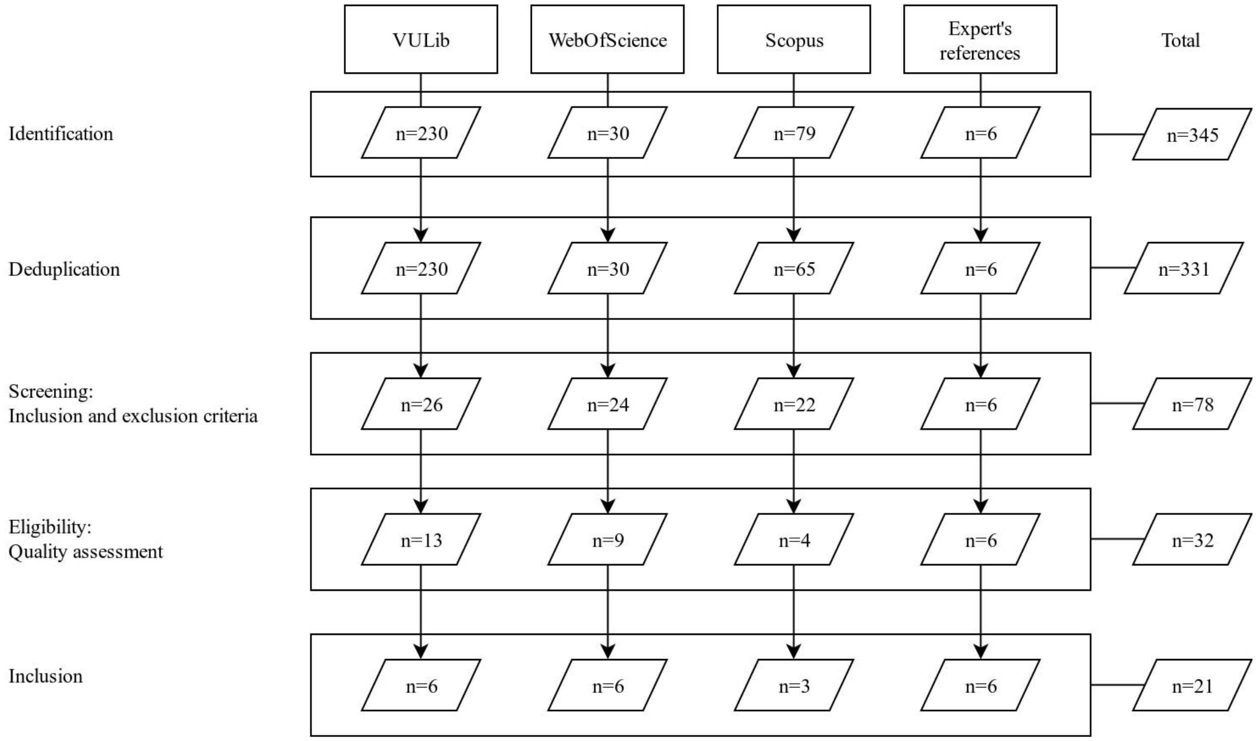

Figure 1. Search process and number of papers per each step and source

Data collection

We extracted data from each paper for both meta-level information as well as relevant answers to our research question.

Data extracted as meta-information:

- Source of paper (i.e. the journal it was published in)
- Publication year
- Name of the authors
- Research question or issue

Data extracted to answer the research question:

- Type of technology, type of data collected
- Nation or country where the study was relevant to (origin country of the data source, not the country of analysis)
- Stated actual or potential negative impacts of the technology or processes
- Practices and policies to protect the data and privacy and citizens

Data analysis

Extracted data was analyzed and categorized, e.g. within the identified actual or potential negative impacts, different subcategories such as faulty prediction or business exploitation were created as columns. The authors split up and reread all papers with regards to the identification of those specific subcategories and mapped each paper to the mentioned sub-categories in a new data analysis spreadsheet (see Appendix).

