



**HAL**  
open science

# Towards a Balanced Natural Language Processing: A Systematic Literature Review for the Contact Centre

Lungile Binza, Adheesh Budree

► **To cite this version:**

Lungile Binza, Adheesh Budree. Towards a Balanced Natural Language Processing: A Systematic Literature Review for the Contact Centre. 17th International Conference on Social Implications of Computers in Developing Countries (ICT4D), May 2022, Lima, Peru. pp.397-420, 10.1007/978-3-031-19429-0\_24 . hal-04601171

**HAL Id: hal-04601171**

**<https://inria.hal.science/hal-04601171v1>**

Submitted on 5 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# TOWARDS A BALANCED NATURAL LANGUAGE PROCESSING: A SYSTEMATIC LITERATURE REVIEW FOR THE CONTACT CENTRE

Balancing the AI Triple Challenge of *opportunity*, *ethics*, and *opportunity cost*!

Lungile Binza  
University of Cape Town  
Department of Information Systems  
[bnzlun003@myuct.ac.za](mailto:bnzlun003@myuct.ac.za)

Adheesh Budree  
University of Cape Town  
Department of Information Systems  
[adheesh.budree@uct.ac.za](mailto:adheesh.budree@uct.ac.za)

## Abstract

Artificial Intelligence (AI), which is the design, development, and utilisation of iterative and complex algorithmic systems that complete tasks which normally required human intelligence, is rapidly gaining momentum throughout the world. Through Machine Learning these AI systems automatically learn and adapt themselves so that they can offer an even more accurate outcome than humans and this offers many exciting benefits to most businesses and economies. But their introduction has raised ethical questions after some unethical conduct on their part like racism, biasness against women, unemployment (through intelligent automation), and inequality. After these incidents of unethical behaviour by some AI technologies around the globe most public, private, and even non-profit organisations embarked on initiatives to address the unethical conduct of AI systems. They produced a range of ethical AI frameworks, guidelines, and principles, but a properly balanced AI, where *opportunity*, *ethics*, and *opportunity costs* intersect, is still to be achieved or realised. Most AI Practitioners are pressured by their shareholders to prioritise commercial interests over ethical considerations.

The findings from this Systematic Literature Review study, which used meta-analyses for qualitative synthesis, demonstrate that a **Balanced** Natural Language Processing (NLP) is possible.

**Keywords:** Artificial Intelligence, Machine Learning, Balanced AI, Natural Language Processing, Ethical AI, Systematic Literature Review

## 1. Introduction

Artificial Intelligence (AI) is the study of the design, development, application, and operationalisation of complex algorithmic systems that train themselves to think and act as intelligently as human beings (Schrader and Ghosh, 2018). Powered by Machine Learning (ML) algorithms, where they adaptively interpret and continuously self-learn through analyzing new large datasets (Walsh et al., 2019; Floridi and Cowls, 2019; Bankins, 2021; Giermindl et al, 2021), these “intelligent virtual agents” recognize and respond to specific environments (Acemoglu and Restrepo, 2019; Teng, 2020) that would have otherwise required human-level intelligence and intervention (Biller-Andorno et al, 2020; Naude and Dimitri, 2020). These environments, which require pattern identification in data and the generation of predictions to assist in decision making processes (Nadimpalli, 2017; Biller-Andorno et al, 2020; Krijger, 2021), include *Knowledge Reasoning*, *Machine Learning (ML)*, *Robotics*, *Computer Vision*, *Visual Perception*, *Speech Recognition*, *Voice Biometrics*, and *Natural Language Processing (NLP)* (Coombs et al, 2020).

NLP as an “interdisciplinary field of AI where the machine can be used to understand and manipulate natural language text or speech” (Bender and Friedman, 2018; Li et al, 2021) includes *text generation*, *speech recognition*, *question answering*, *speech to text*, and *text to speech*. In this research study, the term AI would refer mainly to NLP, since the focus is on a Banking Contact Centre suite of AI technologies.

The AI industry forecasts some huge economic promise, potential, and benefits (Brynjolfsson & McAfee, 2014, 2017), it’s resulting in things getting ‘easier, cheaper, and abundant’ (Naude and Dimitri, 2020). AI has the potential to increase human wellbeing, boost most sectors of the economies, improve various societies, and general good quality customer service (Arambula and Bur, 2020; James and Whelan, 2022). As Dolganova (2021) points out AI increases “customer loyalty, trust, and could boost sales by up 30% per annum”.

But, with all these undoubted economic and commercial success stories, AI has the transformative social and cultural influence in society (Mika et al, 2019). AI has also demonstrated a pattern of entrenching social divides and amplifying or worsening social inequality, particularly among historically marginalized groups (Hagerty and Rubinov, 2019).

Some few examples that demonstrate AI unethical behaviour in the United States are:

- (a). “a job recruiting tools being biased against women” (Hagerty and Rubinov, 2019);
- (b). “Latin and African American borrowers faced with discriminatory credit algorithms” (Hagerty and Rubinov, 2019);
- (c). “bias regarding race, gender, and/or sexual orientation in sentiment analysis systems, and natural language processing technologies” (Hagerty and Rubinov, 2019); and
- (d). “an AI legal system that discriminated against African American and Hispanic men when making decisions about granting parole which has since become infamous” (Taddeo and Floridi, 2018).
- (e). “the US recidivism prediction algorithm that allegedly mislabelled African American defendants as “high-risk” at nearly twice the rate as it mislabelled white defendants” (Angwin et al., 2016; Chouldechova, 2016) from Krijger, 2021).

- 
- (g). “hiring algorithms that, based on analysing previous hiring decisions, penalized applicants from women’s colleges for technology related positions” (Dastin, 2018).
  - (h). “a healthcare entry selection program that exhibited racial bias against African American patients” (Obermeyer et al., 2019)
  - (i). “Google uses AI to identify photos, people, and objects in GOOGLE’s Photos Services and always it will not be able to produce the correct results. Also, if any camera misses mark on racial sensitivity then the predicted criminals will be always the black people” (Alotaibi, 2018).

All of these incidents and instances led to governments, academics, civil society or non-profit organisations, research institutes, and even the private sector to embark on initiatives to ensure ethical AI behaviour. Ethical AI which broadly refers to “the fair and just development, use, and management of AI technologies” (Bankins, 2021) was initiated. Since AI Systems are more socio-technical in nature, considering ethical implications on their design, development, deployment, and operation became of critical importance for the society. Non-governmental organisations like **AI4Good** led to the development of the **AI4People** ethical AI Framework by Floridi et al. in 2018 (Kindylidi and Cabral, 2021); initiatives by European Commission’s High-Level Expert Group on Artificial Intelligence (AI-HLEG) led to the publishing of Ethics Guidelines for **Trustworthy AI or Sustainable AI** (Kindylidi and Cabral, 2021); The Japanese Society for Artificial Intelligence (JSAI) identified the ethical principles to be followed by developers of artificial intelligence systems (Dolganova, 2021); The Atomium – European Institute for Science, Media and Democracy white paper on AI ethics outlines five principles of AI ethics (Dolganova, 2021); **Responsible AI** which is “concerned with the design, implementation, and use of ethical, transparent, and accountable AI technology in order to reduce biases, promote fairness, equality, and to help facilitate interpretability and explainability of outcomes” (Trocin et al, 2021) was also born.

