UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The Potential of Big Data Technologies for the Human Rights and Environmental Due Diligence Process

Duval, A.; Rouas, V.; Ebdon, M.

Publication date
2023

Document Version
Final published version

Link to publication

Citation for published version (APA):
Duval, A., Rouas, V., & Ebdon, M. (2023). *The Potential of Big Data Technologies for the Human Rights and Environmental Due Diligence Process*. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

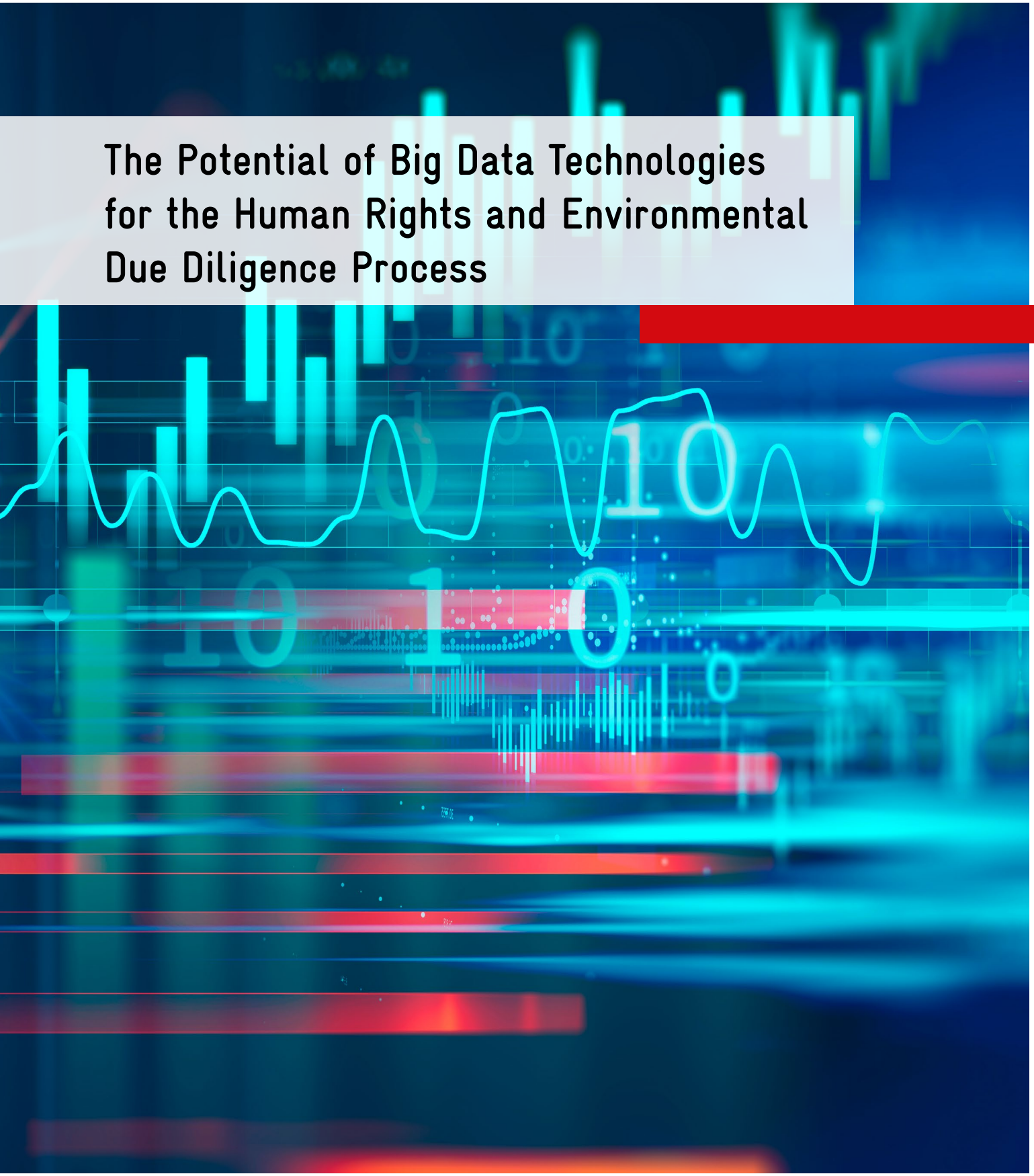
General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:07 Oct 2024

The Potential of Big Data Technologies for the Human Rights and Environmental Due Diligence Process



Deutsche Gesellschaft
für Internationale
Zusammenarbeit (GIZ) GmbH

This study was prepared by:

ASSER INSTITUTE

On behalf of:

IGS INITIATIVE FOR GLOBAL SOLIDARITY

Published by:

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn, Germany

Friedrich-Ebert-Allee 32
53113 Bonn, Deutschland

T +49 228 44 60-0
F +49 228 4460-17 66

E info@giz.de
I www.giz.de

Project:
Inititiative for Global Solidarity

Author:
Dr. Antoine Duval, Dr. Virginie Rouas, Max Ebdon

Design:
Tinkerbelle GmbH, Berlin

Photo credits:
Cover: istock/monsitj

URL-links:
Responsibility for the content of external websites linked in this publication always lies with their respective publishers. GIZ expressly dissociates itself from such content.

On behalf of German Federal Ministry for Economic Cooperation and Development (BMZ)

Directorate 120 – Sustainable transformation of global supply chains,
Reinhard Junker, Berlin

GIZ is responsible for the content of this publication.

May 2023

Table of Contents

Executive Summary		**4**
List of abbreviations		**5**
1.	**Introduction**	**6**
1.1	Objective of the study	6
1.2	Methodology	6
1.3	Structure	6
2.	**Big Data and Big Data Technologies: A Short Introduction**	**7**
2.1	What is big data?	7
2.2.	Big data technologies	8
2.3.	Preconditions for using big data	10
3.	**Big Data Technologies and HREDD**	**11**
3.1.	What is HREDD?	11
3.2.	Using big data technologies in HREDD	12
3.3.	Using big data technologies in supervising compliance with HREDD obligations	21
4.	**Risks and Challenges in Using Big Data Technologies for HREDD.**	**24**
4.1.	Limits to the effectiveness of big data technologies in the HREDD process	24
4.2.	Potential incompatibilities of the use of big data technologies with some requirements of the HREDD process	25
4.3.	Adverse human rights and environmental impacts linked to the use of big data technologies	27
4.4.	The risks of conflicts of interests and the use of big data technologies in HREDD	28
5.	**Conclusions**	**29**
6.	**Annexes**	**30**
6.1.	Bibliography	30
6.2.	List of interviewees	34
6.3.	Overview of relevant big data technology solutions	35

Executive Summary

This study is the first to investigate the potential role of big data technologies in the HREDD process. While big data technologies are already widely used by businesses in order to increase their economic efficiency, these efficiencies could potentially be transferred to the HREDD process. In particular, big data technologies could help businesses navigate the scale and complexity of contemporary supply chains as well as the challenges associated with analysing a wide range of potential adverse impacts.

Moreover, from the point of view of regulators and civil society organisations, big data technologies could facilitate monitoring the compliance with HREDD obligations of a large number of corporations. The aim of this study was thus to map the current usages of big data technologies in the HREDD process, as well as to discuss potential challenges that might arise in this context. The study shows that there are various steps of the HREDD process in which big data technologies are already used or could be of use.

First, we have concluded that big data technologies can play a role in enabling corporation to map much more precisely their extensive and complex upstream and downstream supply chains.

Second, our research shows that big data technologies can play a role at the stage of detecting and assessing human rights and environmental risks or adverse impacts linked to global supply chains. In both of these examples the main added value provided is in the ability of big data technologies to navigate the size and complexity of modern supply chains and to process vast amounts of data in real time covering a wide range of risks.

The third potential use explored in the study concerns the potential of big data technologies to prove beneficial in supporting the monitoring of business compliance with HREDD obligations. More specifically, these technologies could offer advantages in scrutinizing extensive corporate due diligence reports that contain large amounts of qualitative (textual and visual) data.

Nevertheless, the study also identified a number of potential risks and challenges connected to the reliance on big data technologies in the framework of the HREDD process. More specifically, the study outlined challenges related to the effectiveness of these technologies, to their relationship with specific requirements of the HREDD process, such as transparency and stakeholders' engagement, as well as to the human rights and environmental risks linked to their use and the potential conflicts of interests of the companies developing these technologies.

List of abbreviations

AI	Artificial Intelligence
BHRRC	Business & Human Rights Resource Centre
CSR	Corporate Social Responsibility
DIHR	Danish Institute for Human Rights
ESIA	Environmental and social impact assessments
EIA	Environmental impact assessments
EU	European Union
GPS	Global Positioning System
GRI	Global Reporting Initiative
HS Codes	Harmonised System Codes
HREDD	Human rights and environmental due diligence
HRDD	Human rights due dilaigence
HRIA	Human rights impact assessments
IT	Information Technology
ILO	International Labour Organization
IoT	Internet of Things
KYC	Know your consumer
KDD	Knowledge Discovery in Databases
NGO	Non-governmental Organisation
OHCHR	Office of the United Nations High Commissioner for Human Rights
OECD	Organisation for Economic Co-operation and Development
SDG	Sustainable Development Goals
UN	United Nations
UNGPs	United Nations Guiding Principles on Business and Human Rights

1. Introduction

This study was prepared on behalf of the Initiative for Global Solidarity (IGS). The IGS is a project funded by the German Federal Ministry for Economic Cooperation and Development and implemented by the Deutsche Gesellschaft for International Cooperation (GIZ) GmbH. IGS works to implement human rights and environmental due diligence in global supply chains by scaling up and harmonising proven approaches. The IGS enables buying and supplying companies particularly in the garment and electronics sector to exercise shared responsibility for the impacts of their business practices on people and the planet.

1.1 Objective of the study

In the last five years there has been a growing trend of mandatory HREDD legislation across Europe. Most notably, the EU's two biggest economies, France and Germany, have imposed a duty to conduct the HREDD process onto their biggest corporations, while the EU institutions are currently in the final stages of the adoption of a corporate sustainability due diligence directive. Due diligence as a process aimed at preventing human rights harms caused by or linked to businesses was first introduced in the UNGPs in 2011. The substantial scope of application of the due diligence process was widened beyond human rights by its integration into the 2011 OECD Guidelines for Multinational Enterprises. Since then, it has become a touchstone in regulatory initiatives aimed at the responsibility of corporations for the adverse human rights and environmental impacts linked to their transnational economic activities.

The emergence of complex transnational supply chains, which are difficult to control for a single state, is a central background condition to the emergence and spread of HREDD as a regulatory strategy. It is also the reason why conducting HREDD throughout these supply chains in order to identify and prevent adverse impacts linked to a particular product or service will be an arduous task, which will involve analysing thousands of businesses across several continents for a broad range of human rights and environmental risks. It is against this backdrop that this study aims to map the role that big data technologies could play in the HREDD process. Big data technologies are already being used by businesses in order to increase efficiency and thus profits. The hypothesis behind the present study is that these efficiencies could potentially be transferred to the HREDD process in order to support corporations in tackling the scale and complexity of contemporary supply chains as well as the challenges associated with analysing a wide range of potential risks. While, from the point of view of public regulators, big data technologies could facilitate monitoring the compliance with HREDD obligations of a large number of corporations. Our purpose is not to analyse the effectiveness of these big data technologies but to draw links between claims about what these technologies could do and the main steps of the HREDD process, as well as to highlight potential challenges that could arise due to the systematic use of big data technologies in this context.

1.2 Methodology

This study's methodology is grounded in two main processes of collection of qualitative information. First, a comprehensive interdisciplinary literature review was conducted in order to map the existing scholarship on the application of big data technologies to issues directly or indirectly related to HREDD (such as supply chain sustainability, big data for development and human rights) and derive interesting insights that could be transposable to the application of big data technologies in the HREDD context. Second, we conducted a series of semi-structured interviews with a variety of stakeholders, including representatives of companies offering big data technologies services for sustainable supply chains and, in some cases, HREDD; academics and NGOs working on new technologies and business and human rights; and HREDD consultants. These interviews helped identify the current state of use of big data technologies in HREDD processes, as well as the potential opportunities and challenges that may arise from their use.

This study was conducted by legal scholars with business and human rights expertise, we have relied on secondary sources to describe technological processes. Therefore, our contribution is limited to linking claims made by companies and in the literature to the specific requirements of the HREDD process. However, we are not in a position to assess the validity of claims regarding the effectiveness of specific technical solutions.

1.3 Structure

This study will begin in Section 2 with a brief introduction to big data technologies, discussing the four main types of technology and some of the preconditions for the use of such technologies. It will then turn in Section 3 to the core subject matter of this report, the contribution of big data technologies to the HREDD process. In this section, we start by briefly outlining what the HREDD process entails for companies. Thereafter, we discuss the general potential of big data technologies in the context of the HREDD process before outlining three potential use-cases. These use-cases focus on the application of big data technologies to map transnational supply chains, to identify human rights and environmental risks and to monitor the compliance of companies with HREDD obligations. The use-cases are followed in Section 4 by a selection of potential risks and challenges linked to the use of such technologies in relation to the HREDD process. The study concludes in Section 5 by offering some concrete recommendations.

2. Big Data and Big Data Technologies: A Short Introduction

Big data is a term that has had significant use over the last two decades. At its most basic level big data refers to a massive amount of data which can be analysed and used to make decisions.[1] However, this definition does not capture the broad nature of the term big data nor does it point towards any of its actual and potential uses. In this section the meaning of big data will be defined, the various 'big data technologies' will be discussed as well as the preconditions for the effective use of big data.

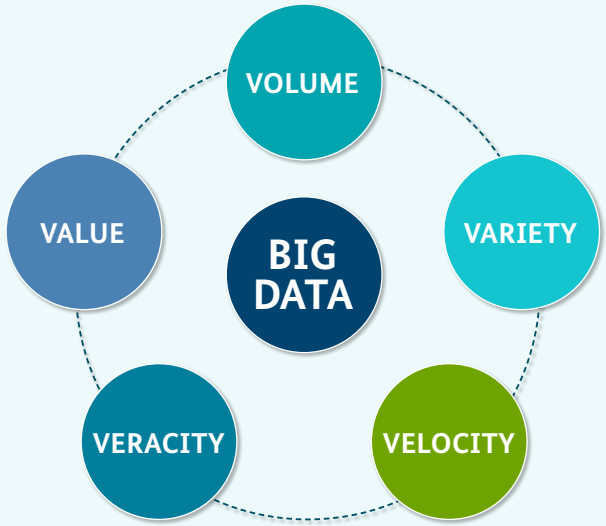
2.1 What is big data?

Big data is a term that was coined by data analyst Douglas Laney in 2001 in his paper '3D Data Management: Controlling Data Volume, Velocity and Variety'.[2] His definition includes three essential concepts: the three V's of Volume, Velocity and Variety.