Results

After creating the data analysis, overviews of these subjects were created. These are shown below in tables 3, 4, 5 and 6. Following the tables are elaborations on the particular data mentioned in the studies. The tables and elaborations help to answer the main research question by providing all the necessary information for the subquestions.

In summary, 17 out of 21 papers discussed 14 types of technologies and data. Moreover, 14 out of 21 papers discussed 15 actual and potential negative impacts of these technologies. Lastly, 9 out of 21 papers mentioned 7 different practices and policies to protect the data and privacy of citizens, while only 4 papers mentioned 8 challenges of practices and policies.

Many papers discussed topics without studying the topic, for instance, the discussion around drones was fairly abundant, but none of the papers read contained a study or experiment around drones. Also, the reader will sometimes find one particular instance of a topic is in regards to an LMIC that is not within SSA, while the paper itself was still included because elsewhere SSA was discussed. It is important to keep this in mind when reviewing the following results.

Type of technology and data collected

The most common type of data used and/or discussed was mobile phone data. Mobile phone data refers to call detail records which contain only information regarding the time and location a person sends or receives a text message or call (Amankwah-Amoah, 2016; Dogo et al., 2019; Mann, 2018; Nalubega & Uwizeyimana, 2019; Pokhriyal & Jacques, 2017; Taylor, 2017; Taylor, 2016; Taylor & Broeders, 2015; The Economist, 2018; Vogel et al., 2015). Both mobile phones and smartphones have great potential for their future use in areas such as agriculture and healthcare alongside satellite imagery (Andersen, 2019). Smartphone data is similar to mobile phone data, with the added bonus of applications (Amankwah-Amoah, 2016; Andersen, 2019; Dogo et al., 2019; Pokhriyal & Jacques, 2017; The Economist, 2018) which for example can help farmers better connect with their market, increase productivity on the farm, as well as track behavior through social media sites (Mann, 2018).

Social media data, such as data from Facebook, has a great number of privacy concerns (Shozi &

Mtsweni, 2017). It can provide the location of people in real-time (Amankwah-Amoah, 2016; Taylor & Broeders, 2015), as well as information on popular beliefs (Mann, 2018).

Digital banking information refers to monetary transactional data (Head et al., 2017), which can be accessed through a biometric identification number for every citizen in India (Taylor & Broeders, 2015). Alongside mobile phone data and smartphone data, digital banking information can be used as a potential evaluation of a person's creditworthiness (Mann, 2018; Nalubega & Uwizeyimana, 2019; The Economist, 2018).

Drones are essentially flying robots and include general robotics (Miailhe, 2018; Taylor, 2017), as well as specific uses of drones to deliver Internet to LMICs (Taylor & Broeders, 2015) or deliver blood to rural areas (Nalubega & Uwizeyimana, 2019).

Health-related data was mostly referring to using mobile phone data to track the movement of sick people in order to better predict and prevent the spread of disease (Hao, 2019; Head et al., 2017; Nalubega & Uwizeyimana, 2019; Vogel et al., 2015) which sometimes involved satellite images as well (Amankwah-Amoah, 2016; Taylor, 2017).

Blockchain is a technology used to address security concerns, such as for the protection of financial information (Dogo et al., 2019; Nalubega & Uwizeyimana, 2019).

Search results are records pulled from people's search history to gather information around their interests and preferences and is particularly useful for marketing to better target customers (Mann, 2018; Taylor & Broeders, 2015).

Satellite images are images produced from satellites (Amankwah-Amoah, 2016; Andersen, 2019; Pokhriyal & Jacques, 2017; Vogel et al., 2015) and alongside mobile-phone data were used to follow general human movement (Taylor, 2017; Taylor, 2016; Taylor & Broeders, 2015). One paper was a study proving that a neural network created from satellite images could estimate the wealth of individuals in small villages (Head et al., 2017).