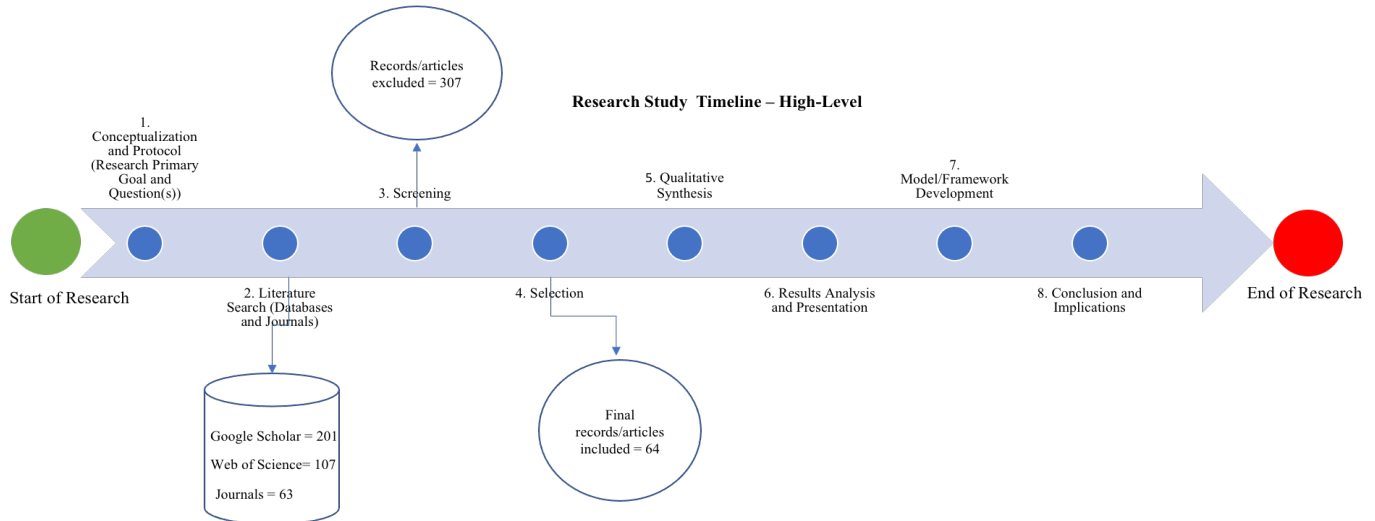
These initiatives were not only in the non-profit, governmental, or research institutes based only but the private sector also came on board:

- (a) IBM launched an “Everyday Ethics for Artificial Intelligence: A practical guide for designers and developers” (Ashby, 2020);
- (b) Microsoft leadership uses the “six principles mainly derived from the AI4People framework to guide its use of their AIs” (Bankins, 2021); and
- (c) Google as also formulated “seven principles of artificial intelligence that the company is following in creating and using AI technologies” (Dolganova, 2021).

Most AI Practitioners are pressured to prioritise commercial interests over public or ethical consideration (Mittelstadt, 2020). This raises a question of whether a **Balanced AI**, where business benefits or commercial viability of AI (*opportunity*); *ethics* in AI design and development, and AI utilisation or operation (*opportunity cost*) are all balanced? So the research question and problem is on whether an ethically designed, developed, implemented, and utilised NLP for Contact Centres in South Africa continue having positive effects on business performance? This question though cannot be fully answered in this SLR publication as the case study is still to be undertaken, but the conceptual framework(s) gained from this exercise will be used as a guide to conduct the case research. The primary goal for this SLR publication was to conduct thematic analysis for academic literature to achieve a *Balanced NLP*.

## 2. Methodology

A systematic literature review (SLR) (Paré et al., 2015) was conducted to identify relevant research papers with meta-analyses for qualitative synthesis. It followed 8 stages namely: conceptualization and protocol identification, searching the literature, screening the articles, selecting the relevant articles, performing some thematic analysis, presenting the results, developing the framework, and then concluding. These stages are illustrated in figure 1 below:



**Figure 1: Systematic Literature Review and Meta-Analyses research flow**

The literature search, screening, and selection was conducted in 5 steps as shown in table 1 below. These are sources searched, search keywords, search strategy, inclusion criteria, and exclusion criteria.

### (i). Sources Searched

Two main databases, Web of Science and Google Scholar, were used as the main sources for the literature search. Also about 6 mainly IS journal websites were also used. These are Journal for Management Information Systems (JMIS); European Journal of Information Systems (EJIS); MIS Quarterly; Information Systems Journal (ISJ); Journal of Strategic Information Systems (JSIS); and Journal of Information Technology (JIT).

### (ii). Search Keywords

The keywords “AI Implementation”; “AI Adoption”, “Ethical AI Implementation”, “NLP”, and “Ethical NLP” were used to conduct the literature search from the above databases and journal websites. A total of over 350 articles were found.

### (iii). Search Strategy

All the websites of the 2 identified Databases and Journals were interrogated. Using the combination of the above keywords, the search strategy was for articles from the last 3 years: January 2018 to December 2021. Interestingly, there were about three relevant articles from *MIS Quarterly*, the *International Journal of Information Management*, and *Critical Social Policy* that were already accepted for publishing in 2022 and those 3 papers were also included.

Step	Step Description	Detailed Step Description	No of Articles
#1	Sources Searched	<b>Databases:</b> (i). Web of Science, and (ii). Google Scholar <b>Journals:</b> (a). Journal for Management Information Systems (JMIS); (b). European Journal of Information Systems (EJIS); ©. MIS Quarterly; (d). Information Systems Journal (ISJ); (e). Journal of Strategic Information Systems (JSIS); and (f). Journal of Information Technology (JIT)	Journals n = 63; Databases n = 308
#2	Search Terms	The Keywords used during the process were: “AI Implementation”; “AI Adoption”, “Ethical AI Implementation”, “NLP”, and “Ethical NLP”	Journals n = 63; Databases n = 308
#3	Search Strategy	Went to all the websites of the 2 identified Databases and Journals. Searched for articles in the last 3 years (2018 to 2021)	Journals n = 20; Databases n = 44
#4	Inclusion Criteria	All publications between Jan 2018 and Dec 2021 with ethical AI principles and guidelines	Journals n = 20; Databases n = 44
#5	Exclusion Criteria	Articles older than 2018 and those that were not entirely relevant to the ethical AI cause	Journals n = 43; Databases n = 264

**Table 1: Steps for searching and selecting the articles**

#### (iv). Inclusion Criteria

With the search strategy discussed in (iii) above, the results were +/- 64 relevant articles selected. From those articles, Google Scholar contributed about 45 articles, which accounts for about 70% of all the included publications, and Web of Science about 19, which accounts for 30% of the articles.

The 2021 calendar year seemed to be the most prominent with over 33% of all ethical AI related articles being published that year. 2020 publications follow with 22% slightly ahead of 2019 articles at 22% were tied second at around 22% each, as shown in table 2 below.

Year	Count	Percentage (%)
2018	11	17%
2019	14	22%
2020	14	22%
2021	22	34%
2022	3	5%
<b>Total</b>	<b>64</b>	<b>100%</b>

**Table 2: Year of article publications included in the SLR**

From the journal, the Association for Computational Linguistics and European Journal of Information Systems had the biggest number of articles with 4 each, as shown in table 3 below. The Institute for Electronic and Electrical Engineers (IEEE), and Minds and Machines came second tied at 3 articles each. There was a very long tail of 1 publication each from various academic institutions, research institutes, MIS Quarterly, and other journals like Journal of Strategic Information Systems.