Volume refers to the sheer amount of data being collected. Before computing power reached its current level, businesses and governments collected a lot of data, but storing it was challenging. Nowadays, increased computing capacity means that data storage is no longer a problem. Moreover, there is an increasing amount of data collected through a variety of sources and processes. Data is constantly being collected via the internet through social media, online retail or online streaming sites or in person through sensors and cameras. It is now routine for data scientists and those working in big data to come across data sets with trillions of elements[3] - such is the volume of data now being collected. Velocity refers to the speed at which data is collected.[4] Whereas data used to be 'lagged', meaning there would be a period of time between actions generating data (e.g. browsing an online store) and data storage and analysis, now data is being collected in real time and at incredibly fast rates.[5] Velocity involves the rapid production and exchange of the huge volume of data in limited time spans.[6]

Variety refers to the type of data being collected. Whereas in the past the type of data being collected was basic (e.g. demographic and geographical information), there is now a much wider range being collected.[7] Additionally, the range of sources from which data is collected and processed has drastically expanded.[8] For example, the political consulting firm Cambridge Analytica, which made use of big data to target undecided voters, claims to have had, using only data from social media, 4,000–5,000 different data points for every voter in the US upon which it could evaluate individuals and offer targeted messaging.[9] The three V's Laney introduced have now been enriched by two further V's: Veracity

Fig. 1 | The five V's of big data defined by Douglas Laney



1 Patrick C. K. Hung, Big Data Applications and Use Cases (Springer Cham 2016).

2 Doug Laney, '3D Data Management: Controlling Data Volume, Velocity and Variety' (META Group 2001). This definition has also been used by the UN Global Pulse in UN Global Pulse, 'Big data for Development: Challenges and Opportunities' (May 2012)

3 Kristian Kersting and Ulrich Meyer, 'From Big Data to Big Artificial Intelligence? Algorithmic Challenges and Opportunities in Big Data' (2018) 32 Künstliche Intelligenz 3.

4 Angela C. Lyons and John Grable J, 'An Introduction to Big Data' (2018) 72(5) Economics & Investment Management 17, at 17.

5 Ibid.

6 Kristian Kersting and Ulrich Meyer, 'From Big Data to Big Artificial Intelligence? Algorithmic Challenges and Opportunities in Big Data' (2018) 32 Künstliche Intelligenz 3.

7 Angela C. Lyons and John Grable J, 'An Introduction to Big Data' (2018) 72(5) Economics & Investment Management 17, at 17.

8 Kristian Kersting and Ulrich Meyer, 'From Big Data to Big Artificial Intelligence? Algorithmic Challenges and Opportunities in Big Data' (2018) 32 Künstliche Intelligenz 3.

9 'Cambridge Analytica parent firm SCL Elections fines over data refusal' (BBC News, 10 January 2019) <https://www.bbc.co.uk/news/technology-46822439> accessed 12 January 2023.

and Value.[10] Veracity refers to how the accuracy of the data inevitably affects the quality of the final result – poorly collected data will mean the analysis is less accurate. Value expresses the ability of big data to create business opportunities by generating values from the processing of information.[11] The 5 V's of Volume, Velocity, Variety, Veracity and Value can help to provide an understanding of what is meant by big data but are of less use concerning how big data is used and which technologies make use of the vast resources of data available.

2.2. Big data technologies

Big data refers only to the amount of data that is now accessible but this data is made effective through big data technologies which make use of this data and provide an outcome or insight. There are four important types of big data technologies: data mining, data storage, data analytics, and data visualisation. It is through these four types of big data technology that big data has become a driver of innovation.[12]

Fig. 2 | Types of big data technologies

2.2.1 Big data mining

Big data mining refers to the extraction of useful patterns and trends from raw data.[13] Other terms used for data mining are Knowledge Discovery in Databases (KDD), knowledge extraction, data archaeology, and data dredging.[14] The data mining process involves a number of steps, from data collection to extracting useful information from large data sets. The Internet of Things (IoT) has become the primary grounds for data mining. IoT refers to the embedded sensory and network capable devices which are now a part of everyday life.[15] Wearable technology, smart home devices, and smartphones are examples of IoT technology; these devices record and track various data points that can then be analysed. They work by communicating together and storing information in the cloud where it can be accessed by data analytic tools. The growing trend of inserting IoT devices into everyday life is known as pervasive computing. [16] Data is also gathered from the internet, with social media being one of the most prominent examples as companies like Facebook and Twitter use insights gained from the minutia of its users behaviour to provide targeting advertising. [17] Data mined from the internet is not limited to social media, though, practically any website can be mined to provide insights. [18] There are many different tools through which data mining occurs; one example is text-mining whereby artificial intelligence 'reads' text and can then sort documents by different criteria.[19]

10 Angela C. Lyons and John Grable J, 'An Introduction to Big Data' (2018) 72(5) Economics & Investment Management 17, at 17; Cornelia Hammer, Diane C. Kostroch and Gabriel Quiros 'Big Data: Potential, Challenges, and Statistical Implications' IMF Staff Discussion Note 2017, https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106, accessed 31 March 2023

11 Kristian Kersting and Ulrich Meyer, 'From Big Data to Big Artificial Intelligence? Algorithmic Challenges and Opportunities in Big Data' (2018) 32 Künstliche Intelligenz 3.

12 Patrick C. K. Hung, Big Data Applications and Use Cases (Springer Cham 2016).

13 Cornelia Hammer, Diane C. Kostroch and Gabriel Quiros 'Big Data: Potential, Challenges, and Statistical Implications' IMF Staff Discussion Note 2017, https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106, accessed 31 March 2023, at 6.

14 Feng Chen et al, 'Data Mining for the Internet of Things: Literature Review and Challenges' (2015) 11(8) International Journal of Distributed Sensor Networks.

15 Patrick C. K. Hung, Big Data Applications and Use Cases (Springer Cham 2016).

16 Ibid.

17 Natasha Singer, 'What You Don't Know About How Facebook Uses Your Data', The New York Times, 11 April 2018 <https://www.nytimes.com/2018/04/11/technology/facebook-privacy-hearings.html> Accessed 20 January 2023.

18 ChatGPT is an example of artificial intelligence utilising a broad range of online sources: https://openai.com/blog/chatgpt/

19 D. Ganesh and P. Kalpana, 'Supply chain risk identification: a real-time data-mining approach' (2022) 122(5) Industrial Management & Data Systems 1333.

2.2.2. Big data storage

As previously stated, big data involves the use of vast quantities of data – more data than traditional methods of storage, such as local hard disk drives, can hold. The ability to store more data has been instrumental in the growth of big data technologies, it has allowed analytics tools to work with a large enough data pool to provide useful insights in various sectors and industries. Big data storage comprises infrastructure which is designed to fetch, store, and manage vast amounts of data. In the context of big data storage, there are two important technologies: Blockchain and Cloud Computing. These technologies allow for data storage and are offered by various businesses, such as Apache Hadoop and MongoDB, through their own software.

Blockchain

Blockchain is a technology that is used to store data in the safest way possible. It is a distributed ledger system that operates in a decentralised environment, allowing it to protect the large amounts of data generated. [20] Blockchain works by placing data in a decentralised ledger, this data is broken up into shared blocks which are 'chained' together with unique identifiers in the form of cryptographic hashes which can be thought of as digital fingerprints identifying the block and its content. This makes the data secure as, due to being decentralised, no single individual can modify the data unilaterally and, due to being chained together via the hashes, there is a record of modification to the data as any modification creates a new hash and adds a new block to the chain. Blockchain was created in 1991 and was intended to timestamp digital documents so they could not be backdated or edited. The three main features of blockchain are that it is decentralised, immutable and built on consensus. Decentralisation refers to the fact that data is transferred from the control of a centralised entity (individual, organisation or group) to a decentralised network – meaning the data is shared over many peer-to-peer networks where each network contains the blockchain. Immutability refers to the fact that the blockchain cannot be changed or altered – this occurs because of the hash system and the fact it is decentralised. Consensus refers to the fact that data can only be added and taken away from the blockchain when the majority of participants in the network give their consent. Blockchain provides a safe and secure way to store big data where it is accessible to all parties but is safe from being tampered with.

Cloud computing

Cloud computing provides an infrastructure which is playing an important role by providing organisations with the ability to store data economically and efficiently. [21] It is used by many organisations irrespective of size because of an adapted pay-per-use model thus providing scalable and cost-effective solutions. Cloud computing at its simplest means data is stored via the internet. A business choosing to 'move to the cloud' means its IT infrastructure is stored offsite at a data centre that is managed and maintained by a cloud computing provider – this is attractive as it offers agility, scale and flexibility without the need for large upfront investment. In the context of big data, cloud computing allows for masses of data to be stored without the need for the upfront investment – once on the cloud the data can be accessed and analysed to provide insights.

2.2.3 Big data analytics

Technologies are used in big data analytics to clean and transform data into information that is actionable, in other words that can be used to drive decisions. This is accomplished through the use of algorithms, models, and specific tools, such as Apache Spark and Splunk. Big data analytics can be divided into four categories: **descriptive analytics**, which helps to create reports and visualise information; **diagnostic analytics**, which can explain why a problem has occurred; **predictive analytics**, which uses past and present data to make predictions; and **prescriptive analytics**, which can provide solutions to problems by relying on AI and machine learning to gather data and use it for risk management. Underlying these types of data analytics is artificial intelligence (AI). AI has become a catchall phrase for applications that complete complex tasks that were once done by humans and is often used interchangeably with its subfields like machine learning and deep learning. AI refers to systems or machines that can mimic human intelligence to perform tasks and can iteratively improve themselves based on the information they collect. AI is, therefore, central to big data analytics and making use of big data gathered through data mining.

Machine learning is a term that is often used in literature on AI and is frequently used interchangeably with deep learning but these two technologies are different, with deep learning being a sub-field of machine learning. The way they learn differs, with 'deep' machine learning able to use both labelled and unlabelled data sets; this means it can ingest unstructured data in its raw form, such as texts and images. Classical or 'non-deep' machine learning, on the other hand, requires more human intervention to learn. Humans must determine the set of features to understand the differences between data inputs. Machine learning and deep learning are two important technologies that aid the data analysis process.

20 Fadi Muheidat et al, 'Emerging Concepts Using Blockchain and Big Data' (2022) 198 Procedia Computer Science 15.

21 Mohamed Bakhouya et al, 'Cloud computing, IoT, and big data: Technologies and applications' (2020) 32(17) Current and Computation: Practice and Experience 159.

2.2.4 Big data visualisation

Data visualisation tools are used to make the outcome of the analysis process easy to understand for humans, this can come in the forms of graphs, text or simply by showing the relevant data. Data visualisation tools retrieve data from the cloud, block-chain or local storage in order to make results understandable. Visualisation is often a part of the service offered by data analysis technologies or can be done externally.

2.3. Preconditions for using big data

There are two main preconditions for making meaningful use of big data: data quantity and data quality. First, the various big data technologies discussed previously need a large amount of data in order to become more accurate and hence produce better results. This is particularly true for data analytics technologies that use AI. However, this comes with a downside as there are drawbacks to data hoarding. Data hoarding is a situation in which data is collected and stored but never utilised to create insights. To begin with, storing data is expensive – whether stored locally, the cloud, or on blockchain, there can be significant economic and environmental costs. Data hoarding also costs time and money spent on experts who need to structure the data – this is especially true when machine learning, which requires labelled or structured data, is used. Deep learning, on the other hand, does not require the previously mentioned data structuring. Data hoarding also means that more energy is used by the various technologies. For example, machine learning, particularly deep learning, consumes a great amount of energy. There must be a balancing act between having enough data to gain accurate insights on the one hand and not using an unnecessary and expensive amount of data on the other.

Second, data quality is an increasingly large problem in big data. Due to online retail, social media, and IoT devices, there is usually no shortage of data, but whether this data is of the desired quality is another matter. Mining data from multiple sources means managing different data types and complex structures, which makes data integration more difficult. Judging data quality is also a resource heavy exercise, particularly for unstructured data, which makes up over 80 % of the total data in existence – this adds to the cost and the time needed to apply big data technologies.

3. Big Data Technologies and HREDD

The core aim of this study is to determine whether big data technologies could play a role in the context of the HREDD process. This section will first briefly elucidate what implementing an HREDD process entails for corporations and then elaborate on the potential contribution of big data technologies in this regard.

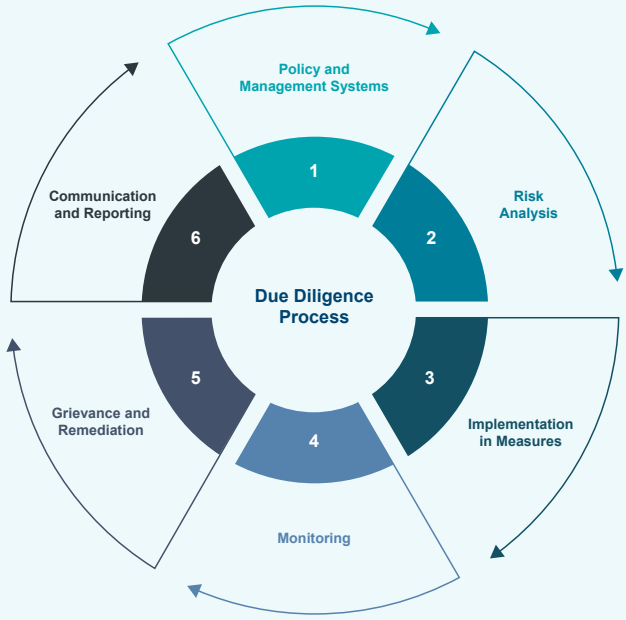
3.1. What is HREDD?

The notion of due diligence is not new, it has been in use in the business context in Anglo-Saxon countries for a long time, where it is referring to the process of evaluation of risks for a particular business and is generally conducted in the context of the acquisition of a company. However, this is not the type of due diligence at the heart of this report. Instead, the study is focused on a different sort of due diligence altogether, a process aimed originally at preventing human rights harms enshrined at the heart of the UNGPs by the former U.N. Special Representative on Business and Human Rights, Prof. John Ruggie.[22] In this section, we will rely on the definition of due diligence for responsible business conduct provided by the OECD, which is widely referred to and remains directly inspired by the UNGPs.[23]

The main difference between HREDD and traditional corporate due diligence is that the former is not focused on the material risks faced by a business, but instead is concentrated on adverse impacts for people and the environment which are linked to the activities of a particular business. HREDD is not limited to the adverse impacts caused or contributed to by a specific business, as it extends to impacts that are directly linked to a company's operations, products or services by its business relationships. Importantly, it means in principle that a particular business will have to conduct HREDD along its entire transnational value chain. However, recently adopted laws, such as the French loi sur le devoir de vigilance and the German Lieferkettengesetz have limited the legal obligation to conduct HREDD to certain tiers or relations in the supply chain.[24] HREDD is defined by both the UNGPs and the OECD Guidance as an ongoing process which is meant to vary in complexity with the size of the business enterprise, the risk of severe human rights impacts, and the nature and context of the operations of a specific business.

The fact that HREDD is risk-based implies that the measures that are to be taken by an enterprise will need to be 'commensurate to the severity and likelihood of the adverse impact' and 'adapted to the nature of the adverse impact'.[25] Furthermore, HREDD emphasizes the need for a business to 'prioritise the order in which it takes action based on the severity and likelihood of the adverse impact'.[26] In short, the most significant impacts must be identified and dealt with before the less significant ones. This process of identification and mitigation of risks is to be informed by meaningful stakeholder engagement and involves ongoing and accessible external communication about the due diligence processes, findings and plans.

Fig. 3 | The OECD's six steps of due diligence



22 UNHRC, 'Guiding Principles on Business and Human Rights' (2011) UN Doc HR/PUB/11/04; OHCHR, 'The Corporate Responsibility to Respect Human Rights: An Interpretive Guide' (2012) UN Doc HR/PUB/12/02.

23 OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018); OECD, Guidelines for Multinational Enterprises (OECD Publishing 2011).

24 The French law restrict the scope of its "plan de vigilance" to the stable business relationships of a particular company, while the German law requires a systematic risk analysis only with regard to direct suppliers and in exceptional circumstances with regard to indirect suppliers.