Food and agricultural data allows farmers to have better stock of their crops to effectively communicate with buyers, keep track of the chains of distribution, and increase their creditworthiness (Mann, 2018; Nalubega & Uwizeyimana, 2019; Pokhriyal & Jacques, 2017). Algorithms are used to predict and prevent crop disease (Andersen, 2019; Hao, 2019).

Sensor data is used for detecting the space occupation and mobility of humans (Taylor & Broeders, 2015; Dogo et al., 2019; Mann, 2018).

Interviews were a type of data collection tool used to discuss the broad impact of AI (Web Foundation, 2017) or Big Data (Taylor, 2016; Mann, 2018) on LMIC, as well as the potentially missed economic opportunities within Africa.

In general, a wide variety of data types such as search results, sensor data, and biometric information were discussed simply to illustrate the corporate value of data in Africa (Mann, 2018; The Economist, 2018).

Lastly, polls are an example of the misuse of data in elections (Fayoyin & Ngwainmbi, 2014).

Type of technology and data collected	Study
Drones and robots	Miailhe, 2018; Nalubega & Uwizeyimana, 2019; Taylor, 2017; Taylor & Broeders, 2015
Blockchain	Dogo et al., 2019; Nalubega & Uwizeyimana, 2019
Search results	Mann, 2018; Taylor & Broeders, 2015
Social media data	Amankwah-Amoah, 2016; Mann, 2018; Shozi & Mtsweni, 2017; Taylor & Broeders, 2015
Mobile phone data	Amankwah-Amoah, 2016; Andersen, 2019; Dogo et al., 2019; Mann, 2018; Nalubega & Uwizeyimana, 2019; Pokhriyal & Jacques, 2017; Taylor, 2017; Taylor, 2016; Taylor & Broeders, The Economist, 2018; Vogel et al., 2015
Smartphone data	Amankwah-Amoah, 2016; Andersen, 2019; Dogo et al., 2019; Pokhriyal & Jacques, 2017; Mann, 2018; The Economist, 2018
Satellite images	Amankwah-Amoah, 2016; Andersen, 2019; Head et al., 2017; Pokhriyal & Jacques, 2017; Taylor, 2017; Taylor, 2016; Taylor & Broeders, 2015; Vogel et al., 2015
Digital banking information	Head et al., 2017; Mann, 2018; Nalubega & Uwizeyimana, 2019; Taylor & Broeders, 2015; The Economist, 2018
Food and agricultural data	Andersen, 2019; Hao, 2019; Mann, 2018; Nalubega & Uwizeyimana, 2019; Pokhriyal & Jacques, 2017
Sensor data	Dogo et al., 2019; Mann, 2018; Taylor & Broeders, 2015
Biometric information	Mann, 2018; The Economist, 2018
Health-related data	Amankwah-Amoah, 2016; Hao, 2019; Head et al., 2017; Nalubega & Uwizeyimana, 2019; Taylor, 2017; Vogel et al., 2015
Interviews	Mann, 2018; Taylor, 2016; Web Foundation, 2017
Polls	Fayoyin & Ngwainmbi, 2014

Table 3 Overview type of technology and data collection and corresponding studies

Actual and potential negative impacts

The table below sheds light on which papers discuss the negative impacts of the technology and data collected. The impact mentioned most was faulty prediction, meaning that by using certain technology or data, inaccurate predictions can be made (Benedikter, 2019; Nalubega & Uwizeyimana, 2019; Taylor & Broeders, 2015). This can include the evaluation of someone's creditworthiness, using data from social media which can lead to discrimination (Mann, 2018) by using parameters such as level of education and income (Andersen, 2019; Taylor, 2017).

Technology and data can also lead to increased inequality, as vulnerable groups such as women can be disproportionately affected (Tomalin & Ullman, 2019; Web Foundation, 2017). Moreover, the data collected is generally not stored in LMICs countries, which also leads to increased inequality since one group has more influence than others (Taylor & Broeders, 2015). As this information is often owned by foreign companies, the power these companies have is reinforced (Mann, 2018; Miailhe, 2018). Consequently, these companies gain access to power more easily, which leads to power being centralized (Taylor & Broeders, 2015).