Publication	Count	H-Index
Association for Computational Linguistics	4	71
AI & SOCIETY	2	
European Journal of Information Systems	4	108
Global AI Ethics	1	
Oxford	2	
Journal of Information Technology	2	79
Minds and Machines	3	39
International Journal of Information Management	2	114

**Table 3: Some Selected journal publications included in the SLR > 1 article**

#### (v). Exclusion Criteria

All publications before January 2018 were excluded and also those that were not entirely relevant to the debate of ethical AI. This resulted in around 307 articles being excluded from the SLR. The field of ethical implementation and utilisation of AI tends to be highly dynamic and Ethical AI is a phenomenon that has recently been thrust into the spotlight, hence the decision to restrict the literature search to the last 3 years.

### 3. Analysis of Literature Identified

All the literature that has been reviewed came from the 64 relevant articles that have been included in the study using the combination of the keywords “*AI Implementation/Adoption*”; “*Ethical AI Implementation/Adoption*”, “*NLP*”, and “*Ethical NLP*”. These are the articles published in the last 3 years, from 2018, that deal with AI ethical related issues.

Floridi and Cowls (2019) posit that the ethical debate around AI is almost as old as the 1950s and 1960s (Samuel, 1960; Wiener, 1960). However, it is only in recent years that impressive advances in the capabilities and applications of AI systems have brought the opportunities and risks of AI for society into sharper focus (Yang et al., 2018). Just to recap, Table 4 below shows some definitions of artificial intelligence from various authors and researchers:



No	AI Definition	Author	Publisher & (Year)
1	AI are individualized applications being developed by businesses and used by consumers to power human decision-making processes, reduce the time needed to complete everyday tasks, and automatically inform us of events and incidents in the social and political world around us.	Schrader and Ghosh	IEEE (2018)
2	AI is a growing resource of interactive, autonomous, and often self-learning agency that can deal with tasks that would otherwise require human intelligence and intervention to be performed successfully.	Floridi and Cowls	Harvard (2019)
3	AI is increasing capability of machines to perform tasks, which have been conducted traditionally by people.	Mika et al	VTT Technical Research Centre of Finland (2019)
4	The theory and development of computer systems able to perform tasks normally requiring human intelligence, is widely heralded as an ongoing “revolution” transforming science and society altogether	Jobin et al	Health Ethics & Policy Lab (2019)
5	AI encompasses various interrelated technologies often underpinned by machine learning algorithms, whereby AI achieves set objectives via supervised (with human guidance) or unsupervised (machine autonomous) learning through analyzing large datasets	Walsh et al.	Ethics and Information Technology (2019)
6	AI refers to ‘machines that act intelligently when a machine can make the right decision in uncertain circumstances.	Naude and Dimitri	AI & Society (2020)
7	AI has the ability to decide what task to execute under diverse circumstances and sometimes beyond human capabilities and understanding and doesn’t use fixed algorithm to perform a tasks.	Teng	<i>AIBotics</i> (2020)

**Table 4: Artificial Intelligence (AI) Definitions**

### 3.1 Ethical Issues AI

Alotaibi (2018) list 9 leading ethical issues related with artificial intelligence: *Unemployment; Inequality; Humanity; Artificial Stupidity; Racist Robots; Security; Evil Genies; Singularity; Robot Rights*. These are exactly the same ethical issues that had been raised by Julia Bossmann at the World Economic Forum (WEF) in 2016 which demonstrated that there’s some unanimity when it comes to ethical AI issues.

**Table 5: Leading Ethical Issues**

<b>Ethical Issue</b>	<b>Brief Description</b>
<b>Unemployment</b>	Is this the end of jobs? What happens after the end of jobs?
<b>Inequality</b>	How do we ensure the income from AI machines is equitable shared?
<b>Humanity</b>	Can the machines influence our behaviours, interaction or relationship building with others?
<b>Artificial stupidity</b>	How do we prevent mistakes, especially in healthcare industry where it could have catastrophic consequences in human life?
<b>Racist robots</b>	How do we eliminate AI bias and injustice?
<b>Security</b>	Cybersecurity is even more important now, how do we keep AI safe from adversaries?
<b>Evil genies</b>	What if AI itself turns against humanity? How do we protect against unintended consequences?
<b>Singularity</b>	Humans controls dangerous animals because of ingenuity and intelligence. Can they do the same for complex and intelligent systems like AI?
<b>Robot rights</b>	Is there a humane way to treat AI systems? How do we know they require some 'human' rights?

**Source: adapted from Bossmann (2016); and Alotaibi (2018)**

On unemployment, the rise of Intelligence Automation (IA) – intelligent automation of knowledge and service work (Coombs et al, 2020) – could lead to job substitution especially for office work and clerical tasks, sales and commerce, transport and logistics; manufacturing industry, construction, some aspects of financial services, and some types of services like translation, and tax consultancy (Degryse, 2016). IA differs from previous forms of automation because of machine learning where AI machines can learn, adapt and improve over time, based on new available data (Coombs et al, 2020; Coombs, 2020), which could lead *job polarization*. Job Polarization is a phenomenon where “advance technology seems to be displacing workers away from middle-skill/middle-pay jobs down to low-skill/low-wage jobs, where these workers further depress low-skill wages, or for a lucky few who are retrained, up to high-skill jobs where the workers enjoy higher productivity and higher wages” (Dau-Schmidt, 2017).

### 3.2 Global AI Ethical Initiatives

Table 5 above on leading AI ethical issues and the unethical AI behaviours listed in the introduction page have led to the global initiatives to combat unethical AI design, development, implementation, and utilisation, especially in the global north. Ethical AI broadly refers to “the fair and just development, use, and management of AI technologies” (Bankins, 2021, 5). Nakata et al (2021) list some major global initiatives led by governments, national and international organisations, research institutes, civil society (non-profit), and private sector pertaining to AI principles for ethics since 2016 (Jobin et al, 2019). Some selected of those initiatives are listed below:

- (i) 5 principles on “*Ethically Aligned Design*” from IEEE in 2017;
- (ii) 6 principles on “*Future Computed*” by Microsoft in 2018;

- 
- (iii) 9 principles on “*Statement on AI, Robotics, and ‘Autonomous’ Systems*” by European Group on Ethics in Science and New Technologies in 2018;
  - (iv) 5 principles on “AI in the UK: ready, willing, and able” by the Select UK Committee on AI in 2018;
  - (v) 5 principles on “*Everyday Ethics for Planning*” by IBM in 2018;
  - (vi) 7 principles on “*Sony Group AI Ethics Guidelines*” by Sony in 2018;
  - (vii) 5 principles on “*An Ethical Framework for a Good AI Society*” by AI4People in 2018
  - (viii) 7 principles on “*Principles of Human-Centric AI Society*” by the Japan’s Cabinet Office in 2019; and
  - (ix) 4 principles on “*Ethics Guidelines for Trustworthy AI*” by the EU’s High-Level Expert Group on Artificial Intelligence on AI (AI-HLEG) in 2019.

**Source: adapted from Nakata et al (2021)**

These initiatives and the subsequent principles and guidelines have led to the emergence of terms like AI4Good and AI4People Framework (Floridi et al., 2018; Cowls et al, 2021), Trustworthy AI (Mökander and Floridi, 2021; Cowls et al, 2021), Responsible AI (Kindylidi and Cabral, 2021), and Emotional AI, (McStay, 2019). Some of these will be discussed in detail below:

### 3.2.1 Responsible AI

Responsible AI is concerned with the design, development, implementation, and utilisation of ethical, transparent, and accountable AI technology in order to reduce biases, promote fairness, equality, and to help facilitate interpretability and explainability of outcomes (Trocin et al, 2021). It is a fairly new phenomena that investigates the ethics of AI to understand the moral responsibility of these technologies (Kindylidi and Cabral, 2021).