25 OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018), at 17.

26 Ibid, at 17.

In practice, HREDD constitutes a 'bundle of interrelated processes'[27] aimed primarily at:

- Identifying and assessing the risk of adverse impacts
- Preventing and mitigating them through specific actions
- Tracking the implementation and effectiveness of the actions taken
- Communicating on how adverse impacts are being addressed

3.2. Using big data technologies in HREDD

As HREDD is progressively being turned into a legal obligation for a number of companies in Europe and beyond, questions related to the proper implementation by businesses of the process will become increasingly relevant. Both (national or European) courts and (national or European) administrative bodies will soon have to determine whether a particular HREDD process implemented by a company has complied with the requirements introduced in national law and soon European legislation. In light of the complexity and transnationality of supply chains, it is likely that businesses will have recourse to technological support to conduct the HREDD process. In this regard, this study aims to discuss the extent to which big data technologies, as defined in section 2, could play a role in discharging this duty to implement HREDD. We will address the issue first in the abstract, by highlighting, based on our literature review, how big data could hypothetically contribute to the HREDD process. This general discussion will be followed by the presentation of use cases in which big data technologies are already being used in order to contribute to different steps of the HREDD process.

3.2.1 General potential of big data technologies for HREDD

The emergence of big data dates back to the early 2000s, but it is only in the past decade that these technologies started to be harnessed to pursue development goals or humanitarian objectives. Similarly, it is only recently that corporations commenced to utilise big data technologies in order to strengthen the sustainability of their supply chains. These emergent usages of big data for the public good can inform the relevance of big data in the context of the HREDD process.

The emergence of big data for development and human rights

While the use of big data technologies by businesses is often a way to increase efficiencies and ultimately profits, big data is also increasingly mobilised to pursue the public interest or, more simply put, to do good. This ambition has been at the heart of the creation by the UN of the Global Pulse, an initiative aimed at leveraging innovations in digital data, rapid data collection and analysis to help decision-makers.[28] More generally, the contribution of big data technologies to development and the fulfilment of the sustainable development goals has also been extensively discussed.[29] From this perspective, big data is to be harnessed primarily by public actors, states and international organisations to guide transformative policy changes aimed at improving the lives of citizens.

At the same time, big data has also been progressively harnessed by a number of civil society organisations in the context of humanitarian activities, leading to the emergence of what has been called 'Digital Humanitarians'.[30] In particular, the tragic 2010 earthquake in Haiti proved a foundational moment in which a set of big data tools were for the first time deployed to guide some of the relief action.[31] As outlined by Aronson, 'the 'digital exhaust' from mobile phone and Internet usage can increasingly be mined for information about social, economic, and political conditions that often precipitate human rights violations if not remedied quickly'.[32] The use of big data has also been trialled to support investigations into international crimes in conflict zones

27 Ibid, at 16

28 See the website of Global Pulse at https://www.unglobalpulse.org.

29 UN Global Pulse, 'Big data for Development: Challenges and Opportunities' (May 2012), <https://www.unglobalpulse.org/document/big-data-for-development-opportunities-and-challenges-white-paper/>. See also, World Economic Forum, Big Data, Big Impact: New Possibilities for International Development, 2012, <https://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf>; Jennifer L. Cohen and Homi Kharas 'Using big data and artificial intelligence to accelerate global development' (Brookings, 15 November 2018) <https://www.brookings.edu/research/using-big-data-and-artificial-intelligence-to-accelerate-global-development/> accessed 24 February 2023. For a sobering analysis of the potential and challenges of big data in this context, see Steve MacFeely, 'The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators' (2019) 10(1) Global Policy 121.

30 Patrick Meier, Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response (Routledge 2015).

31 Recalling the 'Early Battles with Big Crisis Data', see Patrick Meier, 'Big (Crisis) Data Humanitarian Fact-Finding with Advanced Computing' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016), at 480-481.

32 Jay D. Aronson, 'Mobile Phones, Social Media and Big Data in Human Rights Fact-Finding: Possibilities, Challenges, and Limitations' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016)

which are impossible to access physically.[33] In short, big data technologies are perceived as contributing to a new technologically informed approach to enforcing human rights law and engaging into human rights advocacy.[34]

While the assessment of the literature is not without strong caveats about the impacts and effectiveness of these initiatives, be it on the development side[35] or the human rights side,[36] these examples aimed at leveraging big data in the public interest can provide interesting models and insights in considering the relevance of big data in the context of the HREDD process.

The turn to big data in the management of sustainable supply chains

While leveraging big data technologies for development and human rights is a relatively new phenomena, businesses have traditionally been using big data analytics for a variety of purposes, including to better anticipate and influence the behaviour of their customers and to improve the efficiency of their decision-making.[37] Moreover, in recent years, various authors have also envisaged the use of big data in order to improve the sustainability of transnational supply chains.[38] In particular, it has been argued that big data analytics can help predict several social problems (such as workforce safety, fuel consumptions monitoring, workforce health, security, physical condition of vehicles, unethical behaviour, theft, speeding and traffic violations) and enable companies to mitigate social risks throughout their supply chains.[39] Similarly, big data analytics has been found to help with 'remote monitoring of labour issues, including compensation, like the use of forced or child labour and exploitation of communities where the supply chain facilities are located'.[40] In the agricultural sector, researchers have concluded that the use of big data in the coffee supply chain 'can control and reduce waste from the production process and reduce the risks arising from the processing'.[41] All these examples rely more or less directly on the increased surveillance capacity of lead firms in supply chains enabled by big data, and are going in the direction of the emergence of the 'digital sustainability panopticon'[42] envisaged by Seele. The intimate link between these developments and the underlying purpose of the HREDD process should be evident, but none of these contributions considered directly how big data could support and be integrated into the HREDD process.

Towards harnessing big data for HREDD?

Until now big data technologies have not been linked in the literature with the HREDD process. During our research, we did not find any academic article engaging with the role of big data in this context. Yet, in theory at least, the collection of massive amounts of diverse data and its analysis through descriptive, predictive or prescriptive algorithms could play a role in the implementation of HREDD processes. Data collected on the web (such as social media posts, videos or news items), in specialised repositories (such as entries in public registries or customs data), from remote sensing technologies (via satellite imagery, cameras or IoT), or from businesses could be harnessed in the implementation of the due diligence process to improve its effectiveness in preventing the materialization of human rights or environmental risks.[43] More concretely, big data has the potential to help respond to one of the fundamental challenges faced by corporations in implementing the HREDD process: the size and complexity of their supply chains. Many businesses have a limited insight into their supply chains beyond the first tier and will, therefore, struggle to identify and assess adverse impacts linked to their products and services located further upstream. Accordingly, the first step in an effective HREDD process will always be a comprehensive mapping of a corporation's supply chain, its functional territory

33 Patrick Meier, 'Big (Crisis) Data Humanitarian Fact-Finding with Advanced Computing' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016), at 481-487.

34 Molly K. Land and Jay D. Aronson (eds), New Technologies for Human Rights Law and Practice, (CUP 2018) and Galit A. Sarfaty, 'Can Big Data Revolutionize International Human Rights Law?' (2017) 39(1) University of Pennsylvania Journal of International Law 73.

35 See for a strong critique, Miren Gutierrez and John Bryant, 'The Fading Gloss of Data Science: Towards an Agenda that Faces the Challenges of Big Data for Development and Humanitarian Action' (2022) 65 Development 80.

36 See Galit A. Sarfaty, 'Can Big Data Revolutionize International Human Rights Law?' (2017) 39(1) University of Pennsylvania Journal of International Law 73; Roisin Read, Bertrand Taithe & Roger Mac Ginty 'Data hubris? Humanitarian information systems and the mirage of technology' (2016) 37(8) Third World Quarterly 1314; and Ryan Burns, 'Rethinking Big Data in Digital Humanitarianism: Practices, Epistemologies, and Social Relations' (2015) 80(4) GeoJournal 477.

37 See Andrew McAfee and Erik Brynjolfsson, 'Big Data: The Management Revolution' (2012) 90(10) Harvard Business Review 60.

38 Ricardo Chalmeta and Nestor J. Santos de León 'Sustainable Supply Chain in the Era of Industry 4.0 and Big Data: A Systematic Analysis of Literature and Research' (2016) 12(10) Sustainability 4108; Rameshwar Dubey et al, 'Can big data and predictive analytics improve social and environmental sustainability?' (2019) 144 Technological Forecasting and Social Change 534.

39 Venkatesh Mani et al 'Mitigating Supply Chain Risk via Sustainability Using Big Data Analytics: Evidence from the Manufacturing Supply Chain' (2017) 9(4) Sustainability 608.

40 Joash Mageto, 'Big Data Analytics in Sustainable Supply Chain Management: A Focus on Manufacturing Supply Chains' (2021) 13(13) Sustainability 7101.

41 Yotsaphat Kittichotsatsawat, Varattaya Jangkrajarng and Korrakot Yaibuathet Tippayawong, 'Enhancing Coffee Supply Chain towards Sustainable Growth with Big Data and Modern Agricultural Technologies' (2021) 13(8) Sustainability 4593.

42 Peter Seele, 'Envisioning the digital sustainability panopticon: a thought experiment of how big data may help advancing sustainability in the digital age' (2016) 11(5) Sustainability Science 845.

43 For a critical study of such use, see Matthew Archer, 'Imagining Impact in Global Supply Chains: Data-Driven Sustainability and the Production of Surveillable Space' (2021) 19(3) Surveillance & Society 282.

of responsibility. As we will see in section 3.2.2, specific solutions based on big data analysis are starting to emerge in order to support companies in their mapping of their supply chains.

Big data technologies are also widely used in a variety of sectors to predict the existence of risks or adverse impacts (such as by the police in order to predict crimes).[44] The fact that risk assessments are so central to the HREDD process and risk prevention is its main purpose would in principle speak for integrating big data technologies already employed in other contexts to predict risks.[45] In other words, large quantities of data (such as satellite imagery, social media posts or news articles) could be collected, mined and analysed with the help of big data technologies in order to produce risks scores on a variety of metrics relevant to the HREDD process. In fact, as discussed in greater details in section 3.2.3 below, big data technologies are currently being leveraged by a number of service providers in order to identify specific risks or harms, such as deforestation or forced labour, in supply chains.

Finally, big data technologies are likely to possess the potential to support other steps of the due diligence process, even though we have not found evidence that there are being used to do so at this point. For example, the tracking and the assessment of the effectiveness of the actions taken to prevent or mitigate the risks identified through the HREDD process could be supported by similar big data technologies as the risk assessment. Furthermore, in the longer term, it might be possible for AI tools trained on large datasets of past decisions by corporations taken in the context of their HREDD process to offer prescriptive insights in terms of the best course of action concerning a particular risk or with regard to the most appropriate remedy to be offered to affected stakeholders.

In any event, as we will show in the coming sections, big data technologies are already being employed to contribute to the HREDD processes of corporations and it is difficult to imagine that this contribution would diminish or disappear in the years to come.

44 For a comprehensive summary of the existing use of big data in predictive policing, see See Peter Seele, 'Predictive Sustainability Control: A review assessing the potential to transfer big data driven 'predictive policing' to corporate sustainability management' (2017) 153 Journal of Cleaner Production 673.

45 Advocating the transferability of big data technologies from policing to 'sustainability control', see ibid.

3.2.2. Use case 1: Big data technologies and the mapping of the functional territory of HREDD

The centrality of supply chain mapping for HREDD

Supply chain mapping is necessary to complete the first half of the 'scoping exercise' in the abovementioned second step of HREDD. This 'scoping exercise' is where businesses identify the adverse impacts linked to their operations, products and services.[46] In particular, it requires producing a high-level picture of all areas of a specific business, across its operations, relationships and supply chain.[47] It is only when a business has this information that it can begin identifying the adverse impacts associated or linked to its operation, products or services. Scoping is necessary because, as previously mentioned, HREDD covers not only the business' own human rights and environmental impacts but also those to which it is directly linked.[48] The various pieces of legislation and guidance on HREDD have different definitions of the nature and intensity of this link.[49] This scoping can be done through supply chain mapping but there is also a trend of businesses providing commodity chains mapping services whereby they map the flow of commodities and can be useful for businesses to know the geography of their supply chain, though is not business-specific like a supply chain.[50]

The fact that HREDD extends beyond the impacts caused or contributed to by a specific business to cover also the adverse impacts to which it is linked through its supply chain implies that the mapping of these supply chains is a necessary first step to the assessment. Without a detailed knowledge of the parties involved in their supply chain, businesses would lack the ability to identify and assess the extent of the risks linked or associated with their activities.

The challenge with supply chain mapping

Mapping supply chains is a challenging task for businesses of all sizes.[51] There are many factors that combine to make this a hard task, for example, a study by Gaur, Osadchity and Udenio concluded that two largest challenges were a lack of upstream visibility as firms rarely know their suppliers beyond tier 1 and the interlinkage of supply chains in a vast global network.[52] The challenges faced can be split into three fundamental characteristics of modern supply chains: they are vast, complex, and changeable.[53]

First, the vastness of modern supply chain refers to the sheer number of firms that businesses have relationships with, even those that could be considered high in the supply chain. For example, a raw material extraction firm can report having over 14,000 suppliers in their chain despite being traditionally high in the chain.[54] Other firms will, of course, have even more suppliers. Furthermore, supply chains go both upstream as well downstream; the European Commission proposal for a directive on Corporate Sustainability Due Diligence explicitly stated its intention to apply in both directions of the supply chain.[55] The OECD Due Diligence Guidance also references the use of HREDD on mid-stream and upstream suppliers.[56] The sheer size of modern transnational supply chains makes mapping them a practical challenge for every business initiating a sustainability due diligence process.

Furthermore, the structure of modern supply chains composed of a wide diversity of intermediary business relationships makes it very difficult to map the entire network of entities involved in a particular supply chain. More often than not, businesses will be ignorant of many of the companies involved in their supply chains. For example, a business will ordinarily never know, through looking at their official suppliers list, that their supplier in Shanghai is potentially sourcing its products from a company active in Xinjiang.

46 OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018), at 25-28.

47 Ibid., at 25.

48 Ibid

49 The OECD Due Diligence Guidance calls for the mapping of an enterprise's operations, suppliers, and other business relationships, including associated supply chains. Highlighting that scoping is intended to be broad and to serve as an initial exercise to also enable prioritisation. It explicitly refers to the relevance of supply chain mapping in this initial scoping exercise, stating that it "will be important for enterprises to map their general operations and the structure of the supply chains". The UNGPs' definition of HRDD likewise goes beyond a business's adverse human rights impacts from its own activities and covers also those which may be directly linked to its operations, products or services by its business relationships. The German Lieferkettengesetz calls for HREDD to cover the actions of the enterprise in its own business area, the actions of direct suppliers and, to a more limited extent, the actions of indirect suppliers.

50 Some businesses that offer commodity mapping are: The Sustainability Consortium and S&P Global

51 Vishal Gaur, Nikolay Osadchiy and Maximiliano Udenio, 'Research: Why It's So Hard to Map Global Supply Chains' (Harvard Business Review, 31 October 2022) <https://hbr.org/2022/10/research-why-its-so-hard-to-map-global-supply-chains> accessed 16 December 2022.

52 Ibid.

53 Lisa Smit et al, 'Human rights due diligence in global supply chains: evidence of corporate practice to inform a legal standard' (2021) 25(6) The International Journal of Human Rights 945, 949-951.