Concluding that a form of colonization is at play because the power largely is not held by Africans (Mann, 2018; Miailhe, 2018).

People's perceptions, as well as public policies, can change from technology and data (Head et al., 2017; Miailhe, 2018; Taylor, 2017). For example, people may vote differently, thus the level of democracy may decline and corruption can rise (Benedikter, 2019; Fayoyin & Ngwainmbi, 2014; Miailhe, 2018; Web Foundation, 2017).

Another negative impact of these modern technologies may be that they can be used for personal identification (Andersen, 2019; Mann, 2018, The Economist, 2018).

Even in cases where information is removed on the individual level, there are methods to identify an individual as belonging to a specific group (Mann, 2018; Taylor 2017; Taylor 2016). Both types of identification enable surveillance (Andersen, 2019; Mann 2018; Taylor 2017; Taylor 2016). This data could be used for value evasion, which means the added value of technology use in LMICs will be captured by large corporations, including platforms like Google or Facebook (Mann, 2018; Miailhe, 2018). This is also linked to business exploitation, which means companies have commercial motivation for collecting data and exploiting it (Mann, 2018; The Economist, 2018). Exploitation could happen by using the data to shape people's behavior or by dominating new markets (Mann, 2018; Taylor & Broeders, 2015).

A well-known phenomenon is function creep. This refers to the scenario when data collected for a specific purpose is used for other, less well-intended purposes (Taylor, 2016). Also, Fayoyin and Ngwainmbi (2014) stated there is a misuse of data in advocacy and journalism. Another potential negative impact is the risk of fabricated outputs as the output of a machine learning algorithm is often not questioned. Andersen (2019) states that the output is sometimes far from objective.

The implementation of technology could also influence the job market in Africa. Two articles show a concern regarding job availability since technology could be replacing jobs (Knott, 2019; Web Foundation, 2017). This could also be linked to over-reliance on AI by humans and consequently free themselves of a responsibility (Andersen, 2019; Web Foundation, 2017).

Lastly, civilians could be persuaded to use technologies, which happened in Sierra Leone where the government switched to digital payment providers, thus the employees were required to open a bank account (The Economist, 2018).

Actual and potential negative impacts	Study
Policy influence	Head et al., 2017; Miailhe, 2018; Taylor, 2017
Fabricated evidence	Andersen, 2019
Business exploitation	Mann, 2018; Taylor & Broeders, 2015; The Economist, 2018
Personal identification	Andersen, 2019; The Economist, 2018; Mann, 2018
Surveillance	Andersen, 2019; Mann, 2018; Taylor, 2017; Taylor, 2016
Group identification	Mann, 2018; Taylor, 2017; Taylor, 2016
Function creep [from good to bad uses]	Fayoyin & Ngwainmbi, 2014; Taylor, 2016
Faulty prediction	Andersen, 2019; Benedikter, 2019; Mann, 2018; Nalubega & Uwizeyimana, 2019; Taylor, 2017; Taylor & Broeders, 2015
Centralization of power	Mann, 2018; Miailhe, 2018; Taylor & Broeders, 2015
Persuasion and addiction	The Economist, 2018
Increased inequality	Mann, 2018; Miailhe, 2018; Taylor & Broeders, 2015; Tomalin & Ullmann, 2019; Web Foundation, 2017
Over-reliance on AI	Andersen, 2019; Web Foundation, 2017
Weakening democracy / increasing corruption	Benedikter, 2019; Fayoyin & Ngwainmbi, 2014; Miailhe, 2018; Web Foundation, 2017
Value evasion (to developed countries)	Mann, 2018; Miailhe, 2018
Job stability and security	Knott, 2019; Web Foundation, 2017

Table 4 Overview of actual and potential negative impacts and corresponding studies

Practices and policies

Table 5 shows papers that contain information on the practices and policies for protecting the data and privacy of citizens in SSA countries. Anonymization techniques are perused to review social media sites. These include suppression, generalization, aggregation, data swapping, random noise, and synthetic data (Shozi & Mtsweni, 2017). Anonymized call detail records were used for research purposes (Taylor & Broeders, 2015; Vogel et al., 2015). In some papers, anonymized data meant the removal of demographic information (Taylor, 2017).