### 3.2.2 Trustworthy AI

Ethics Guidelines on Artificial Intelligence (‘the Guidelines’) were released by the European Commission’s High-Level Expert Group on Artificial Intelligence on AI (AI-HLEG) for Trustworthy AI. developing trustworthy AI by ensuring that systems are lawful, ethical and robust (European Commission, 2019).

In general, these frameworks consist of several key themes, exemplified by the HLEG Guidelines for trustworthy AI that lay out seven key requirements: (European Commission, 2019; Cowls et al, 2021; Krijger, 2021):

- (i) **Human Agency and Oversight**, “AI systems should allow humans to make informed decisions and be subject to proper oversight.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021);
- (ii) **Technical Robustness and Safety**, “AI systems need to be resilient, secure, safe, accurate, reliable, and reproducible.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021);
- (iii) **Privacy and Data Governance**, “Adequate data governance mechanisms that fully respect privacy must be ensured.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021);
- (iv) **Transparency**, “The data, system and AI business models should be transparent and explainable to stakeholders.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021);

- (v) **Diversity, Non-discrimination and Fairness**, “Unfair bias must be avoided to mitigate the marginalisation of vulnerable groups and the exacerbation of discrimination.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021);
- (vi) **Societal and Environmental well-being**, “AI systems should be sustainable and benefit all human beings, including future generations.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021);
- (vii) **Accountability** “Responsibility and accountability for AI systems and their outcomes should be ensured.” (European Commission, 2019; Cowls et al, 2021; Krijger, 2021).

### 3.2.3 Research Institutes on Ethical AI

Below are just some selected research institutes that also developed some ethical AI principles.

(a). The **Capgemini Research Institute** also designed their six core principles and characteristics of ethical (Dolganova, 2021): (a). “ethical actions from design to application” (Capgemini Research Institute, 2019); (b). “transparency” (Capgemini Research Institute, 2019); (c). “explainability of the functioning of AI” (Capgemini Research Institute, 2019); (d). “the interpretability of the results” (Capgemini Research Institute, 2019); (e). “fairness, lack of bias” (Capgemini Research Institute, 2019); (f). “the ability to audit” (Capgemini Research Institute, 2019).

(b). The **European Institute for Science, Media and Democracy** published a white paper outlining five principles of AI ethics (Floridi et al, 2020; Dolganova, 2021): (a). Promoting human well-being (Floridi et al, 2020); (b). Harmlessness (“confidentiality, security, and ‘attention to opportunities’”) (Floridi et al, 2020); (c). Autonomy (“the right of people to make their own decisions”) (Floridi et al, 2020); (d). Fairness (“respect for the interests of all parties that can be influenced by the actions of the system with AI, the absence of discrimination, the possibility of eliminating errors”) (Floridi et al, 2020); (e). Explainability (“transparency of the logic of artificial intelligence, accountability”) (Floridi et al, 2020).

Other fields, especially in Health Care, have developed their Frameworks for ethical AI considerations like in Pathology: Transparency, Accountability, and Governance (Chauhan and Gullapallyz, 2021) are of critical importance in an AI System and in Otolaryngology: Consent and Patient Privacy, Beneficence, Non-Maleficence, and Justice (Arambula and Bur, 2020) are the principles adopted.

### 3.2.4 Private Sector Ethical AI Initiatives

The private sector, especially the Tech Companies, also developed some practical standards and guidelines in designing and developing AI within their companies:

(a). IBM’s “Everyday Ethics for Artificial Intelligence: A practical guide for designers and developers” (IBM, 2018; Ashby, 2020) were published in 2018 and are in full operation: (i). Purpose expressed as unambiguously prioritized goals (IBM, 2018); (ii). Truth about the past and present (IBM, 2018); (iii). Variety of possible actions (IBM, 2018); (iv). Predictability of the future effects of actions (IBM, 2018); (v). Intelligence to choose the best actions (IBM, 2018); (vi). Influence on the system being regulated (IBM, 2018); (vii). Ethics expressed as unambiguously prioritized rules (IBM, 2018); (viii). Integrity of all subsystems (IBM, 2018); (ix). Transparency of ethical behaviour (IBM, 2018).

(b). Microsoft also developed and published five principles that their leadership uses in order to guide their use of AI (Bankins, 2021). These are mainly linked to the AI4People framework (Bankins & Formosa, 2021; Bankins, 2021) and are: (i). **fairness** (“aligned to the justice norm”) (Microsoft, n.d.); (ii). **reliability and safety** (“aligned to the beneficence norm”) (Microsoft, n.d.); (iii). **privacy and security** (“aligned to the non-maleficence norm”) (Microsoft, n.d.); (iv). **inclusiveness** (“aligned to the justice norm”) (Microsoft, n.d.); and (v). **transparency and accountability** (“aligned to the explicability norm”) (Microsoft, n.d.).

(c). Google as also developed and published *seven* principles that they use to design, develop, and utilise these AI type technologies (Dolganova, 2021): (i). “AI should be socially useful” (Pichai, 2018); (ii). “it is necessary to strive to avoid unfair influence on people” (Pichai, 2018); (iii). “application of best security practices” (Pichai, 2018); (iv). “responsibility for the actions of AI in front of people” (Pichai, 2018); (v). “ensuring guarantees of confidentiality, proper transparency and control over the use of data” (Pichai, 2018); (vi). “maintaining standards of excellence” (Pichai, 2018); (vii). “limiting the use of potentially harmful and offensive software products” (Pichai, 2018).

### 3.3 AI4People Framework

The AI4Good initiative led to the development AI4People Unified Framework of Principles for AI in Society (Floridi et al., 2018; Floridi and Cowls, 2019; Floridi et al, 2020). This framework seem to encapsulate all the earlier ethical AI principles and consolidated them into 5 unified ethical AI principles: (a). **Beneficence**: “Promoting Well-Being, Preserving Dignity, and Sustaining the Planet” (Floridi et al., 2018; Floridi and Cowls, 2019; Beil et al, 2019; Floridi et al, 2020); (b). **Non-maleficence** “Privacy, Security and ‘Capability Caution’” (Floridi et al., 2018; Floridi and Cowls, 2019; Beil et al, 2019; Floridi et al, 2020); (c). **Justice**: “Promoting Prosperity and Preserving Solidarity” (Floridi et al., 2018; Floridi and Cowls, 2019; Beil et al, 2019; Floridi et al, 2020); (d). **Explicability**: “Enabling the Other Principles Through Intelligibility and Accountability” (Floridi et al., 2018; Floridi and Cowls, 2019; Beil et al, 2019; Floridi et al, 2020); (e). **Autonomy**: “The Power to Decide (to Decide)” (Floridi et al., 2018; Floridi and Cowls, 2019; Beil et al, 2019; Floridi et al, 2020).

### 3.4 AI Ethical Design

The robustness of the design and the agility of the architecture of Trustworthy AI is extremely important and must be built into the system from the beginning (Mökander and Floridi, 2021). That could also lead to a process of continuous improvement of the re-design process. The IEEE has setup 4 Ethically Aligned Design (EAD) standards to address the ethical concerns of AI (Alotaibi, 2018): (a). “Model Process for Addressing Ethical Concerns during System Design” (Alotaibi, 2018); (b). “Transparency of Autonomous Systems” (Alotaibi, 2018); (c). “Data Privacy Process” (Alotaibi, 2018); and (d). “Algorithmic Bias Considerations” (Alotaibi, 2018).