54 Ibid.

55 Proposal for a Directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive (EU) 2019/1937 [2022]

56 OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018), at 47.

This complexity is exacerbated by the unwillingness of first-tier suppliers to disclose information about their own supply chain, a problem which is prominent throughout industries but particularly with suppliers that act as consolidators of raw material from smaller firms.[57] The complexity of supply chains might also be enhanced by the inclusion of informal relationships as recognised in the OECD Due Diligence Guidance.[58] Overall, the complexity of the production networks behind the supply of services and products in our globalized economy can make it a very arduous task to reconstitute the path followed by a particular product (or service) until it reaches its end user and to identify all the entities (as well as people and environments) affected by this process.

Finally, contemporary supply chains are extremely dynamic as businesses are constantly onboarding new suppliers and innovating. Consequently, these complex and vast networks of production are also changing at a fast-pace. Furthermore, modern supply chains are volatile in nature due to their susceptibility to political instability, legal developments and technological changes – making it difficult for businesses to maintain a complete picture of their supply chain.[59] The ephemeral nature of contemporary supply chains is further driven by the ease with which companies appear and disappear by changing their form or place of incorporation. The UNGPs took into account this issue by highlighting that the HREDD process is dynamic and needs to adapt to changes in the way in which businesses operate and structure their activities.[60] In a similar vein, the OECD Due Diligence Guidance for Responsible Supply Chains in the Garment and Footwear Sector emphasised the short-term nature of the buyer-supplier relationships in the sector.[61] This volatile nature means supply chain maps can quickly become outdated and difficult to keep a track of.

The vastness, complexity, and changeability of modern supply chains therefore make them difficult to map. Were a business to simply contact their suppliers to ask for their own list of suppliers it is likely that, even in the unlikely event that they would receive a complete list of their suppliers' business relationships, this list would be outdated by the time they receive it. It is against this backdrop that big data technologies could provide corporations engaging in an HREDD process some needed support in the mapping of their supply chains. In particular, we have identified a number of service providers relying on AI to analyse vast amounts of data published by customs authority or national registers of corporations, which claim to be able to produce an accurate map of the supply chain of a particular company.[62]

Use case: Leveraging big data technologies to map supply and value chains

Company A is a technology-based start-up which provides a software as a service (SaaS) in the form of an interactive map of supply chains. It utilises cloud computing and observations of transactions to build a map of the global supply chain over which it flags forced-labour risks.[63] In an international transaction these observations can come from customs authorities that register the exportation and importation of goods. This publicly available information can provide details on who is shipping to who, and some information around what is being shipped and when, including details like HS codes.[64] This data is purchased from 3rd party providers and stems from the custom authorities of specific states, including India and the US. Such data is not available, however, for domestic shipments or in jurisdictions where such information is not available for purchase – like the EU and China. Company A also receive data from partners that operate in the shipping, warehousing and logistics industry. Another purchased 3rd party data source is ownership data which is used to connect clusters of companies that are linked but are legally separate entities, for example there are many separate Adidas corporations which otherwise would be seen as separate by AI. Other 3rd party data sources include geographical, financial and country-specific data.

To address the data gaps in 3rd party data sources, Company A employs a federated learning model to develop their own 1st part data. Company A's customers run their own instance of their map in a secure data enclave, where their data is enriched and fused with the chain, and where machine models are trained to send analytics and learnings back to the core map model to improve its accuracy. The federated aspect refers to the fact that these insights are incorporated into the map without clients having direct access to the datasets shared. This is being done by only taking the 'learnings' from this data, through aggregate

57 Lisa Smit et al, 'Human rights due diligence in global supply chains: evidence of corporate practice to inform a legal standard' (2021) 25(6) The International Journal of Human Rights 945

58 OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018), at 36.

59 Lisa Smit et al, 'Human rights due diligence in global supply chains: evidence of corporate practice to inform a legal standard' (2021) 25(6) The International Journal of Human Rights 945.

60 UNHRC, 'Guiding Principles on Business and Human Rights' (2011) UN Doc HR/PUB/11/04, Principle 18.

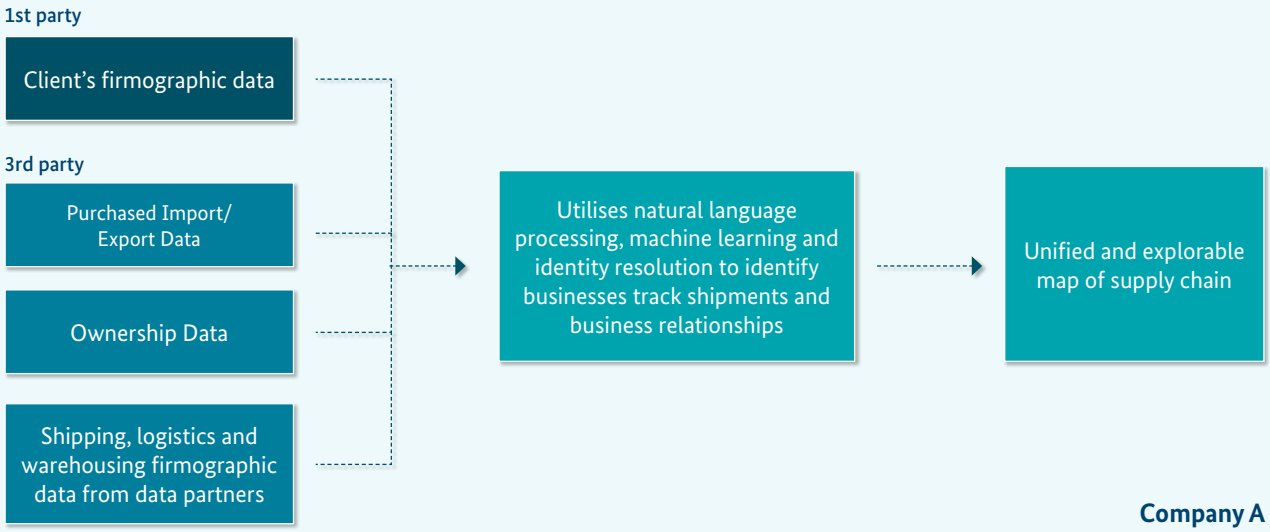
61 OECD Due Diligence Guidance for Responsible Supply Chains in the Garment and Footwear Sector, at 17

62 See in particular Altana AI (https://www.altana.ai) and Sayari (https://sayari.com).

63 For the purpose of this case study only Company A's supply chain mapping capabilities are discussed.

64 HS Codes are commodity codes used by customs authorities worldwide to categorise products.

Fig. 4 | Use case: Leveraging big data technologies to map supply and value chains



information commonly known as firmographic data. This model is commonly deployed elsewhere, such as in smartphone maps or texting applications. Taking the latter as an example, the spell check application on a smartphone learns how its user types, but entire text messages are not sent to the smartphone maker for analysis. Instead, the typing patterns, common mistakes, and other pieces of aggregate information are analysed — allowing for both a tailored spell-check for every user and a better understanding of language patterns for all user applications. This private, closed-source data can allegedly fill in some of the gaps left by the open-source data and in a way that enriches the map without businesses sharing proprietary information with the public or their competitors. Identity resolution and machine learning is then used to combine all these data sources, creating a unified and explorable view of the supply chain that is always growing. Company A claim their map can navigate the issue of the enormity of modern supply chains. The issue is simply, as will be expanded in more detail later, access to enough data to be able to do this. With regard to complexity, Company A claims to be able to find business relationships that are not obvious or disclosed. By seeing which businesses are sending goods to each other, for example, the company claims to be able to identify the business networks of firms to see whether a supplier has a sister company based elsewhere from which it is receiving goods that would otherwise be invisible to businesses using traditional methods. Furthermore, Company A claim they can match the shipments and products involved in the various stages of the chain to the end-product, for example, matching cotton and thread which is shipped between businesses to an apparel product made using that cotton. They do this through AI tools using natural-language processing to match the product descriptions or HS codes of goods shipped to a business and then create a likelihood score of those goods having a relationship to the final outcome. The alleged ability to map value chains would mean, in the HREDD context, that businesses can carry out human rights and environmental risk assessments on those businesses that have contributed to the products they sell – not businesses that are unrelated but happen to be in a business relationship with the supplier.

Model example of a product specific value chain

However, Company A's data is limited to international exportation and importation from and to specific countries or data received from clients and data partners – this will leave some business relationships unaccounted for. Shipments within one country can only be picked up where the shipment is done via one of their data partners or through open source information. This is usually the first step in a supply chain – where raw materials are transported to an exporter ready for exportation, meaning this step in the supply chain is likely often missed. The harvesting of raw materials that are used later in the supply chain is an area where human rights and environmental risks are prevalent. For example, reports have suggested that 570,000 people in the Xinjiang region of China are being forced to pick cotton through coercive labour training and transfer schemes.[65] Where such materials are first transported within China, Company A will only be aware if it happens to be done via a data partner or an open source of information.

Furthermore, the changeability of modern supply chains may present a challenge. Company A accepts that there is a time delay between a new entity appearing in a supply chain and its ability to identify it. This is also true when businesses create avatars after

65 Adrian Zenz, 'Coercive Labor in Xinjiang: Labor Transfer and the Mobilization of Ethnic Minorities to Pick Cotton', Intelligence Brief, Center for Global Policy, December 2020 <https://media.business-humanrights.org/media/documents/20201214-PB-China-Zenz-11.pdf> Accessed 10 January 2023.

they have been identified as contributing to human rights or environmental risks. Businesses will often cease their activities, and a brand-new entity will appear in its place engaging with all the old business-relationships. However, in such a case, Company A will need several weeks for enough data to accumulate and accrete to confirm that the new business is just a guise for the old one.
In conclusion, this example shows how big data technologies have the potential to help with supply chain mapping, as they have the capacity to overcome some of the problems related to the vast, complex, and changeable nature of supply chains. Company A is not the only provider of this type of services and it seems big data combined with AI analysis could increase supply chain transparency and visibility and, therefore, enable lead companies to conduct HREDD on a more accurate and comprehensive set of entities and locale directly linked to their products or services.

3.2.3 Use case 2: Big data technologies and risk identification and analysis in HREDD

The centrality of risk identification and assessment in HREDD

The HREDD process aims at preventing the manifestation of certain adverse impacts affecting human rights or the environment. In order to do so, it requires primarily that companies determine the actual or potential adverse impacts linked to their activities or supply chains. The identification of these actual or potential adverse impacts requires that each company engages in a complex discovery process in order to identify the specific impacts connected with its economic activity. This impact assessment, as provided by the OHCHR Interpretive Guide to the Responsibility to Respect, will need to draw on 'various sources'.[66] The OHCHR Interpretive Guide identifies a variety of sources relevant to the assessment, such as:

- A grievance mechanism
- News or expert reports on particular operating contexts or industry developments
- Campaigns by non- governmental organizations (NGOs) or other third parties

Ultimately, the Guide foresees that 'processes for assessing human rights impact should be systematic so that the various elements add up to a coherent overview of actual and potential human rights impact associated with an enterprise's activities and relationships and can accurately inform the subsequent steps in the due diligence process'.[67]

The OECD Due Diligence Guidance provides a more detailed description of what is expected from corporations in this regard. First, it foresees a broad scoping exercise, aimed at identifying 'all areas of the business, across its operations and relationships, including in its supply chains, where RBC [Responsible Business Conduct] risks are most likely to be present and most significant'.[68]The exercise calls for the analysing of information about sectoral, geographic, product and enterprise risk factors, which can be gathered from governments, international organisations, civil society organisations, workers' representatives and trade unions, national human rights institutions (NHRIs), media or other experts. Information is also to be obtained through consultation with relevant stakeholders, as well as through early warning systems and grievance mechanisms. Second, the business must conduct in-depth assessments of operations, suppliers and other business relationships focusing on the significant areas of risk identified in the scoping exercise. It must assess the nature and extent of actual and potential impacts linked to prioritised operations, suppliers or other business relationships by using where available:

- Information from environmental impact assessments (EIA)
- Environmental and social impacts assessments (ESIA)
- Human rights impact assessments (HRIA)
- Legal reviews
- Compliance management systems regarding corruption
- Financial audits (for disclosure)
- Occupational, health and safety inspections
- Any other relevant assessments of business relationships carried out by the enterprise or by industry and multi-stakeholder initiatives, including environmental, social and labour audits, corruption assessments and Know Your Customer (KYC) processes

The impacts identified must be reassessed at regular intervals, in particular in relation to changes in the activity or operating environment of the business. Furthermore, when human rights impacts are concerned, businesses must consult and engage impacted and potentially impacted rightsholders to gather information on adverse impacts and risks, taking into account potential barriers to effective stakeholder engagement.
In short, the identification and assessment of potential adverse impacts is at the heart of the HREDD process and a central responsibility of corporations implementing it.

66 OHCHR, 'The Corporate Responsibility to Respect Human Rights: An Interpretive Guide' (2012) UN Doc HR/PUB/12/02, at 40.

67 Ibid., at 41.

68 OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018), at 25.

The challenges posed by risk identification and assessment in HREDD

We have discussed in the previous section how big data technologies can help businesses to map their supply chains by tracking links between companies and retracing the journey from particular raw material to end product. As we will see, there is also potential for big data technologies to support the risk identification and evaluation phase of the due diligence process. Indeed, once the mapping of a specific supply chain has been completed, it remains to be determined what human rights and environmental risks are prevalent along that supply chain. To conduct such an assessment, a business will need to analyse vast quantities of information on the different companies and sites involved in its supply chain. In this regard, big data technologies have the advantage, compared to human analysts, to be able to process vast quantities of data at a fast pace. Indeed, as most businesses are dealing with massive, complex and dynamic transnational supply chains, it is unlikely that (relatively small) teams of human analysts specialized in human rights or environmental issues will be able to qualitatively identify and assess a significant share of the adverse impacts linked to them. In fact, even when companies are able to engage auditors to investigate for example compliance with labour rights at specific factories, there are well-known limits to their capacity to identify the failings of a particular employer or building.[69] In this context, big data technologies could offer a powerful support for companies to assess, on a recurrent basis, the existence of risks of adverse impacts linked to their supply chains without having to organize in-person visits of each and every factory, mine or farm and without being dependent on auditors, who might be at best ineffective and at worse captured. In other words, the ability of big data technologies to gather, process and analyse vast amounts of data at a relatively quick pace could ensure a swifter and more comprehensive identification and analysis of the risks of adverse impacts linked to a particular supply chain.[70]

More concretely, the data produced by users of social media or the information stemming from news organizations around the world could be gathered and analysed by powerful algorithms (such as AI systems specialized in language analysis), which could then detect specific patterns in complaints of social media users directed against a particular business or industrial site or identify negative assessments in news items potentially unavailable to human analysts due to the sheer size of the data collected and linguistic barriers.[71]

Another potential use of big data in this context relates to the use of environmental data stemming from remote sensors or satellite imagery to assess risks or ongoing instances of, for example, deforestation. Indeed, powerful algorithms trained to assess satellite imagery can sift through massive quantities of data to pinpoint locations in which deforestation is highly likely to be occurring.[72] For example, it has been argued that 'advances in remote sensing and cloud computing (the use of networks of remote servers hosted on the Internet to store, manage, and process data) are making it possible to monitor changes in forest cover and condition, as well as the extent of fires and the impacts of natural disasters more cost-effectively and more frequently than forest patrols and surveys are able to'.[73] The same could be true of the health and safety of workers, which could be monitored at a distance by gathering data through wearable bracelet sensors.[74] In short, the relative ease of collection of massive amounts of diverse data from a wide variety of sources linked with the capacity to process that data swiftly and effectively could in theory help alleviate the physical limits faced by human analysts when they have to conduct a wide-ranging risk analysis (in terms of the variety of adverse impacts on human rights or the environment that need to be considered) on an even wider terrain (the entire supply chain of a specific business).