National data protection laws are implemented in less than 50% of African nations (Shozi & Mtsweni, 2017). The laws differ greatly amongst nations, and there is no one truth around data protection within SSA as a whole.

Bias awareness can be defined as the practice of developers to consider their own potential biases before creating an AI system. This tool is used often within public sector monitoring and evaluation (Nalubega & Uwizeyimana, 2019). Sometimes companies consider their algorithms "bias aware" simply because machine learning algorithms use math, and thus are seen as being without bias from human input (Andersen, 2019). Pokhriyal & Jacques (2017) were aware of biases affecting the results in their study about poverty in Senegal.

Principles for Digital Development are important within the ICT4D community. The seven principles discussed were: design with the user; understand the ecosystem; design for scale; build for sustainability; be data-driven; use open standards, open data, open source, open innovation; and reuse/improve (Andersen, 2019). Similar to these are the FAIR principles, a framework for the collection, use, and reuse of data (van Reisen et al., 2019).

UN Global Pulse, the World Economic Forum (WEF) and the GSMA all play a role in their efforts to change the practices and policies around this data globally. Global Pulse is a data lab based in New York with operations out of Uganda and Indonesia. The GSMA is an organization that represents mobile operators worldwide (Mann, 2018).

Practices and policies	Study
National data protection laws	Shozi & Mtsweni, 2017
Anonymization	Shozi & Mtsweni, 2017; Taylor, 2017; Taylor & Broeders, 2015; Vogel et al., 2015
Principles for Digital Development	Andersen, 2019; van Reisen et al., 2019
Bias awareness	Andersen, 2019; Pokhriyal & Jacques, 2017; Nalubega & Uwizeyimana, 2019
UN Global Pulse	Mann, 2018
World Economic Forum	Mann, 2018
GSMA	Mann, 2018

Table 5. Overview of contemporary practices and policies and corresponding studies

Challenges of practices and policies

Table 6 gives an overview of current challenges related to the practices and policies mentioned in table 5. Responsibility touches upon the problem of ownership of an AI system when multiple parties are involved in the development. Related is the problem of accountability which means it is not always stated who is ultimately accountable for the outcomes of AI systems. Lastly, the policy should include how the explainability of the system is ensured in order to be sure how an AI system comes to its results (Andersen, 2019). Challenges mentioned regarding data protection laws include the lack of enforceability of laws (Taylor, 2016) and the lack of policy coordination within and between governments (van Reisen et al., 2019). Regarding principles for digital development, the problem of information isolation causes governance models to be not fully inclusive (Dogo et al., 2019). Besides, principles and policies from Western countries could not be copied and implemented because they lack the perspective of African countries (Dogo et al., 2019; van Reisen et al., 2019). Lastly, a challenge regarding the practice of anonymization is that it can be re-identified with the use of Big Data due to the ability to cross-match with non-personal data points in the same data source or combined with other data sources (Taylor, 2016).

Challenges of practices and policies	**Study**
Re-identification (anonymization)	Taylor, 2016
Lack of enforceability	Taylor, 2016
Lack of coordination	van Reisen et al., 2019
Importing from Western countries	Dogo et al., 2019; van Reisen et al., 2019
Information isolation	Dogo et al., 2019
Accountability	Andersen, 2019
Responsibility	Andersen, 2019
Explainability	Andersen, 2019

Table 6. Overview of the challenges related to practices and policies and corresponding studies

Discussion

The following discussion was restricted by the limitations of the assignment at hand, forcing the exclusion of further research.