Also, the Japanese Society for Artificial Intelligence (JSAI) articulated some ethical principles to be followed by designers and developers of the AI systems (Dolganova, 2021): These include (a). respect for human rights and respect for cultural diversity (JSAI, 2017); (b). compliance with laws and regulations, as well as not harming others (JSAI, 2017); (c) respect for privacy (JSAI, 2017);

- (d). justice (JSAI, 2017);
- (e). security (JSAI, 2017);
- (f). good faith (JSAI, 2017);
- (g). accountability and social responsibility (JSAI, 2017);
- (h). self-development and promotion of understanding of AI by society (JSAI, 2017).

Leidner and Plachouras (2017) had originally proposed 6 ethical by design principles: (i). “proactive and not reactive”; (ii). “ethical set as a default setting”; (iii). “ethics embedded into the process”; (iv). “end to end ethics”; (v). “visibility and transparency”; and (vi). “respect for user values”. These were further reiterated by Chauhan and Gullapalliyyz (2021) in their *Inclusive AI Design and No Bias* design principles. Schrader and Ghosh (2018) also proposed an ethical AI development Framework which includes the following five components: (a). “Identify ethical issues of AI”; (b). “Improve human awareness of AI”; (c). “Engage in dialogical collaboration with AI”; (d). “Ensure the accountability of AI”; (e). “Maintain the integrity of AI”.

### **3.5 NLP Business Benefits**

In most businesses their primary goal or objective when implementing an AI technology or AI innovation is to enable product and service innovation, enhance the customer experience, improve customer service, increase efficiency and productivity, and also Improve decision making (Amazon, 2020). In some other cases, like financial services industry, risk assessment and fraud detection and prevention are another priority (Nadimpalli, 2017). Because the focus area in this AI study is the NLP, which is mainly utilised in the services sector, the AI business benefits of interest would be revenue generation and cost optimisation, customer service, efficiency improvements and high productivity, and competitiveness.

#### **3.5.1 Revenue Generation & Cost Optimisation:**

Revenue and sales forecasts with the deployment of the NLP solution are astronomically. For instance, the NLP can process a thousand times more the queries and transactions that would have otherwise been done by over 100 people (Nadimpalli, 2017). That coupled with some cost savings and optimisation use cases like “a Bank that saved over \$9 million to customers after implementing an AI tool to scan small business transactions for fake invoices” (Quest et al, 2018). The scaling of these AI technologies is also simple but the benefits are so huge. Through AI implementation in the Anti-Money-Laundering space in banking, the incidents of suspicious activity increased to more than 20 times, but the resolution is also quicker (Quest et al, 2018).

#### **3.5.2 Customer Experience**

The deployment of the AI in customer services environment has significantly enhanced customer experience (Dolganova, 2021) and this became a huge competitive advantage. The introduction of the NLP in the financial services improved the Customer Satisfaction Index (CSI) score, improved the Complaints to Compliments ratio, and the Net Promoter Score (NPS) (Quest et al, 2018). With the forecast that over 80% of consumers are likely to buy a brand that offers personalised service, the need for AI to enhance customer experience is even bigger (Arambula and Bur, 2020; Amazon, 2020).

### 3.5.3 Efficiency Improvements and enhance Productivity

The implementation of AI in various areas improves accuracy and efficiencies like in the contact centres the NLP empowered customers through self-service, shortens the Average Handle Time (AHT), especially on query resolution, significantly improves contact quality since upfront authentication enabled, and reduces customer effort (Leins et al, 2020). In the health care sector, AI improves efficiency and accuracy through image analysis, robot-assisted surgery, and drug discovery by augmented the ability to provide quality health care (Hague, 2017). These AI models do not only predict the image analysis outcomes, but also identify the specific areas on an image for examination and showing the level of confidence in the prediction (Hague, 2017).

### 3.5.4 Competitive Advantage or Strategic Competition

NLP is a strategic competition tool as it allows you to gain new customers, but also significantly enhance service offering the existing customers, which is a huge competitive advantage (Arambula and Bur, 2020).

## 4. Analysis, Synthesis, and Interpretation

After all the ethical AI initiatives undertaken, Jobin et al (2019) contended that there are around 84 ethical AI guidelines and around 11 principles: transparency; justice and fairness; non-maleficence; responsibility; privacy; beneficence; freedom and autonomy; trust; dignity; sustainability; and solidarity. But Floridi et al (2020) team, which were part of the AI4Good initiative, agreed that there's some evidence of convergence of the 11 principles into around 5 principles which are part of the AI4People Ethical AI principles: transparency; justice and fairness; non-maleficence; responsibility; and privacy.

Of the 84 AI guidelines, 43, which makes just over 50%, is shared between 5 countries who are more economically developed: USA, UK, Germany, France, and Finland (Jobin et al, 2019). The rest of the guidelines is shared between EU, EC, Australia, Canada, and EU institutions.

### 4.1 The AI Triple Challenge

The triple AI challenge is trying to ensure that the AI is designed, developed, and operated in such a way that the business *opportunity*, the *ethical* considerations of the AI, and the *opportunity cost* are all balanced (Floridi et al, 2018). If these three components or themes, as we call them in this thematic analysis study, are fully considered then a Balanced AI (NLP) will be born. The *ethics* - avoidance of overuse or wilful misuse of the AI - and the business benefits - *opportunity* from the NLP (AI) innovation - emanating from the adoption of the AI have been discussed above, but the opportunity cost (underuse) hasn't been fully explored. The primary objective is to avoid the misuse and the underuse of the NLP (AI) technologies (Floridi et al, 2018).

The opportunity cost of the AI is the underuse of the AI technology (Floridi et al, 2018). This is the potential benefit that can be forgone or missed out because of an option that was not chosen in AI. Since it's unseen it is easy to overlook and opportunity cost. Within AI, the broader risk is that it may be underused out of fear of overuse or misuse, hence the need to balance these in the Balanced NLP (AI) (Floridi et al, 2018).

### 4.1.1 Opportunity Theme

For the Balanced NLP (AI) to be realized, the NLP must satisfy all the necessary business case metrics, especially on return on investment. Within the NLP realm, Customer Experience, Revenue Generation & Cost Optimization, Competitiveness, and Efficiency Improvements & High Productivity are the best business metrics to satisfy.

**Table 6: Summary of the SLR Outcome of the Opportunity Theme**

Theme	Principle	Guidelines	Authors
Opportunity	<b>Customer Experience</b>	<ul style="list-style-type: none"> <li>■ Improved Customer Satisfaction index;</li> <li>■ Good Customer Service</li> </ul>	Dolganova, 2021; Quest et al, 2018
	<b>Revenue Generation &amp; Cost Optimization</b>	<ul style="list-style-type: none"> <li>■ Sales revenues up;</li> <li>■ Costs optimised</li> </ul>	Quest et al, 2018
	<b>Competitiveness</b>	<ul style="list-style-type: none"> <li>■ Gain new customers;</li> <li>■ Retain existing customers</li> </ul>	Arambula and Bur, 2020
	<b>Efficiency Improvements &amp; High Productivity</b>	<ul style="list-style-type: none"> <li>■ Average Handle Time on queries improved</li> <li>■ empowered customers through self-service</li> <li>■ Good Contact Quality</li> </ul>	Hague, 2017; Leins et al, 2020

### 4.1.2 Ethics Theme

The outcome of the SLR ethics theme of the AI triple challenge is summarised in table 7 below. Beneficence & Dignity; Justice & Fairness; Accountability & Responsibility; Human Agency & Oversight; Non-Maleficence, Governance, Privacy, and Security; and Transparency, Autonomy, & Freedom are the ethical principles that are consolidated into.