Use case: Big data analysis of publicly available information on the Internet in order to identify and assess human rights and environmental risks

Our first use case on risk identification and analysis concerns Company B, which claims to be capable to employ AI and automated language processing in order to identify a wide range of risks in every tier of the supply chain of its customers. In order to do so, specialized data collectors scan publicly available information on the Internet from sources in more than 50 languages, including from local news media, NGOs websites, social media, and other publicly available databases. The information is collected automatically in blocks every 5 or 6 hours based on keywords related to specific event types corresponding to specific risks, including in particular human rights and environmental risks. The data collected is automatically assessed to ensure data quality and avoid duplicates.

69 Genevieve LeBaron, Jane Lister & Peter Dauvergne, 'Governing Global Supply Chain Sustainability through the Ethical Audit Regime' (2017) 14(6) Globalizations 958.

70 For similar conclusion on the contribution of big data analysis to human rights fact-finding, see Jay D. Aronson, 'Mobile Phones, Social Media and Big Data in Human Rights Fact-Finding: Possibilities, Challenges, and Limitations' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016), at 448.

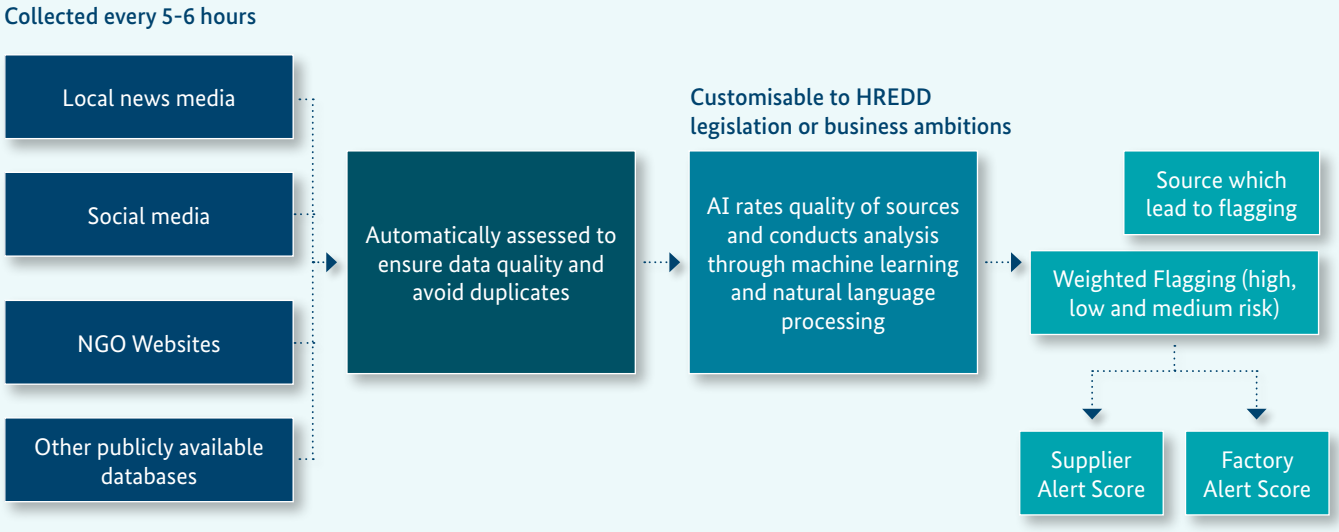
71 Already a number of service providers are claiming to be able to conduct this type of automated risk analysis, see our list of big data technology solutions in annex.

72 R.Taylor, C. Davis and J. Brandt, 'The rise of big data and supporting technologies in keeping watch on the world's forests' (2020) 22(1) International Forestry Review 129.

73 Ibid, at 130.

74 For such an attempt, see Kenzen (https://kenzen.com).

Fig. 5 | Use case: Analysing publicly available information to identify and assess human rights and enviromental risks



It is then analysed by an AI employing machine learning and natural language processing. Company B claims that the AI is able to accurately identify data pointing at specific risks linked to their customers' suppliers. In the framework of this analysis, the AI rates the quality of the underlying sources and tries to detect whether other sources point at the same risk linked to the same company or industrial site. If the AI comes to the conclusion that the data collected supports the existence of a risk it will issue an alert which will be visualised for the customers of Company B through an online interface. The alerts will either be flagged as low, medium, high risk and red flag alerts depending on the customer's customization of the system and the specific characteristic of the event identified (taking, for example into account the number of persons potentially affected) or of the company concerned (if it is a company that has already been flagged for this risk). The customers have then the opportunity to review the underlying data which led the AI to issue the alert and can decide how to proceed next to prevent the risk to materialise or to mitigate the harm suffered. Furthermore, the customers of Company B have also access to an overall alert score reflecting the AI risk scoring of suppliers (and their factories) on a scale of 0-100 (with 0 being the lowest score) based on the frequency, types and severity of alerts produced by the AI in a time frame of 24 months.

A short case study published by Company B on extractive practices in Ethiopia and their effects on local communities illustrates the mode of operation of this technology. As presented by the company, its technology was capable of detecting social unrests caused by the environmental and health impacts of a particular mining company in social media posts and news items published in Amharic (in this case Twitter posts). This led Company B to label the mining company (and in particular its mining site in Ethiopia) as high risk in its platform even before the government decided to remove its license. The example illustrates how, by gathering publicly available sources on a specific production site or company and by being able to process it quickly, Company B might be able to rapidly flag certain risks and harms and attract the attention of its customers to a weak link in their supply chains.

In the course of our research, we have encountered a number of other companies claiming to engage in similar approaches than Company B in order to flag risks emerging in the supply chain of their clients. Most of these companies are either buying large data bundles of social media posts or new items from specialized providers and then analysing them with their in-house technology or, like Company B, dredging the internet to collect the data themselves. This type of approach to the use of big data analytics in the context of HREDD is premised primarily on the capacity of machine learning algorithms to analyse vast quantities of text and to identify correlations between the textual content and specific risks. However, we are not in a position to assess the effectiveness of such technologies in achieving their aim, and we have not encountered any publications having done so at this point. Nevertheless, we can conclude that they are rapidly occupying a market segment and being aggressively marketed as a suitable technical solution to comply with due diligence obligations stemming from emerging mandatory HREDD legislation.

Use case: Big data analysis of satellite imagery to detect deforestation

The second company profiled in this section presents itself as a 'geospatial sustainability company' which provides insights on deforestation, supply chains, land degradation, and downstream scope 3 emissions. In terms of technology, its approach is to rely on data collected from remote sensing technology in the guise of satellite imagery.

In particular, for the purpose of this case study, we are focusing on the company's use of satellite imagery to identify risks of deforestation (but the company is also monitoring tree age, carbon stock, burned area, and crop performance indicators). In terms of data, the company relies on open data from satellites (Santine-1, Landsat (5,7,8) and Sentinel-2) which are capable of looking through clouds and analyses the radar signals that they collect. The changes of the levels of the signal over time (the higher the forest density, the higher the signal is) provide a good indication of changes in the forest environment and, therefore, potentially of ongoing deforestation.

In order to process these photographs Company C relies on a change detection algorithm using 'Iterative Bayesian Updating', which calculates the probability that an area is deforested and 'stacks' these probabilities in an iterative way to have a higher degree of certainty about the deforestation event. Ultimately, the tool is delivering to its customers a probability assessment that deforestation has been taking place.

The total area monitored by Company C reaches approximately three million square kilometres meaning that three billion pixels are being processed on a near-real-time basis by the company's algorithms. There is no way a human analyst could process such a quantity of data on a daily basis. In fact, the running of the algorithm requires massive amounts of processing power which is secured through a specific cloud system. Ultimately, the detection of changes in forest cover by the algorithm does not automatically trigger an alert being sent to the customers. The change in forest coverage is first compared to the historical record based on Landsat satellite archive dating back to 1984 in order to make sure that the area in question has never been deforested before. In addition, each customer of Company C can customize a list of additional conditions determining when alerts are to be sent to them.

Unlike Company B, Company C is thus relying on data capturing the state of the environment through sensory captors (in this case satellites, but other types of sources of data, such as video cameras could also be used), which is then processed by algorithms in order to determine whether the data is pointing at the existence of deforestation. This approach could potentially, under the condition that relevant data can be collected from sensory captors, be applicable to capture other types of environmental issues, such as degradation of biodiversity or pollution linked to particular farming, industrial or mining processes.

3.3. Using big data technologies in supervising compliance with HREDD obligations

3.3.1 Monitoring HREDD compliance[75]

Legislation may require national authorities to monitor companies' compliance with HREDD obligations. The German Supply Chain Act, for example, is officially monitored and enforced by the Federal Office for Economic Affairs and Export Control.[76] As such, this authority must review the annual reports that businesses are required to submit on the fulfilment of their due diligence obligations. It has also far-reaching supervisory powers. For instance, it can inspect and examine documents, and prescribe specific measures to remedy problems. It may also require enterprises to take concrete steps to meet their obligations, which it may enforce by imposing financial penalties and administrative fines.

Furthermore, private third-parties, such as NGOs, investors, and business partners, may seek to monitor and evaluate companies' HREDD processes and their effectiveness.[77] For example, due to a lack of public enforcement of the French Act on the Duty of Vigilance by French authorities, NGOs have played an important role in monitoring companies' compliance with their obligations, particularly by analysing the vigilance plans that companies must publish each year.[78]

75 It is important to note that monitoring in the context of HREDD has both an external and internal dimension. While external HREDD monitoring is performed by third parties (eg public authorities and private third-parties), internal HREDD monitoring is performed by the company itself and is an important part of the HREDD process for businesses. This case study focuses on the potential of big data technologies for external monitoring. However, big data technologies could also be relevant for internal monitoring by assisting businesses in tracking the effectiveness of HREDD measures they have implemented.

76 Bundesamt für Wirtschaft und Ausfuhrkontrolle.

77 Claire M. O'Brien and Olga Martin-Ortega, 'EU Human Rights Due Diligence Legislation: Monitoring, Enforcement and Access to Justice for Victims' (Briefing N°2, European Parliament 2020), at 3.

78 See Radar de devoir de vigilance, https://plan-vigilance.org/.

3.3.2. The potential role of big data technologies in monitoring HREDD compliance

In theory, big data technologies could support the monitoring of companies' compliance with mandatory HREDD obligations. Indeed, the functionalities of big data technologies sought by companies in order to implement their HREDD obligations, as discussed in Section 3.2, may also be useful to national authorities and private third-parties seeking to ensure effective compliance with due diligence legislation. Broadly speaking, adopting a monitoring approach based on big data could increase resources efficiency for monitoring authorities and private third-parties. As described in this study, big data technologies are being used for the mapping of supply and value chains or the identification of human rights and environmental risks, a task that would otherwise likely require greater financial, technical, and human resources. Given that authorities and NGOs have limited resources, the use of big data technologies can enable them to allocate these more efficiently.

Additionally, using big data may provide benefits for reviewing corporate due diligence reports. HREDD legislation frequently requires companies to submit an annual report outlining the risks identified in their supply chain as well as the measures taken to prevent those risks. Depending on the legislation's personal and material scope (ie the types of companies and their business or ownership relationships covered by the legislation), as well as its global reach, those reports may contain a variety of complex information. In practice, this means that authorities and private third-parties will have to review a large amount of complex data if they want to assess the content of the reports of all, or even some, companies subject to HREDD obligations. In this context, some AI-based processes could assist the review and analysis of those reports. For example, natural language processing, which enables computer to process human language in the form of text or voice data and to understand its full meaning,[79] could be used to select, extract, categorize, and organise relevant information from corporate due diligence reports. This could make it easier for monitoring authorities and private third-parties to identify relevant or missing information.

3.3.3 Status quo of big data technologies in HREDD compliance monitoring

Based on our scoping research, big data technologies do not appear to have been systematically used to monitor business compliance with HREDD thus far. Having said that, in recent years, a number of projects led by data scientists and NGOs have attempted to use AI-based technologies in order to better understand how those technologies can help monitor how companies comply with their non-financial reporting obligations. Several pieces of legislation, including the UK Modern Slavery Act, and the EU Non-Financial Reporting Directive[80] and its amending Directive on Corporate Sustainability Reporting,[81] require certain companies to disclose their impacts on sustainability issues, such as modern slavery and climate change. Although non-financial reporting and HREDD are distinct processes, companies must be transparent in both cases by making publicly available information on the human rights and environmental impacts of their activities. Given the similarities between non-financial reporting and HREDD, experiments using big data technologies to monitor corporate implementation of non-financial reporting or modern slavery statements can provide valuable insights into the potential use of these technologies to monitor business compliance with HREDD obligations.

Use case: Project AIMS

Project AIMS (Artificial Intelligence against Modern Slavery) was launched in May 2020 by the Future Society and Walk Free. The aim of Project AIMS is to understand whether AI can be used to improve the efficiency of assessing compliance with the UK and Australian Modern Slavery Acts[82] and policymaking by providing actionable insights to governments, businesses, and civil society organisations. The project's overarching goal is to identify and share best practises in modern slavery reporting, as well as to identify specific sectors where reporting falls short. It will make recommendations to businesses on how to improve compliance with the UK and Australian Modern Slavery Acts, as well as to governments considering similar legislation on how to maximise its impact.[83]

For those purposes, Project AIMS employs AI to analyse modern slavery statements and evaluate business compliance with the UK and Australian Modern Slavery Acts. More specifically, it employs data science and machine learning techniques, with a particular emphasis on natural language processing and computational

79 'What is natural language processing ?' (IBM) <https://www.ibm.com/topics/natural-language-processing>.

80 Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups [2014] OJ L330/1.

81 Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting [2022] OJL322/15.

82 'Project AIMS (Artificial Intelligence against Modern Slavery)' (The Future Society, 17 March 2021) <https://thefuturesociety.org/2021/03/17/project-aims-artificial-intelligence-against-modern-slavery/>.

83 WalkFree, The Future Society, WikiRate and BHRRC, 'AI Against Modern Slavery. Digital insights into modern slavery reporting – challenges and opportunities' (2021) < https://cdn.walkfree.org/content/uploads/2021/05/05173207/20210428-digital-insights-into-modern-slavery.pdf>, at 5.

linguistics, to delve deeply into each report while also assessing and improving the overall understanding of the quality of the reports produced under the UK and Australian Modern Slavery Acts.