Another limitation is that while AI is a broad field, not many papers covered research on AI systems in Sub-Saharan Africa, thus limiting the scope and possibly the external validity of the research. Even those which did cover AI did not necessarily discuss privacy concerns or data protection of citizens. The papers on Big Data did capture privacy concerns, but whether or not one can extrapolate those issues from Big Data to AI can be debated.

The papers collected for this research cover a wide range of AI techniques and also gather many different types of data. Therefore, the findings of the study cannot be generalized for every type of system using AI. The challenges and dangers identified in these papers are not applicable for every type of technology.

It is hard to make predictions about the negative effects of AI on privacy issues based on the literature review simply because Big Data is not always related to AI algorithms. The practices and policies addressed in the results section may not be applicable in real-life, because most of the proposed policies are not examined in practice in the papers of this research. Another limitation of the implementation of the proposed policies is the dependency on data protection laws. Data protection laws are implemented in less than 50% of African nations and the lack of policy coordination within and between governments hinders the effectiveness of the proposed practices and policies. However, the challenges of the practices and policies can help to implement AI systems in African nations by taking the predefined problems into account.

Conclusion

Africa is becoming more digital and concerns are growing around the invasion of personal privacy. Data collected and used within AI technology could be used by Western corporations, as well as foreign countries, who are above all interested in the monetary value and potential power. This puts the citizens of SSA at risk since they are likely unaware their data is collected and by whom. The lack of policy within individual nations leaves people vulnerable to data-exploitation and data-colonization.

The hope is with the increased adoption of technology in SSA, policy implementation will follow. Governments within SSA could potentially build upon FAIR principles, Principles for Digital Development, as well as build in privacy and security into the technological design. However, local African citizens should be included in this process to make the principles relevant to what is needed in SSA, since the AI landscape is currently biased towards extra continental forces. Thus, more research into privacy- and context-aware AI in SSA is required in the future.

References

Barocas, S. (2014). Data mining and the discourse on discrimination. In Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining.

Bryman, A. (2016). Interviewing in qualitative research. Social research methods. New York: Oxford University Press, 5, 465-499.

Buchanan, D., & Bryman, A. (2009). The Sage handbook of organizational research methods. Sage Publications Ltd.

Ferrein, A., & Meyer, T. (2012). A Brief Overview of Artificial Intelligence in South Africa. Artificial Intelligence Magazine, 33(1), 99-103.

FocusEconomics. (2019). The Fastest Growing Economies in the World (2019-2023). Retrieved December 5, 2019, from
https://www.focus-economics.com/blog/fastest-growing-economies-in-the-world.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*(2004), 1-26.

Lehmacher, W., Betti, F., Beecher, P., Grotemeier, C., & Lorenzen, M. (2017). Impact of the Fourth Industrial Revolution on Supply Chains. In World Economic Forum, Geneva, Switzerland, REF (pp. 061117-00035015).

Machanavajjhala, A., & Reiter, J. P. (2012). Big privacy: protecting confidentiality in Big Data. XRDS: Crossroads, The ACM Magazine for Students, 19(1), 20-23.

Makulilo, A. B. (2016). African Data Privacy Laws. Springer International Publishing.

Poole, D., Mackworth, A., & Goebel, R. (1998). Computational Intelligence: a logical approach.

Reis, J., Santo, P. E., & Melao, N. (2019). Impacts of AI on Public Administration: A Systematic Literature Review. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI).

The World Bank. (2019). World Bank Country and Lending Groups. Retrieved December 5, 2019, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

United Nations. (2003). Millennium Development Indicators: World and regional groupings. Retrieved December 5, 2019, from
https://unstats.un.org/unsd/mi/africa.htm.