#### 4.1.2.1 Beneficence & Dignity

For the AI to achieve the Balanced AI status in the global south it must benefit everyone. In South Africa in particular there's a crisis of *inequality*, *unemployment*, and *poverty* and any AI to be trusted it must create jobs and no job substitution or staff reduction. Job creation and augmentation must be at the centre instead of staff reduction or job polarization. The AI shouldn't further exacerbate the inequality problem instead everyone must benefit equally: company, employees, and customers. Basically, the AI must be seen to addressing some of the socio-economic issues, which are really endemic in this society.

#### 4.1.2.2 Justice & Fairness

The AI must also be seen to be helping out in ensuring fairness in society. The global south comes from some painful past of colonisation and exploitation and the AI must contribute in correcting some of those injustices. Racism was rife and AI shouldn't be seen to be furthering some racist behaviours. The languages, dialects, and accents vary from region to region and AI must adept to these through machine learning. There should be no biasness and proper pronunciation of African names would to an AI that could be trusted.



#### 4.1.2.3 Non-Maleficence, Governance, Privacy, and Security

Most countries in the global south are underdeveloped because of lack of service delivery and corruption by elected officials. In order for the Balanced AI to be trustworthy, it must be seen to be exposing these corrupt practices and not to be misused by the corrupt elite. It must comply with the laws like POPIA in South Africa and should stand in solidarity with the marginalised.

**Table 7: Summary of the SLR Outcome of the Ethics Theme**

Theme	Principles	Guidelines	Authors
Ethics	<b>Beneficence &amp; Dignity</b>	Promoting Well-Being, Preserving Dignity, and Sustaining the Planet	Floridi et al., 2018; Beil et al, 2019; Floridi and Cows, 2019; Jobin et al, 2019 European Commission, 2019; Arambula and Bur, 2020 Floridi et al, 2020
	<b>Justice &amp; Fairness</b>	<ul style="list-style-type: none"> <li>■ Unfair bias must be avoided to mitigate the marginalisation of vulnerable groups and the exacerbation of discrimination.</li> <li>■ Respect for the interests of all parties that can be influenced by the actions of the system with AI, the absence of discrimination, the possibility of eliminating errors</li> </ul>	Floridi et al., 2018; Pichai, 2018 ; Beil et al, 2019; Floridi and Cows, 2019; Jobin et al, 2019; Capgemini Research Institute, 2019; European Commission, 2019; Arambula and Bur, 2020; Floridi et al, 2020; Bankins, 2021; Dolganova, 2021; Krijger, 2021; Microsoft, n.d.
	<b>Human Agency &amp; Oversight</b>	<ul style="list-style-type: none"> <li>■ Promoting human well-being</li> <li>■ AI systems should allow humans to make informed decisions and be subject to proper oversight</li> </ul>	Jobin et al, 2019 European Commission, 2019; Floridi et al, 2020; Cows et al, 2021; Krijger, 2021
	<b>Non-Maleficence, Governance, Privacy, and Security</b>	<ul style="list-style-type: none"> <li>■ No misuse or overuse of the AI systems</li> <li>■ AI systems need to be resilient, secure, safe, accurate, reliable, and reproducible</li> <li>■ confidentiality, security, and 'attention to</li> </ul>	Floridi et al., 2018; Jobin et al, 2019; Pichai, 2018 ; European Commission, 2019; Beil et al, 2019; Arambula and Bur, 2020;

		opportunities' Adequate data governance mechanisms that fully respect privacy must be ensured	Floridi et al, 2020; Bankins, 2021; Cowls et al, 2021; Dolganova, 2021; Krijger, 2021; Microsoft, n.d.
	<b>Transparency, Autonomy, &amp; Freedom</b>	<ul style="list-style-type: none"> <li>■ The data, system and AI business models should be transparent and explainable to stakeholders</li> <li>■ Explainability of the functioning of AI</li> <li>■ The right of people to make their own decisions</li> <li>■ The power to decide</li> </ul>	IBM, 2018; Pichai, 2018; Floridi et al., 2018; Schrader and Ghosh, 2018; Jobin et al, 2019; European Commission, 2019; Capgemini Research Institute, 2019; Beil et al, 2019; Floridi and Cowls, 2019; Arambula and Bur, 2020; Floridi et al, 2020; Bankins, 2021; Krijger, 2021; Chauhan and Gullapallyz, 2021; Microsoft, n.d.

### 4.1.3 Opportunity Cost Theme

From the reviewed literature, the principles that could be grouped under the opportunity cost theme are *Transparency* and *Explainability*, as well as *Accountability* and *Responsibility*. The transparency principle is as an opportunity cost as is the ethics and makes a second appearance. For the balanced NLP (AI) to be a reality, these AI systems should be have people to explain their inner workings (features and functionality) so that they are not underused or underutilised.

**Table 8: Summary of the SLR Outcome of the Opportunity Cost Theme**

Theme	Principles	Guidelines	Authors
	<b>Transparency &amp; Explainability</b>	<ul style="list-style-type: none"> <li>■ The data, system and AI business models should be transparent and explainable to stakeholders</li> <li>■ Explainability of the functioning of AI</li> </ul>	IBM, 2018; Pichai, 2018; Floridi et al., 2018; Schrader and Ghosh, 2018; Jobin et al, 2019; European Commission, 2019; Capgemini Research Institute, 2019;

<b>Opportunity Cost</b>			Beil et al, 2019; Floridi and Cowls, 2019; Arambula and Bur, 2020; Floridi et al, 2020; Bankins, 2021; Krijger, 2021; Chauhan and Gullapalliyyz, 2021; Microsoft, n.d.
	<b>Accountability &amp; Responsibility</b>	Responsibility and accountability for AI systems and their outcomes should be ensured.	Schrader and Ghosh, 2018; Beil et al, 2019; Chauhan and Gullapalliyyz, 2021; Dolganova, 2021; European Commission, 2019; Cowls et al, 2021; Krijger, 2021

## 4.2 Ethical AI Design

Based from the data of the SLR the ethical AI design principles can be consolidated into of *Data Privacy and Security*, *No Algorithmic Bias*, *Integrity of Subsystems*, and *Interpretability and Robustness of the AI*, as shown in table 9 below. All other ethical design considerations can be accommodated with these principles. Therefore, for a Balanced AI to be achieved, these ethical AI design principles must be adhered to.

**Table 9: Summary of the SLR Outcome of the Ethical AI Design**

<b>Theme</b>	<b>Principles</b>	<b>Guidelines</b>	<b>Authors</b>
<b>Ethical AI Design</b>	<b>Data Privacy and Security</b>	■ Security for personal data and access control while working with AI	IEEE, 2009; Alotaibi, 2018;
	<b>No Algorithmic Bias</b>	■ Inclusive AI Design and no Bias	IEEE, 2009; Shahriari and Shahriari (2017); Chauhan and Gullapalliyyz, 2021
	<b>Integrity of subsystems</b>	■ Visibility and transparency: design process published for public scrutiny and criticism	Leidner and Plachouras, 2017; IBM, 2018; Schrader and Ghosh, 2018;
	<b>Interpretability and Robustness of the AI</b>	■ Build interpretability and robustness into AI systems from the start	IEEE, 2009; Shahriari and Shahriari (2017);

		<p>■ Ethics embedded in the process end to end (E2E)</p>	<p>Leidner and Plachouras, 2017; Mökander and Floridi, 2021;</p>
--	--	--	--

### 4.3 Balance NLP Conceptual Framework

Based on the above Ethical AI Challenge themes identified and the ethical AI design principles, Figure 1 below represents the recommended conceptual framework for the Balanced NLP from this SLR.