Project AIMS builds on the work of Walk Free, WikiRate, and the Business & Human Rights Resource Centre (BHRRC) to evaluate statements produced under the UK Modern Slavery Act. It draws from the BHRRC Modern Slavery Registry to develop an AI algorithm that 'reads' and 'assesses' the statements produced by companies under supply chain transparency legislation. This algorithm will assess statements using 18 metrics designed by Walk Free, in line with the UK Home Office guidance, and will be integrated with the WikiRate platform to allow for ongoing human verification of the automated data collection and analysis. Project AIMS is divided into four phases, with the first two focusing on designing and fine-tuning the technology. While the first phase, which is now complete, focused on accessing, gathering, and structuring data from existing company statements in order to create a large and publicly available text corpus of modern slavery statements, the second phase, which is still in progress, aims to design an automated labelling function using weak supervision tools to increase the amount of available labelled data. Assessment of the modern slavery statements against specific metrics and publication of the results will take place during the third and fourth phases.[84]

Use case: DIHR's algorithm

The second project was conducted by the Danish Institute for Human Rights (DIHR). The DIHR notes that existing private initiatives to analyse how companies report on human rights and assess their human rights performance, such as the Alliance for Corporate Transparency and the Corporate Human Rights Benchmark, have done so through manual analysis of company reporting. Each of these projects produced data on the state of human rights reporting by large companies that can be used to assist a range of stakeholders in identifying trends and gaps in current company human rights practice. The DIHR noticed, however, that 'efforts to undertake large scale qualitative analysis of company reports are limited by the resource intensive nature of the review, requiring manual review of company reports which provide data in often quite different formats and without reference to common standards', making 'qualitative analysis challenging to scale up'.[85]

To address this challenge, the DIHR developed a project which aimed to use algorithm-assisted analysis of a large number of company sustainability reports against a set of sustainability and human rights indicators. More specifically, the DIHR developed an AI-based text mining algorithm that can be applied to large datasets. This text mining algorithm is the engine behind DIHR's SDG-Human Rights Data Explorer,[86] which currently classifies approximately 145,000 UN system recommendations to the 169 SDG targets. The DIHR accessed company sustainability reports via the GRI's Sustainability Disclosure Database, which contains links to around 50,000 company reports that used GRI reporting standards during 2010 and 2020.[87] The GRI had already tagged the reports in the dataset according to their size, sector, and geography, enabling the DIHR to segment the data at the analysis stage. The DIHR first converted the reports from pdf-files into analysable 'text bits', usually the length of a paragraph. The result was 3,391,615 text bits from the 9,374 sustainability reports included in the analysis. Furthermore, the DIHR developed a set of 18 human rights indicators which the reports were to be classified to. The human rights indicators focus on a range of issues, including disclosure of the policies and processes used by companies to identify and address human rights impacts, as well as how companies report on fundamental labour rights, the needs of a range of stakeholders, and human rights impacts in the supply chain. The text mining algorithm identified relevant text bits in company reports and generated data that was tested against a series of hypotheses, allowing the DIHR to identify trends among companies of a particular category. Then, the DIHR trained the algorithm on a subset of text bits by tagging those that responded to a specific indicator. A human expert was responsible for this tagging. The algorithm analyses the patterns in these text bits, which it then uses to evaluate and categorise the entire set of text bits.

Ultimately, the DIHR was able to observe a number of high-level trends in company human rights reporting. It claims that in some cases, it was able to make observations indicating that trends in company reporting may be responsive to regulatory developments. Furthermore, it was able to segment the data and explore whether there were any observable trends based on size, geography, and sector. According to the DIHR, the full range of data analysed by the algorithm could be of use to researchers and other actors interested in exploring company reporting on human rights issues and could potentially be a valuable supplement to quantitative analysis. [88]

84 WalkFree, The Future Society, WikiRate and BHRRC, 'AI Against Modern Slavery. Digital insights into modern slavery reporting – challenges and opportunities' (2021) < https://cdn.walkfree.org/content/uploads/2021/05/05173207/20210428-digital-insights-into-modern-slavery.pdf>, at 5.

85 Gabrielle Holly and Sarah Kajander Holmer, 'Sustainability reporting and human rights: What can big data analysis tell us about corporate respect for human rights?' (Danish Institute for Human Rights, 2022), at 4.

86 'Making the link between human rights and sustainable development' (Danish Institute for Human Rights) <https://sdgdata.humanrights.dk/en>. Accessed 24 February 2023.

87 Gabrielle Holly and Sarah Kajander Holmer, 'Sustainability reporting and human rights: What can big data analysis tell us about corporate respect for human rights?' (Danish Institute for Human Rights, 2022), at 6.

88 Ibid, at 19.

4. Risks and Challenges in using Big Data technologies for HREDD

This study has so far concentrated on the potential of using big data technologies to drive the HREDD process. However, we believe there are several risks and challenges that could be associated with this use. It should be noted that some of the risks and challenges identified below stem from a broader consideration of the risks associated with big data technologies or their use in the prevention of human rights violations.

4.1. Limits to the effectiveness of big data technologies in the HREDD process

4.1.1. The quality and incompleteness of the data used

First, the underlying data itself poses challenges to the use of big data technologies in the human rights context and therefore in the HREDD context. Ensuring the quality of the data collected seems to present a well-recognised challenge for big data analysis.[89] In order to gain useful insights, the data used must be reliable, as, put simply, bad or incomplete data results in poor quality insights – this is known as the garbage in garbage out principle.[90] This risk is particularly evident in data obtained from social media, as it is not produced under controlled conditions and is rarely subjected to rigorous scrutiny (ie it is not necessarily representative of a given situation).[91] Human rights related data in general has been identified as imperfect in nature due to the challenges faced in its collection[92]

Having reliable data can also be difficult (at least for public authorities or NGOs) due to another challenge: the inaccessibility of privately-held data. Despite the value of open-source, publicly available data, there is likely to be even more value in data held by corporations. However, those corporations can be reluctant to share the data they hold. As a result, accessing enough data of the correct quality can be a difficult task for certain organizations.[93] Furthermore, open-source data is not as accessible as one would presume – there are technical challenges associated with retrieving data which, while open-source and publicly available, can be in non-machine-readable formats amongst other technical challenges.[94]

Finally, there is the issue of data gaps or data deserts, which might fundamentally bias the data sample relied on by private providers or public authorities. For example, a large part of the world's population has no access to social media (or uncensored social media): how are these people and the risks they might face accounted for if no relevant data is collected from them?[95] In other words, particular environmental issues or social groups might become invisible in HREDD processes due to the limited data available (publicly or privately) about them. Only activities, communities, environments, or beings which are releasing collectible data or which are traced by sensors can be subjected to a big data analysis, the others will remain out of sight of a data-led HREDD process. In short, an HREDD process exclusively driven by big data technologies would likely be blind to specific risks, communities and/or environments about which limited data is available.

In sum, relying on trustworthy and comprehensive data is a necessary pre-condition to the use of big data technologies in the context of the HREDD process.

89 Jesse Nishinaga and Faris Natour, 'Technology Solutions for Advancing Human Rights in Global Supply Chains' Human Rights and Business Initiative University of California (2019), at 19; Roger Clarke, 'Big data, big risks' (2016) 26(1) Information System Journals 77, at 78.

90 Robert N. Boute and Maxi Udenio, 'AI in Logistics and Supply Chain Management' in Rico Merkert and Kai Hoberg (eds.) Global Logistics and Supply Chain Strategies for the 2020s: Vital skills for the Next Generation (Springer 2022), at 16.

91 Jay D. Aronson, 'Mobile Phones, Social Media and Big Data in Human Rights Fact-Finding: Possibilities, Challenges, and Limitations' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016), at 445-446.

92 See John Emerson, Margaret L. Satterthwaite and Anshul Vikram Pandey, 'The Challenging Power of Data Visualization for Human Rights Advocacy' in Molly K. Land and Jay D. Aronson (eds), New Technologies for Human Rights Law and Practice, (Cambridge University Press 2018), at 169.

93 UN Global Pulse, 'Big data for Development: Challenges and Opportunities' (May 2012).

94 WalkFree, The Future Society, WikiRate and BHRRC, 'AI Against Modern Slavery. Digital insights into modern slavery reporting – challenges and opportunities' (2021) < https://cdn.walkfree.org/content/uploads/2021/05/05173207/20210428-digital-insights-into-modern-slavery.pdf>, 8; Gabrielle Holly and Sarah Kajander Holmer, 'Sustainability reporting and human rights: What can big data analysis tell us about corporate respect for human rights?' (Danish Institute for Human Rights, 2022), at 6.

95 See Sarah Giest and Annemarie Samuels, "For good measure': data gaps in a big data world' (2020) 53 Policy Sciences 559 and Miren Gutierrez and John Bryant, 'The Fading Gloss of Data Science: Towards an Agenda that Faces the Challenges of Big Data for Development and Humanitarian Action' (2022) 65 Development 80, at 89.

4.1.1. The risk of algorithmic bias in the analysis of the data

There is often a perception amongst the public and amongst developers that technology is inherently neutral and objective.[96] However, as many authors have argued, technology reflects the values and interests of its creators and the data upon which it functions, both of which are fundamentally shaped by society.[97] In fact, it has been argued that a misplaced faith in the neutrality of technology can contribute to discriminatory outcomes.[98] The reasons for the lack of technology neutrality are two-fold: the data fed into big data algorithms and the subjective intervention of those developing these algorithms.

We have discussed the potential issues with biased or incomplete data in the previous section, here we are concerned with the hidden assumptions embedded in the algorithms tasked with mining and analysing the data.[99] Algorithms are as good as their developers program and train them, if trained on a biased sample they will reproduce this bias in their analysis of the data. For example, a number of recent studies have highlighted that algorithms can reinforce racial inequalities and discriminations.[100]

The effectiveness and fairness of the use of big data technologies in the context of the HREDD process will likely as well be affected by the biases of developers and by the choices they make to deal with what they perceive as the key dimensions of the process. In short, big data technologies are never value free and each use of these technologies comes with fundamental decisions, which are controversial and will influence the outcome of the HREDD process in one way or another, to the detriment of, for example, certain countries, companies or risks. This issue could be heightened in the HREDD context because the majority of big data technologies identified in this study are produced by companies (as well as data and computer scientists) headquartered in the global north.

4.2. Potential incompatibilities of the use of big data technologies with some requirements of the HREDD process

4.2.1. The corporations' ownership of the HREDD process

An important issue raised by stakeholders interviewed during this study is the responsibility of the company to take ownership of its own HREDD process. As foreseen by the UNGPs and the OECD Guidelines, it is essential that a company takes ownership and responsibility for its HREDD process, integrates it throughout its operations and develops internal capacity to support the process. With the adoption of big data technologies in this context, there is a risk that companies will simply consider paying to offload their due diligence obligations to the technology and its developers, not unlike they have done with auditing firms in the past. Indeed, the apparent neutrality and objectivity of technology and the simplicity of the conclusions reached by big data technologies might lead corporations to simply defer to their assessments and could lead them to refrain from internalizing HREDD concerns inside their corporate processes. In short, big data technologies could become a new way for corporations to fully externalize HREDD processes and to limit their direct involvement and ownership of the process thus risking defeating one of the purposes of HREDD, which is to change the way corporations consider their own responsibility for their adverse impacts on human rights and the environment. Ultimately, big data algorithms might liberate companies from the need to engage more directly with their adverse impacts and to take full responsibility for the decisions taken to tackle them.

4.2.2. The opacity of big data technologies and the transparency requirements of the HREDD process

The opacity linked with the use of big data technologies might also be seen as incompatible with the HREDD processes. The issue of the opacity of big data technologies and algorithmic accountability

96 UNHRC 'Racial discrimination and emerging digital technologies: a human rights analysis, Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance' (18 June 2020) UN Doc A/HRC/44/57. 12.

97 As put by Burns, '[t]he knowledge produced through Big Data technologies, data, and practices is always partial and reflects the geographical and social contexts of the people producing those knowledges.' See Ryan Burns, 'Rethinking Big Data in Digital Humanitarianism: Practices, Epistemologies, and Social Relations' (2015) 80(4) GeoJournal 477, at 484.

98 Sarah Myers West, Meredith Whittaker and Kate Crawford, Discriminating systems: gender, race and power in AI (New York, AI Now Institute, 2019).

99 Galit A. Sarfaty, 'Can Big Data Revolutionize International Human Rights Law?' (2017) 39(1) University of Pennsylvania Journal of International Law 73, at 93.

100 See, e.g., Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (New York, Penguin, 2016); Ruha Benjamin, Race After Technology (Polity Press, 2019); and Safiya Noble, Algorithms of Oppression: How Search Engines Reinforce Racism (New York, New York University Press, 2018).

has been recognized in the literature,[101] such as in the context of risk assessments by insurance companies in the US.[102] Indeed, big data technologies used by corporations are proprietary to the businesses which have developed them and are therefore not subjected to public scrutiny. In other words, it is currently very difficult for the public, regulators or the affected stakeholders to lift the hood of the technology and get an insider perspective on the algorithms and the analytical steps they follow to reach a particular conclusion. In the context of self-learning algorithms this might even be impossible, as AI technologies are not always explainable. Nevertheless, the findings of these algorithms will be relied upon by companies to take crucial decisions in the HREDD process or to justify certain decisions vis-a-vis public regulators. In many instances it will be impossible for the users, the regulators or the affected stakeholders to fully understand the reasons why a particular technology came to a specific risk assessment, to a specific conclusion regarding the level of the threat posed by a supplier, or how it identified the structure and composition of a supply chain. This might pose an issue in terms of the quality of the information communicated by the company in the framework of its HREDD process and raises the question whether information which does not allow the affected stakeholders or the regulators to understand how a particular company came to a specific conclusion is sufficient to satisfy to the transparency requirements of the HREDD process.[103] Accordingly, there is a risk that the use of big data technologies could turn HREDD into an opaque process, a black box, which would run counter to the transparency requirements embedded in the UNGPs and the OECD Due Diligence Guidance.

4.2.3. Big data technologies and the centrality of stakeholders' engagement in the HREDD process

Finally, engagement with rightsholders and, more broadly, relevant stakeholders is a critical component of the HRDD process. For instance, the process by which companies identify and assess actual or potential adverse human rights impacts should involve meaningful consultation with potentially affected groups and other relevant stakeholders.[104] Business enterprises, in particular, should seek to understand the concerns of potentially affected stakeholders by consulting them directly in a manner that takes into account language and other potential barriers to effective engagement.[105] Similarly, stakeholder engagement with digital technologies can be critical when collecting information on human rights risks, as relevant affected stakeholders, such as workers or local communities, are more likely to have concrete information on local instances of adverse human rights and environmental impacts.

While it is possible to envisage that technologies, including big data technologies, could be used to amplify the voice of, or to reach out to, certain stakeholders, who would be difficult to engage with otherwise,[106] there is also a potential risk that big data technologies will marginalise alternative modes of stakeholders' engagement. As companies have limited resources dedicated to their HREDD processes, and will compete to produce the most cost-efficient processes in order to maintain their profit margins, it is not excluded that businesses will decide to prioritise less costly big data technologies to the detriment of costly on-site engagement with stakeholders. This fear was shared by a number of our interviewees. In doing so, the quality of the engagement of companies with stakeholders, a key measure of an effective and legitimate HREDD process, might be affected. Consequently, the spread of big data technologies as supporting tools for HREDD might relegate direct stakeholder engagement to the margins of the process, while it was considered a fundamental pillar of its integrity by the father of HRDD, John Ruggie.[107] In other words, there is a potential for big data technologies to 'reinforce and expand a colonial-style relationship'[108] by empowering companies in the Global North to shape HREDD without really meaningfully consulting local stakeholders.[109]

101 See amongst many others Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press, 2016) and Karen Yeung, 'Algorithmic regulation: A critical interrogation' (2017) 12(4) Regulation & Governance 505.

102 Danielle Keat Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89(1) Washington Law Review 1.

103 See UNHRC, 'Guiding Principles on Business and Human Rights' (2011) UN Doc HR/PUB/11/04, Principle 21.

104 Ibid., Principle 18.

105 Furthermore, when tracking the effectiveness of their response to adverse human rights impacts, companies should draw on feedback from affected stakeholders (Principle 20). Business enterprises should also communicate how they address their human rights impacts in a way that does not endanger affected stakeholders or employees (Principle 21).