References Search Process

Amankwah-Amoah, J. (2016). Emerging economies, emerging challenges: Mobilising and capturing value from Big Data. *Technological Forecasting and Social Change*, *110*, 167-174.

Andersen, L. (2019). Artificial Intelligence in International Development: Avoiding Ethical Pitfalls, *Journal of Public and International Affair*s. Retrieved December 4, 2019, from https://jpia.princeton.edu/news/artificial-intelligence-international-development-avoiding-ethical-pitfalls.

Benedikter, R. (2019). New technologies can be a force for good in Africa if they're developed from the ground. Retrieved December 4, 2019, from https://theconversation.com/new-technologies-can-be-a-force-for-good-in-africa-if-theyre-developed-from-the-ground-121916.

Dogo, E. M., Salami, A. F., Aigbavboa, C. O., & Nkonyana, T. (2019). Taking Cloud Computing to the Extreme Edge: A Review of Mist Computing for Smart Cities and Industry 4.0 in Africa. In *Edge Computing* (107-132). Springer, Cham.

The Economist. (2018). Exclusive access. *The Economist, 427*, S3-S4. Retrieved December 4, 2019, from https://search-proquest-com.vu-nl.idm.oclc.org/docview/2034648398?accountid=10978

Fayoyin, A., & Ngwainmbi, E. (2014). Use and misuse of data in advocacy, media and opinion polls in Africa: Realities, challenges and opportunities. *Journal of Development and Communication Studies*, *3*(2), 528-545.

Hao, K. (2019). The future of AI research is in Africa. Retrieved December 4, 2019, from https://www.technologyreview.com/s/613848/ai-africa-machine-learning-ibm-google/.

Head, A., Manguin, M., Tran, N., & Blumenstock, J. E. (2017). Can human development be measured with satellite imagery? In *ICTD* (pp. 8-1).

Knott, S. (2019). African Women in Tech Look to Artificial Intelligence. Retrieved December 4, 2019, from https://www.voanews.com/africa/african-women-tech-look-artificial-intelligence.

Mann, L. (2018). Left to other peoples' devices? A political economy perspective on the Big Data revolution in development. *Development and Change*, *49*(1), 3-36.

Miailhe, N. (2018). The geopolitics of Artificial Intelligence: The return of empires? *Politique étrangère*, (3), 105-117.

Nalubega, T., & Uwizeyimana, D. E. (2019). Public sector monitoring and evaluation in the Fourth Industrial Revolution: Implications for Africa. *Africa's Public Service Delivery and Performance Review*, *7*(1), 1-12.

Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences*, *114*(46), E9783-E9792.

Shozi, N. A., & Mtsweni, J. (2017). Big data privacy in social media sites. In *2017 IST-Africa Week Conference (IST-Africa)* (pp. 1-6). IEEE.

Taylor, L. (2017). Safety in numbers? Group privacy and Big Data analytics in the developing world. In *Group Privacy* (pp. 13-36). Springer, Cham.

Taylor, L. (2016). No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environment and Planning D: Society and Space*, *34*(2), 319-336.

Taylor, L., & Broeders, D. (2015). In the name of development: Power, profit and the datafication of the global South. *Geoforum*, *64*, 229-237.

Tomalin, M., & Ullmann, S. (2019). AI could be a force for good – but we're currently heading for a darker future. Retrieved December 4, 2019, from https://theconversation.com/ai-could-be-a-force-for-good-but-were-currently-heading-for-a-darker-future-124941.

van Reisen, M., Stokmans, M., Mawere, M., Basajja, M., Ong'ayo, A. O., Nakazibwe, P., Kirkpatrick, C. & Chindoza, K. (2019). FAIR Practices in Africa. *Data Intelligence*, 246-256.

Vogel, N., Theisen, C., Leidig, J. P., Scripps, J., Graham, D. H., & Wolffe, G. (2015). Mining mobile datasets to enable the fine-grained stochastic simulation of Ebola diffusion. *Procedia Computer Science*, *51*, 765-774.