The ethical AI design is fundamental in ensuring the future ethical behaviour of the Balanced AI. The integrity of the data and it’s privacy is critical since AI depends on the availability of data to continue learning and adapting. Those will lead to robust algorithms that won’t end up performing unethical conduct. So the design is the input and the start of this conceptual framework.

The opportunity, ethics, and the opportunity are the outcomes of the model. From an *ethics* perspective, the Balanced AI recommends 5 ethical AI principles instead of the 11 ethical AI principles as suggested by Jobin et al (2019) and the 5 solid principles suggested by Floridi et al (2018, 2020), which is not really far from the Floridi et al (2018, 2020) teams: (i) Beneficence & Dignity; (ii). Justice & Fairness; (iii). Human Agency & Oversight; (iv). Non-Maleficence, Governance, Privacy, and Security; and (v). Transparency, Autonomy, & Freedom. As discussed in their guidelines above, if these ethical principles are adhered to, the Balanced AI would be achieved.

From an *opportunity cost* perspective, there’s two main principle to consider that should be the outcome of the Balanced AI: (i). *Transparency and Explainability*, and (ii). *Accountability and Responsibility*. For fear of underutilising the AI System, the ability to explain and understand all the features and functionality of the system is fundamental, hence transparency and explainability.

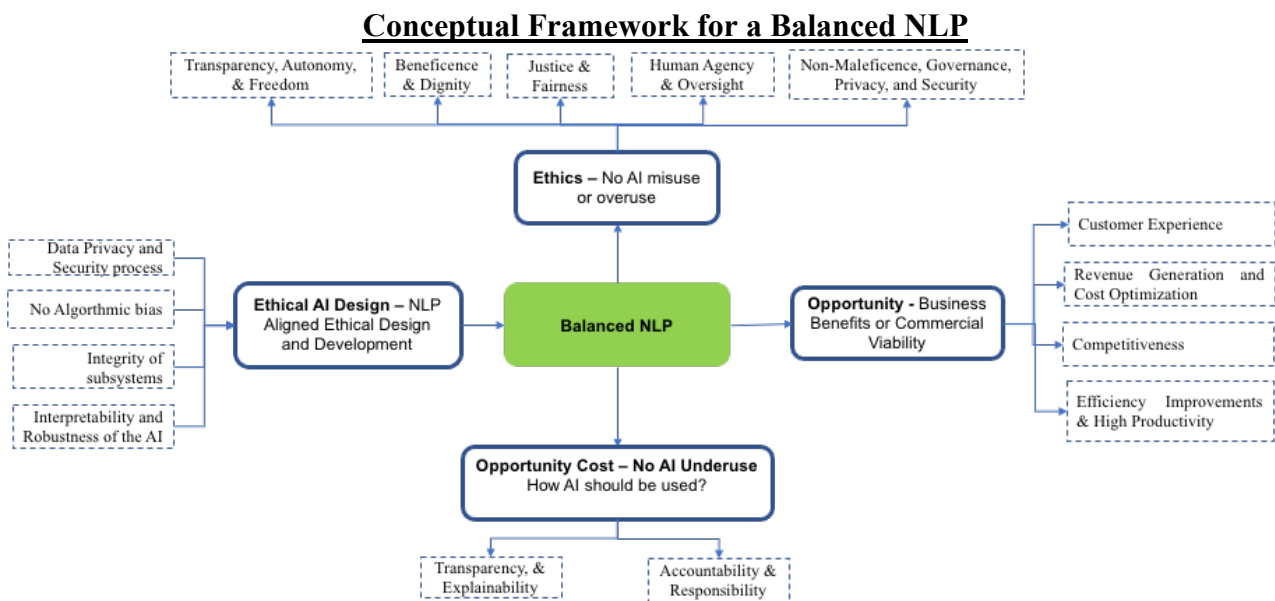


Figure 2: Conceptual Framework for a Balanced NLP

The Opportunity theme or business benefits, the Balanced NLP, should be able to meet the return on investment (ROI) targets of Customer Experience, Revenue Generation & Cost Optimization, Competitiveness, and Efficiency Improvements & High Productivity. There are various metrics to test and validate these as part of the operational environment and they would need to be regularly re-visited to check the continued sustainability on the Balanced AI investment.

## 5. Recommendation

The above conceptual framework in figure 2 will be used as a guide for the case research study where the influence of ethical AI will be tested against the business models of the Contact Centre. The approach of the proposed case research study will be inductively conducted with qualitative data collection technique and thematic data analysis to test for the conceptual framework's validity.

### 5.1 Balanced NLP

For the Balanced NLP to be achieved, the intersection of ethics, opportunity, and opportunity cost should be proven, as seen in figure 3 below:

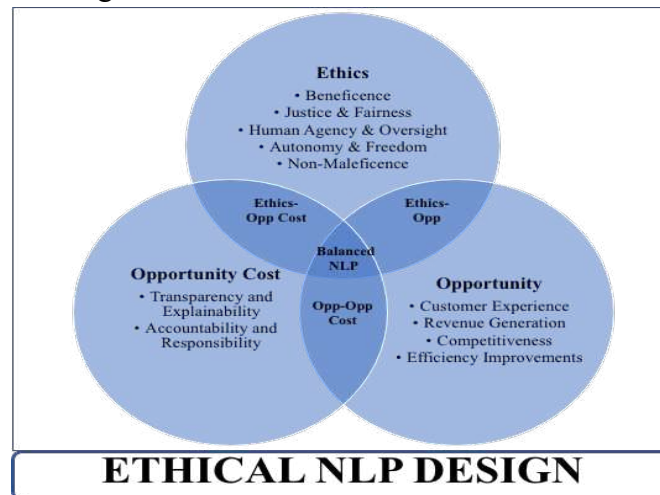


Figure 3: Balanced NLP - Intersection of ethics, opportunity, and opportunity cost

#### (i). Ethics - Opportunity

From this SLR, the link or intersection between the opportunity and ethics themes is critically important for the Balanced NLP to be achieved. This couldn't be clearly demonstrated in this review as the case research isn't finalised yet, but it will need to be fully explored and explained.

#### (ii). Ethics - Opportunity Cost

The link between ethics and opportunity cost themes has been demonstrated in this SLR where ethics and opportunity cost principles converge. The *Transparency* and *Explainability* principles, and *Accountability* and *Responsibility* principles are seen as part of the opportunity cost theme, but most authors classify them as ethics principles. In the final research case study, not much emphasis will be put to this intersection.

**(iii). Opportunity - Opportunity Cost**

From this SLR, the link or intersection between the opportunity and opportunity cost themes is also of critically importance for the Balanced NLP to be achieved. This couldn't be clearly demonstrated in this review as the case research isn't finalised yet, but it will need to be fully explored and explained.

**(iv). Balanced NLP**

The opportunity theme plays a central role in achieving this Balanced NLP. Its relationship with both the ethics and opportunity cost theme will determine whether a Balanced NLP - where ethics, opportunity, and opportunity cost intersect - is possible. This will be fully explored and explained in the case research study.

**6. Concluding Remarks**

The ethical AI design process is as important as any of the above themes in ensuring a Balanced NLP outcome. The ethical AI design is encompassing and is overarching for all the 3 other themes as its application is important in guaranteeing the ethicalness of the AI.