106 See Jesse Nishinaga and Faris Natour, 'Technology Solutions for Advancing Human Rights in Global Supply Chains' Human Rights and Business Initiative University of California (2019), at 17.

107 On the centrality of the participation of affected stakeholders in the HRDD process, see Antoine Duval, 'Ruggie's Double Movement: Assembling the Private and the Public Through Human Rights Due Diligence' (2023) Nordic Journal of Human Rights.

108 Miren Gutierrez and John Bryant, 'The Fading Gloss of Data Science: Towards an Agenda that Faces the Challenges of Big Data for Development and Humanitarian Action' (2022) 65 Development 80, at 89.

109 On the risk that data technology be based on an 'ersatz participative logic in which local communities feed data into the machine (either through crowd sourcing, or by being enumerators or subjects in most traditional surveys) but have little leverage on the design or deployment of the technology', see Roisin Read, Bertrand Taithe & Roger Mac Ginty 'Data hubris? Humanitarian information systems and the mirage of technology' (2016) 37(8) Third World Quarterly 1314, at 1324.

4.3. Adverse human rights and environmental impacts linked to the use of big data technologies

While this study focuses on the potential of big data technologies to prevent adverse human rights and environmental impacts in the supply chain, the human rights and environmental risks posed by these technologies must also be acknowledged.

4.3.1. Big data technologies and their adverse human rights impacts

The first well charted risk posed by big data technologies in terms of human rights is linked to the right to privacy.[110]The collection of large amounts of data from an array of digital sources and sensors, including personal data emitted by individuals as part of their daily lives (eg posting pictures on social media, navigating websites, or using a smartphone with GPS tracking operating in the background) may conflict with the rights to privacy, informed consent, and data autonomy.[111] Individuals may not have given consent for the collection or use of certain data in specific contexts, or may not be truly aware of what data they are sharing, the organizations that have access to it, and the uses to which the data will be put.[112] Some stakeholders interviewed for this study expressed concern that data collection should not violate the privacy of workers or anyone else linked to a particular supply chain. In addition, the existence of a commodified data market may exacerbate tensions regarding privacy and informed consent rights. A company can collect data directly or through a transaction with a third-party data provider. However, the practices of third-party data providers, also known as data brokers, who acquire, merge, analyse, and share data (including personal data) with other recipients, are largely shielded from public scrutiny and only marginally inhibited by existing legal frameworks.[113] As put by Latonero, 'it is unclear whether the uncertain benefits of long-term human rights monitoring and advocacy can outweigh the very concrete risks to privacy that accompany the use of big data'.[114] These concerns regarding the unwanted and monetized surveillance intimately connected with big data technologies extend a fortiori to the use of these technologies as surveillance mechanisms in the framework of the HREDD process.
Another human rights risk linked to the use of big data technologies is the one of exposing the messenger(s) to retaliation. Big data analysis could lead to the identification of individuals or groups which are complaining or opposing a particular corporation or project and denouncing its adverse impact on their rights or environments.[115] It is possible that this opposition once identified through a big data analysis and communicated to the connected corporations as part of their HREDD process will lead to the identification of the complainants and to retaliation. In short, the surveillance through big data analysis of local whistle-blower and activists in order to identify risks of adverse human rights or environmental impacts will need to be done in such a way as to protect their anonymity and safety.

The final, less charted, potential human rights issue that we identified is linked to the development of big data technologies related to the training of algorithms, which can be very labour intensive and might lead to abuses amounting to violations of fundamental labour rights.[116] This would especially be the case when algorithms are being trained to identify sensitive material, such as images or words amounting to human rights violations. Therefore, there is a nontrivial risk that the labour conditions under which big data technologies are developed to support HREDD processes will be themselves conducive of adverse impacts on the human rights of those involved in the development process. In sum, it is essential to ensure that the companies behind big data technologies for HREDD are themselves implementing rigorous HREDD processes and preventing the human rights risks arising out of the development of such technologies.

110 Even staunch advocates of the use of big data for humanitarian purposes consider data privacy and protection 'the biggest challenge', see Patrick Meier, 'Big (Crisis) Data Humanitarian Fact-Finding with Advanced Computing' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016), at 495. On this challenge, see also UNHRC 'The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights' (13 September 2021) UN Doc A/HRC/48/31, para 2; Mark Latonero, 'Big Data Analytics and Human Rights: Privacy Considerations in Context' in Molly K. Land and Jay D. Aronson (eds), New Technologies for Human Rights Law and Practice, (Cambridge University Press 2018); and Steve MacFeely, 'The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators' (2019) 10(1) Global Policy 121, at 130.

111 Alexander Kriebitz and Christoph Lütge, 'Artificial Intelligence and Human Rights: A Business Ethical Assessment' 5(1) Business and Human Rights Journal (2020) 1.

112 Galit A. Sarfaty, 'Can Big Data Revolutionize International Human Rights Law?' (2017) 39(1) University of Pennsylvania Journal of International Law 73, at 88-89.

113 UNHRC 'The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights' (13 September 2021) UN Doc A/HRC/48/31, para 13.

114 Mark Latonero, 'Big Data Analytics and Human Rights: Privacy Considerations in Context' in Molly K. Land and Jay D. Aronson (eds), New Technologies for Human Rights Law and Practice, (Cambridge University Press 2018), at 156.

115 On this double edge of big data analysis in the human rights context, see Jay D. Aronson, 'Mobile Phones, Social Media and Big Data in Human Rights Fact-Finding: Possibilities, Challenges, and Limitations' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016).

116 Billy Perrigo, 'OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make Chat GPT Less Toxic', Time Magazine, 18 January 2023 < https://time.com/6247678/openai-chatgpt-kenya-workers/> accessed 26 February 2023.

4.3.2. Big data technologies and their environmental footprint

Our interviewees have also stressed the fact that big data technologies, such as AI and cloud computing, can have substantial environmental impacts. For instance, AI can consume a significant amount of energy, particularly when training large AI models to conduct natural language processing, and has therefore a significant carbon footprint.[117] Furthermore, the rise of AI places a significant strain on the demand for rare earth metals and other raw materials needed to power significant additional computing resources, putting supplementary environmental pressure on resource extraction.[118] Similarly, cloud computing, which is the backbone of big data, and data centres have a significant environmental footprint characterised by high energy consumption, waste production, and CO_2 emissions.[119] The storage of voluminous amounts of data requires extensive facilities that use natural resources such as water and energy. Furthermore, the manufacturing and disposal of devices used for data collection and processing emit polluting substances. In this context, it has been argued that the 'data revolution, advocated as a vehicle to achieve sustainable development, is supported by technologies that endanger sustainability and the environment by other means'.[120]

Hence, if the core aim of HREDD is to ensure the sustainability and human rights compatibility of economic activities of corporations, it should be clearly demonstrated that the use of big data technologies has an overall positive impact on a company's environmental and human rights footprint.[121]

4.4. The risks of conflicts of interests and the use of big data technologies in HREDD

Finally, the use of big data technologies might give rise to conflicts of interest affecting the providers of such technological solutions. On the one hand, not unlike like auditing firms, the providers of big data technologies for HREDD will be remunerated by companies. Accordingly, it is likely that a big data technology solution which would identify many more risks than its competitors on the market, and therefore increases the compliance costs of its customers, might face difficulties in finding customers. In fact, in our interviews we were regularly told that the technology solutions could be tailored to fit the risk savviness of customers. In other words, it is possible that the companies offering such big data technologies will tweak them to be as attractive as possible to customers and not necessarily to be as comprehensive or as effective as possible in identifying risks to potentially affected stakeholders. The economic incentives and profit-driven nature of the companies devising these technologies might collide with the fundamental objective of HREDD, which is not to prevent adverse impacts to companies but to stakeholders and the environment. As mentioned, the same dilemma exists for auditing services,[122] which have strong economic incentives to under-report non-compliance or risks if they wish to keep being attractive to businesses. However, in this case combined with the opacity of the analytical process, this bias might be even more difficult to evidence in practice.

On the other hand, there is also a risk of conflict of interest when the same companies are providing identical or similar technical solutions to both the regulators and the regulated. Indeed, it is not unforeseeable that in the near future the regulators will be using similar big data technologies than companies to map supply chains and to assess the types of risks that should have been identified and prevented through the HREDD process of a specific company. However, in these types of situations, it is likely that the company providing these technical solutions to the regulator will be in a very favourable situation to market its technologies to companies. In such a situation, there is a risk that the clients of the tech company working with the regulator would get a more favourable assessment than other companies, as they are using the same algorithms as the agency in charge of overseeing them.

Most developers of big data technology for the HREDD process are driven primarily by profit motives. This means that the risk of conflict of interests is heightened when they are supposed to be acting in the interests of both their clients and third parties, such as affected stakeholders and the environment. Big data technologies, due to the objectivity and neutrality often attached to them, might give the impression that conflicts of interest would be less relevant than with, for example, auditing firms, this is not necessarily the case and public authorities should be aware of the potential for algorithmic opacity to be abused by private businesses to favour their economic interests to the detriment of stakeholders or the environment.

117 Emma Strubell, Ananya Ganesh and Andrew McCallum, 'Energy and Policy Considerations for Deep Learning in NLP' (2019) 34(9) EAAI Conference on Artificial Intelligence 13693.

118 Robert N. Boute and Maxi Udenio, 'AI in Logistics and Supply Chain Management' in Rico Merkert and Kai Hoberg (eds.) Global Logistics and Supply Chain Strategies for the 2020s: Vital skills for the Next Generation (Springer 2022), at 58.

119 Federica Lucivero, 'Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives' (2020) 26 Science and Engineering Ethics 1009, at 1010.

120 Ibid, 1013.

121 For a similar cautious approach to the impact of AI on sustainability, see Peter Dauvergne, AI in the Wild (MIT Press, 2020).

122 Highlighting the risk of capture of auditors, see Galit A. Sarfaty, 'Translating Modern Slavery into Management Management Practice' (2020) 45(4) Law and Social Inquiry 1027.

5. Conclusions

This study provided a first mapping of the potential for the use of big data technologies in the context of the HREDD process, it is informed by semi-structured interviews with experts and stakeholders and an extensive multidisciplinary literature review. In this process, we have identified both potential opportunities and challenges in harnessing big data technologies for HREDD.
On the one hand, the study has shown that there are essential steps of the HREDD process for which big data technologies could be or are already in use, such as for the mapping of the extensive upstream and downstream supply chains of corporations. Indeed, a number of companies are leveraging big data technologies to offer what they portray as a more comprehensive and reactive mapping of value, supply or commodity chains than would be possible without them. We have also identified that big data technologies are being used to detect and assess human rights or environmental risks or harm linked to the supply chains of companies. The most important added value of big data technologies in both instances lies in their capacity to cut through the intricacies of supply chains and to swiftly process vast amounts of data to identify a variety of risks or adverse impacts linked to the products or services of a particular company. In a global economy structured around complex transnational networks of exchanges and chaotic trade links, the use of big data technologies seems almost inevitable in order to monitor the impact of corporations on this extensive and highly flexible space overlapping with multiple communities and environments. Furthermore, big data technologies could also be helpful for the supervision of business compliance with HREDD requirements by public authorities and private third parties (such as NGOs or investors). In particular, they may be advantageous when reviewing corporate due diligence reports containing a large amount of complex data to analyse.

On the other hand, we have also identified potential risks or challenges connected with a widespread adoption of big data technologies in the framework of HREDD. First, we discussed challenges linked to the effectiveness of these tools. In particular, the veracity and the partiality of the data available might be leading to misdiagnoses and blind spots in HREDD processes. Furthermore, biases embedded in the algorithms might also vitiate their analyses and affect the quality of the findings. Second, we show that the use of big data technologies could be difficult to reconcile with certain requirements of the HREDD process, such as the required transparent communication on the process and comprehensive engagement with affected stakeholders. The opacity of big data analysis constitutes a formidable potential obstacle to the former, and the potentiality of entrusting most (if not all) the HREDD process to big data technologies could undermine the latter. Third, we believe there is a potential for the development and use of big data technologies to be linked with a number of environmental and human rights risks, which need to be carefully balanced against the benefits they provide. Fourthly, we see a risk of conflict of interests on the side of the companies developing these technologies, as they will have to mediate between the interests of their clients and the interests of the affected stakeholders and environments which the HREDD process is tailored to protect. Moreover, if the same technologies are being used by both regulated entities and regulators, this might also raise issues of conflicts of interests.

In conclusion, this study has shown that big data technologies have both in theory and in practice a potential role to play in the HREDD process. However, this development comes with risks and challenges which ought to be taken seriously by companies and public authorities. Gaining a more granular understanding of the effectiveness of the use of big data technologies in the HREDD processes, as well as of their potential adverse impacts will require further interdisciplinary research in years to come.

6. Annexes

6.1. Bibliography

Academic resources

- Matthew Archer, 'Imagining Impact in Global Supply Chains: Data-Driven Sustainability and the Production of Surveillable Space' (2021) 19(3) Surveillance & Society 282.
- Jay D. Aronson, 'Mobile Phones, Social Media and Big Data in Human Rights Fact-Finding: Possibilities, Challenges, and Limitations' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016)
- Mohamed Bakhouya et al, 'Cloud computing, IoT, and big data: Technologies and applications' (2020) 32(17) Current and Computation: Practice and Experience 159.
- Ruha Benjamin, Race After Technology (Polity Press, 2019)
- Robert N. Boute and Maxi Udenio, 'AI in Logistics and Supply Chain Management' in Rico Merkert and Kai Hoberg (eds.) Global Logistics and Supply Chain Strategies for the 2020s: Vital skills for the Next Generation (Springer 2022)
- Ryan Burns, 'Rethinking Big Data in Digital Humanitarianism: Practices, Epistemologies, and Social Relations' (2015) 80(4) GeoJournal 477
- Ricardo Chalmeta and Nestor J. Santos de León 'Sustainable Supply Chain in the Era of Industry 4.0 and Big Data: A Systematic Analysis of Literature and Research' (2016) 12(10) Sustainability 4108
- Feng Chen et al, 'Data Mining for the Internet of Things: Literature Review and Challenges' (2015) 11(8) International Journal of Distributed Sensor Networks
- Danielle Keat Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89(1) Washington Law Review 1
- Roger Clarke, 'Big data, big risks' (2016) 26(1) Information System Journals 77
- Peter Dauvergne, AI in the Wild (MIT Press, 2020)
- Rameshwar Dubey et al, 'Can big data and predictive analytics improve social and environmental sustainability?' (2019) 144 Technological Forecasting and Social Change 534
- Antoine Duval, 'Ruggie's Double Movement: Assembling the Private and the Public Through Human Rights Due Diligence' (2023) Nordic Journal of Human Rights
- John Emerson, Margaret L. Satterthwaite and Anshul Vikram Pandey, 'The Challenging Power of Data Visualization for Human Rights Advocacy' in Molly K. Land and Jay D. Aronson (eds), New Technologies for Human Rights Law and Practice, (Cambridge University Press 2018)
- D. Ganesh and P. Kalpana, 'Supply chain risk identification: a real-time data-mining approach' (2022) 122(5) Industrial Management & Data Systems 1333
- Sarah Giest and Annemarie Samuels, ''For good measure': data gaps in a big data world' (2020) 53 Policy Sciences 559
- Miren Gutierrez and John Bryant, 'The Fading Gloss of Data Science: Towards an Agenda that Faces the Challenges of Big Data for Development and Humanitarian Action' (2022) 65 Development 80
- Patrick C. K. Hung, Big Data Applications and Use Cases (Springer Cham 2016)
- Kristian Kersting and Ulrich Meyer, 'From Big Data to Big Artificial Intelligence? Algorithmic Challenges and Opportunities in Big Data' (2018) 32 Künstliche Intelligenz 3
- Yotsaphat Kittichotsatsawat, Varattaya Jangkrajarng and Korrakot Yaibuathet Tippayawong, 'Enhancing Coffee Supply Chain towards Sustainable Growth with Big Data and Modern Agricultural Technologies' (2021) 13(8) Sustainability 4593