Web Foundation. (2017). Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries. *A Smart Web for a More Equal Future,* (1-20)

Appendix

			Type of technology and data collected														Actual and potential negative impacts															Practices and policies							Challenges with practices and policies							
Author(s)	**Year**	**Title**	Drones and robots	Blockchain	Search results	Social media data	Mobile phone data	Smartphone data	Satellite images	Digital banking information	Food and agricultural data	Sensor data	Biometric information	Health-related data	Interviews	Polls	Policy influence	Fabricated evidence	Business exploitation	Personal identification	Surveillance	Group identification	Function creep [from good to bad uses]	Faulty prediction	Centralization of power	Persuasion and addiction	Increased inequality	Over reliance on AI	Weakening democr. / incr. corrupt.	Value evasion (to developed countries)	Job stability and security	National data protection laws	Anonymization	Principles for Digital Development	Bias awareness	UN Global Pulse	World Economic Forum	GSMA	Re-identification	Lack of enforceability	Lack of coordination	Importing from Western countries	Information isolation	Accountability	Responsibility	Explainability
Knott, S.	2019	African Women in Tech Look to Artificial Intelligence																													x															
Tomalin, M., & Ullmann, S.	2019	AI could be a force for good – but we're currently heading for a darker future																									x																			
Shozi, N. A., & Mtsweni, J.	2017	Big data privacy in social media sites				x																										x	x													
Head et al.	2017	Can Human Development be Measured with Satellite Imagery?							x	x				x			x																													
Fayoyin, A., & Ngwainmbi, E.	2014	Use and Misuse of Data in Advocacy, Media, and Opinion Polls in Africa: Realities, Challenges, and Opportunities														x							x						x																	
Web Foundation	2017	Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries													x												x	x	x		x															
Amankwah-Amoah, J.	2016	Emerging economies, emerging challenges: Mobilising and capturing value from big data				x	x	x	x					x																																
Dogo et al.	2019	Taking Cloud Computing to the Extreme Edge: A Review of Mist Computing for Smart Cities and Industry 4.0 in Africa		x			x	x				x																														x	x			
Hao, K.	2019	The future of AI research is in Africa									x			x																																
Anderson, L	2019	Artificial Intelligence in International Development: Avoiding Ethical Pitfalls					x	x	x		x							x		x	x			x				x						x	x									x	x	x
Pokhriyal, N., & Jacques, D. C.	2017	Combining disparate data sources for improved poverty prediction and mapping					x	x	x		x																								x											
The Economist	2018	Exclusive access					x	x		x			x						x	x						x																				
Mann, L.	2018	Left to Other Peoples' Devices? A Political Economy Perspective on the Big Data Revolution in Development			x	x	x	x		x	x	x	x		x				x	x	x	x		x	x		x			x						x	x	x								
Taylor, L.	2016	No place to hide? The ethics and analytics of tracking mobility using mobile phone data					x		x						x						x	x	x																x	x						
Nalubega, T., & Uwizeyimana, D. E.	2019	Public sector monitoring and evaluation in the Fourth Industrial Revolution: Implications for Africa	x	x			x			x	x			x										x											x											
Taylor, L.	2017	Safety in numbers? Group privacy and big data analytics in the developing world	x				x		x					x			x				x	x		x									x													
Miailhe, N.	2018	The geopolitics of artificial intelligence: The return of empires?	x														x								x		x		x	x																
Taylor, Broeders	2015	In the name of development: Power, profit and the datafication of theglobal South	x		x	x	x		x	x		x							x					x	x		x						x													
Vogel et al.	2015	Mining Mobile Datasets to Enable the Fine-grained Stochastic Simulation of Ebola Diffusion					x		x					x																			x													
Benedikter, R.	2019	New technologies can be a force for good in Africa if they're developed from the ground																			x								x																	
Van Reisen et al.	2019	FAIR Practices in Africa																																x							x	x				