- Alexander Kriebitz and Christoph Lütge, 'Artificial Intelligence and Human Rights: A Business Ethical Assessment' 5(1) Business and Human Rights Journal (2020) 1
- Mark Latonero, 'Big Data Analytics and Human Rights: Privacy Considerations in Context' in Molly K. Land and Jay D. Aronson (eds), New Technologies for Human Rights Law and Practice, (Cambridge University Press 2018)
- Genevieve LeBaron, Jane Lister & Peter Dauvergne, 'Governing Global Supply Chain Sustainability through the Ethical Audit Regime' (2017) 14(6) Globalizations 958
- Federica Lucivero, 'Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives' (2020) 26 Science and Engineering Ethics 1009
- Angela C. Lyons and John Grable J, 'An Introduction to Big Data' (2018) 72(5) Economics & Investment Management 17
- Steve MacFeely, 'The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators' (2019) 10(1) Global Policy 121
- Joash Mageto, 'Big Data Analytics in Sustainable Supply Chain Management: A Focus on Manufacturing Supply Chains' (2021) 13(13) Sustainability 7101
- Venkatesh Mani et al 'Mitigating Supply Chain Risk via Sustainability Using Big Data Analytics: Evidence from the Manufacturing Supply Chain' (2017) 9(4) Sustainability 608
- Andrew McAfee and Erik Brynjolfsson, 'Big Data: The Management Revolution' (2012) 90(10) Harvard Business Review 60
- Patrick Meier, Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response (Routledge 2015)
- Patrick Meier, 'Big (Crisis) Data Humanitarian Fact-Finding with Advanced Computing' in Philip Alston and Sarah Knuckey (eds.), The Transformation of Human Rights Fact-Finding (Oxford 2016)
- Fadi Muheidat et al, 'Emerging Concepts Using Blockchain and Big Data' (2022) 198 Procedia Computer Science 15
- Safiya Noble, Algorithms of Oppression: How Search Engines Reinforce Racism (New York, New York University Press, 2018)
- Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (New York, Penguin, 2016)
- Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press, 2016)
- Sarah Myers West, Meredith Whittaker and Kate Crawford, Discriminating systems: gender, race and power in AI (New York, AI Now Institute, 2019)
- Roisin Read, Bertrand Taithe & Roger Mac Ginty 'Data hubris? Humanitarian information systems and the mirage of technology' (2016) 37(8) Third World Quarterly 1314
- Galit A. Sarfaty, 'Can Big Data Revolutionize International Human Rights Law?' (2017) 39(1) University of Pennsylvania Journal of International Law 73
- Galit A. Sarfaty, 'Translating Modern Slavery into Management Management Practice' (2020) 45(4) Law and Social Inquiry 1027
- Peter Seele, 'Envisioning the digital sustainability panopticon: a thought experiment of how big data may help advancing sustainability in the digital age' (2016) 11(5) Sustainability Science 845
- Peter Seele, 'Predictive Sustainability Control: A review assessing the potential to transfer big data driven 'predictive policing' to corporate sustainability management' (2017) 153 Journal of Cleaner Production 673
- Lisa Smit et al, 'Human rights due diligence in global supply chains: evidence of corporate practice to inform a legal standard' (2021) 25(6) The International Journal of Human Rights 945
- Emma Strubell, Ananya Ganesh and Andrew McCallum, 'Energy and Policy Considerations for Deep Learning in NLP' (2019) 34(9) EAAI Conference on Artificial Intelligence 13693
- R.Taylor, C. Davis and J. Brandt, 'The rise of big data and supporting technologies in keeping watch on the world's forests' (2020) 22(1) International Forestry Review 129
- Karen Yeung, 'Algorithmic regulation: A critical interrogation' (2017) 12(4) Regulation & Governance 505

Reports and other publications

- Jennifer L. Cohen and Homi Kharas 'Using big data and artificial intelligence to accelerate global development' (Brookings, 15 November 2018) <https://www.brookings.edu/research/using-big-data-and-artificial-intelligence-to-accelerate-global-development/> accessed 24 February 2023
- Gabrielle Holly and Sarah Kajander Holmer, 'Sustainability reporting and human rights: What can big data analysis tell us about corporate respect for human rights?' (Danish Institute for Human Rights, 2022)
- Cornelia Hammer, Diane C. Kostroch and Gabriel Quiros 'Big Data: Potential, Challenges, and Statistical Implications' IMF Staff Discussion Note 2017, https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106, accessed 31 March 2023
- Doug Laney, '3D Data Management: Controlling Data Volume, Velocity and Variety' (META Group 2001)
- Jesse Nishinaga and Faris Natour, 'Technology Solutions for Advancing Human Rights in Global Supply Chains' Human Rights and Business Initiative University of California (2019)
- Claire M. O'Brien and Olga Martin-Ortega, 'EU Human Rights Due Diligence Legislation: Monitoring, Enforcement and Access to Justice for Victims' (Briefing N°2, European Parliament 2020)
- UN Global Pulse, 'Big data for Development: Challenges and Opportunities' (May 2012)
- WalkFree, The Future Society, WikiRate and BHRRC, 'AI Against Modern Slavery. Digital insights into modern slavery reporting – challenges and opportunities' (2021) < https://cdn.walkfree.org/content/uploads/2021/05/05173207/20210428-digital-insights-into-modern-slavery.pdf>
- Adrian Zenz, 'Coercive Labor in Xinjiang: Labor Transfer and the Mobilization of Ethnic Minorities to Pick Cotton', Intelligence Brief, Center for Global Policy, December 2020 <https://media.business-humanrights.org/media/documents/20201214-PB-China-Zenz-11.pdf> Accessed 10 January 2023

Online resources

- 'Cambridge Analytica parent firm SCL Elections fines over data refusal' (BBC News, 10 January 2019) <https://www.bbc.co.uk/news/technology-46822439> accessed 12 January 2023.
- 'ChatGPT: Optimizing Language Models for Dialogue' (OpenAI) <https://openai.com/blog/chatgpt/> Accessed 12 January 2023.
- 'Empowering people with open ESG data' (Wikirate) <https://wikirate.org/> accessed 12 February 2023.
- 'Making the link between human rights and sustainable development' (Danish Institute for Human Rights) <https://sdgdata.humanrights.dk/en> Accessed 24 February 2023.
- 'Project AIMS (Artificial Intelligence against Modern Slavery)' (The Future Society, 17 March 2021) <https://thefuturesociety.org/2021/03/17/project-aims-artificial-intelligence-against-modern-slavery/>. Accessed 24 February 2023.
- 'Open Supply Hub: Explore global supply chain data' (Open Supply Hub) https://opensupplyhub.org/> accessed 22 February 2023.
- 'What is natural language processing ?' (IBM) <https://www.ibm.com/topics/natural-language-processing> Accessed 24 February 2023.
- Vishal Gaur, Nikolay Osadchiy and Maximiliano Udenio, 'Research: Why It's So Hard to Map Global Supply Chains' (Harvard Business Review, 31 October 2022) <https://hbr.org/2022/10/research-why-its-so-hard-to-map-global-supply-chains> accessed 16 December 2022
- Billy Perrigo, 'OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make Chat GPT Less Toxic', Time Magazine, 18 January 2023 < https://time.com/6247678/openai-chatgpt-kenya-workers/> accessed 26 February 2023.
- Natasha Singer, 'What You Don't Know About How Facebook Uses Your Data', The New York Times, 11 April 2018 <https://www.nytimes.com/2018/04/11/technology/facebook-privacy-hearings.html> Accessed 20 January 2023.

UN Documents

- UNHRC 'Racial discrimination and emerging digital technologies: a human rights analysis, Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance' (18 June 2020) UN Doc A/HRC/44/57.
- UNHRC 'The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights' (13 September 2021) UN Doc A/HRC/48/31
- UNHRC, 'Guiding Principles on Business and Human Rights' (2011) UN Doc HR/PUB/11/04
- OHCHR, 'The Corporate Responsibility to Respect Human Rights: An Interpretive Guide' (2012) UN Doc HR/PUB/12/02

OECD Documents

- OECD, Due Diligence Guidance for Responsible Business Conduct (OECD Publishing 2018).
- OECD, Guidelines for Multinational Enterprises (OECD Publishing 2011)
- OECD, Due Diligence Guidance for Responsible Supply Chains in the Garment and Footwear Sector (OECD Publishing 2018)

6.2. List of interviewees

Grischa Beier, Institute for Advanced Sustainable Studies, 14/11/2022

Luke Smitham, Kumi Consulting, 15/11/2022

Adriana Bora, QUT Centre for Data Science, 16/11/2022

Lottie Lane, University of Groningen, 16/11/2022

Shannon Stewart & Juan Pinto, AltanaAI, 17/11/2022 & 25/1/2023

William Wallis, GroupTree, 25/11/2022

Laureen van Breen, WikiRate, 1/12/2022

Jesse Nishanaga, Article One, 2/12/2022

Gabrielle Holly, Omar Layachi & Stinne Skriver Jørgensen, Danish Institute of Human Rights, 5/12/2022

Farley Mesko, Sayari, 6/12/2022

Kate Robinson, Sedex, 6/12/2022

Leontien Hasselman-Plugge, ImpactBuying B.V., 6/12/2022

Lisa Hsin, Bonavero Institute of Human Rights, 8/12/2022

Tobias Streich, Transparency One, 9/12/2022

Gayatri Khandhadai, Business and Human Rights Resource Centre, 13/12/2022

Tim Bouten, Dow Jones Factiva, 14/12/2022

Isabella Bossi Fedrigotti & Thomas Loeber, Prewave, 14/12/2022 and 7/2/2023

Duncan Warner, Asda , 15/12/2022

Kush Wadhwa, Trilateral Research, 5/1/2022

Nanne Tolsma, Satelligence, 9/1/2023

Kiti Mignotte, Mana Vox, 11/1/2023

Joanne Bauer, Rights Colab, 27/1/2023

Alberto Zamora, Osapiens, 2/2/2023

Florian Woitek, Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA), 8/2/2023

Diederek Delen & Fatemah Ghaderinezhad, Amfori, 9/2/2023

6.3. Overview of relevant big data technology solutions

COMPANY	PRODUCT	DESCRIPTION
Altana AI	Atlas	The Altana Atlas is a supply chain map powered by an AI model of the supply chain, which learns from billions of data points describing businesses, facilities, supply chain flows, and ownership relationships worldwide
IFC (World Bank)	Malena	IFC's Machine Learning ESG analyst is an AI powered platform that aims to extract meaningful insights from unstructured ESG data at scale.
Elevate	EIQ	EIQ Sentinel service scans the web and media sources and sends alerts for supplier and vendor controversies relating to labor, health and safety, environment, business ethics and management systems
Mana Vox	MANA-Vox	MANA-Vox uses an AI-enhanced platform that can monitor individual corporations' involvement in local ecological controversies based on real-time information compiled from public information automatically collected from trusted sources.
Sayari	Sayari Graph	Sayari Graph uses big data technologies to identify and visualize supply chain networks as well as to provide risk analysis.
Prewave	Supply Chain Risk Intelligence	Uses big data and AI to provides risk alerts in line with the German Supply Chain Due Diligence law.
Satelligence	Satelligence	Satteligence combines satellite data with supply chain linkage data to provide insights on performance of agricultural production and supply chain risks
FACTSET	ESG Investing Solution	Provides access to ESG databases and ESG workflows through interface via artificial intelligence-powered analytics
I.point systems	I.Point Suite	Collects and analyses human rights-relate data from the supply chain and can be adjusted for company-specific requirements.
Orbital Insight	Orbital Insight Supply Chain Intelligence	Orbital Insight's Supply Chain Intelligence solution applies advanced analytics to a variety of data sources in order to improve the visibility and traceability of commodity supply chains.
Osapiens	Osapiens Hub	The Osapiens HUB claims to offer services to fully automate the risk evaluation in compliance with the German Supply Chain Law.
Sourcemap	Due Diligence	Supply chain mapping software automates the process of engaging stakeholders across the supply chain, including direct suppliers, indirect suppliers, and raw material producers. It also automates the process of risk and impact assessment by scoring every stakeholder using self-reported and third-party data, including the Department of Labor's List of Products of Forced and Child Labor and region-level forced labor risk heat maps.
Trilateral Research	STRIAD: Honeycomb	Honeycomb surfaces patterns, trends, and insights across huge volumes of text data and structured data in order to counter human trafficking. In doing so, it uses natural language processing (NLP) tools and data visualisations.
Trilateral Research	STRIAD: Airquality	Air quality monitoring in an area of interest claim to show where and when air pollution is at its worst. Track different pollutants over time and understand the trends where it matters.
Kharon	Clearview	Kharon ClearView uses large datasets to provide know your customers, investigations, and analysis to check if customers, supply chains, or other individuals or entities of interest are associated with sanctioned or trade-restricted parties.

COMPANY	PRODUCT	DESCRIPTION
Predik	Supply Chain Mapping Tool	Uses big data to visualise all the entities in the supply chain operation and identify patterns, events and behaviours.
Refinitiv	Due Diligence Centre	Provides real-time risk scoring for third parties and monitors and assesses risk areas to provide compliance screening.
Sphera	Supply Chain Risk Management	Uses internal and external data sources to assess through the use of AI supply chain risks. Claims to enable compliance with the German Supply Chain Law.
Earthqualizer	Spatial Intelligence Labs	Monitor agriculture and forestry supply chain No-deforestation, No-peat and No-exploitation commitments using a variety of data sources (High-resolution satellite imagery and aerial monitoring) and technologies (Artificial Intelligence, machine learning, deep learning).
Descartes Labs	Descartes Labs	Monitor deforestation and carbon equivalent emissions across the agricultural supply chain: Powered by remote sensing, machine learning, and multiple datasets.
Trase earth	Supply Chain	Follows trade flows to identify sourcing regions, profile supply chain risks and assess opportunities for sustainable production.
Starling Verification	Starling	Starling uses a combination high-resolution satellite imagery, advanced processing and machine learning for regular deforestation monitoring.
Global Fishing Watch	Global Fishing Watch	Uses tracking data from the publicly available identification system (AIS) and integrates this with information acquired through vessel monitoring systems, vessel registry data and satellite imagery. Utilises machine learning to flag when a boats' location beacons are intentionally turned off, suggesting unregulated fishing activity.
OpenSC	OpenSC	Uses data science and machine learning to verify ethical production claims, and internet of things technologies to trace products across the supply chain.
Kenzen	Kenzen	Uses a wearable device for industrial workers which can detect risks such as heat stress and fatigue that could lead to injuries.

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 32 + 36
53113 Bonn, Germany
T +49 228 44 60-0
F +49 228 44 60-17 66

Dag-Hammarskjöld-Weg 1 - 5
65760 Eschborn, Deutschland
T +49 61 96 79-0
F +49 61 96 79-11 15

E info@giz.de
I www.giz.de