Finally, to constantly check if the NLP continues to meet the required standards of Balanced AI, the ethics-based auditing process would be recommended. This a governance mechanism that is used by businesses that design and deploy AI systems so as to control the influence or behaviour of the AI system (Mökander and Floridi, 2021). This is an audit process that continuously evaluate and monitor system outputs and report on the system's performance characteristics (Mökander and Floridi, 2021).

**6.1 Future Research areas**

As already indicated above, the conceptual framework in figure 2 above will be validated in the research case study to determine the influence of ethical AI on business performance or business models in the Contact Centre. Other areas of study will be influenced by the outcome of that research case.

**References**

- Acemoglu, D. & Restrepo, P. (2017): Robots and Jobs: Evidence from US Labor Markets.
- Acemoglu, D. and Restrepo, P. (2018). Artificial Intelligence, Automation and Work: NBER Working Paper No. 24196 from <http://www.nber.org/papers/w24196>
- Acemoglu, D. and Restrepo, P. (2019): The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand, IZA Discussion Papers, No. 12292, Institute of Labor Economics (IZA), Bonn
- Alotaibi, S. S. (2018). Ethical Issues and Related Considerations Involved with Artificial Intelligence and Autonomous Systems. *International Journal of Advanced Computer Science and Applications, Vol. 9, No. 4*
- Amazon, (2020). "How Amazon Is Using AI To Better Understand Customer Search Queries" <https://analyticsindiamag.com/how-amazon-is-using-ai-to-better-understand-customer-search-queries/> [Accessed on 15 January 2022]
- Arambula, A. M., and Bur, A. M. (2020). Ethical Considerations in the Advent of Artificial Intelligence in Otolaryngology. *Head and Neck Surgery Vol. 162(1) 38–39*

- 
- Ashok, M.; Madan, R.; Joha, A.; and Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management* 62, 102433 <https://doi.org/10.1016/j.ijinfomgt.2021.102433>
- Ashby, M. (2020). Ethical Regulators and Super-Ethical Systems. *Systems*, 8, 53; doi: <https://doi.org/10.3390/systems8040053>
- Bertin, M. and Songül, T. (2018): Will this time be different? A review of the literature on the Impact of Artificial Intelligence on Employment, Incomes and Growth. JRC Digital Economy Working Paper, No. 2018-08, European Commission, Joint Research Centre (JRC), Seville
- Bai, Z., Jain, N., Kurdyukov, R., Walton, J., Wang, Y., Wasson, T., & Chircu, A. M. (2019). Conducting Systematic Literature Reviews in Information Systems: An Analysis of Guidelines. *Issues in Information Systems*, 20(3).
- Bankins, S. (2021). The ethical use of artificial intelligence in human resource management: a decision-making framework. *Ethics and Information Technology* Vol 23, pp. 841–854 <https://doi.org/10.1007/s10676-021-09619-6>
- Beil, M., Proft, I., van Heerden, D., Sviri, S., and van Heerden, P. V. (2019). Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensive Care Medicine Experimental* 7:70. <https://doi.org/10.1186/s40635-019-0286-6>
- Bender, E. M. and Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, vol. 6, pp. 587–604
- Biller-Andorno, N.; Andrea Ferrario, A.; Joebges, S.; Krones, T.; Massini, F.; Barth, P.; Arampatzis, G.; and Krauthammer, M. (2020). AI support for ethical decision-making around resuscitation: proceed with care. <https://doi.org/10.1101/2020.08.17.20171769>
- Bossmann, J. (2016). Top 9 ethical issues in artificial intelligence. <https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/> [accessed on the 10<sup>th</sup> Jan 2022]
- Brynjolfsson, E. & McAfee, A. (2012): *Race Against The Machine: How The Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and The Economy*. MIT Sloan School of Management
- Capgemini Research Institute. (2019). *Why addressing ethical questions in AI will benefit organizations*. Available at: [https://www.capgemini.com/wp-content/uploads/2019/08/AI-in-Ethics\\_Web.pdf](https://www.capgemini.com/wp-content/uploads/2019/08/AI-in-Ethics_Web.pdf) (accessed 11 January 2022).
- Chauhan, C., and Gullapallyz, R. R. (2021). Ethics of AI in Pathology Current Paradigms and Emerging Issues. *The American Journal of Pathology*, Vol. 191, No. 10
- CNBC, (2017) “Jack Ma’s Artificial Intelligence and Automation Warning” [Video]. <https://youtu.be/cgAPG3Lh3gQ?t=70/>, [Accessed on 14 January 2022].
- Coombs, C., Hislop, D., Taneva, S.K., & Barnard, S. (2020): The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems*, 101600
- Coombs, C. (2020). Will COVID-19 be the tipping point for the Intelligent Automation of work? A review of the debate and implications for research. *International Journal of Information Management*. Reader in Information Systems, School of Business and Economics, Loughborough University, LE11 3TU, United Kingdom

- 
- Cowls, J., Taddeo, M., Tsamados, A., and Floridi, L (2021). Achieving a ‘Good AI Society’: Comparing the Aims and Progress of the EU and the US. SSRN Electronic Journal <https://doi.org/10.2139/ssrn.3851523>
- Dau-Schmidt, K. G. (2017) "The Impact of Emerging Information Technologies on the Employment Relationship: New Gigs for Labor and Employment Law". Articles by Maurer Faculty. 2657
- Degryse, C. (2016): Digitilisation of the economy and its impact on labor markets - European Trade Union Institute - Working Paper, 2016
- Dolganova, O. I. (2021). Improving customer experience with artificial intelligence by adhering to ethical principles. BUSINESS INFORMATICS Vol. 15 No 2 pp. 34–46 <https://doi.org/10.17323/2587-814X.2021.2.34.46>
- European Commission. (2019, April 8). *Ethics guidelines for trustworthy AI* [Text]. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- Floridi L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Rossi, U. P. F., Schafer, B., Valcke, P., Vayena, E. (2018): AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations: Minds and Machines. 28:689–707
- Floridi, L., & Cowls, J. A. (2019). Unified framework of five principles for AI in society. Harvard Data Science Review Issue 1.1. <https://doi.org/10.1162/99608f92.8cd550d1>
- Floridi L., Beltrametti M., Burri T., Chatila R., Chazerand P., Dignum V., Madelin R., Luetge C., Pagallo U., Rossi F., Schafer B., Valcke P., Vayena E. (2020) *AI 4People’s ethical framework for a good AI society: Opportunities, risks, principles, and recommendations*. Available at: <https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf> (accessed 10 January 2022).
- Frey, C. B. and Osborne, M. A. (2013): The Future Of Employment: How susceptible are jobs to computerisation? Oxford Martin Programme on Technology and Employment
- Giermindl L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2021): The dark sides of people analytics: reviewing the perils for organisations and employees, European Journal of Information Systems
- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines Vol 30: pp 99–120 <https://doi.org/10.1007/s11023-020-09517-8>
- Hague, D. C. (2017). Benefits, Pitfalls, and Potential Bias in Health Care AI
- Hirschheim, R., & Klein, H.K. (2003). Crisis in the IS Field? A Critical Reflection on the State of the Discipline: *Journal of the Association for Information Systems Vol. 4 No. 10, pp.237-293/October 2003*
- Huysman, M. (2020). Information systems research on artificial intelligence and work: A commentary on “Robo-Apocalypse cancelled? Reframing the automation and future of work debate”. *Journal of Information Technology 2020, Vol. 35(4) 307–309, SAGE*
- IBM Developer Staff (2018) *AI Fairness 360*. Available at: <https://developer.ibm.com/technologies/artificial-intelligence/projects/ai-fairness-360/> (accessed 08 January 2022).
- IEEE (2009). “Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems”. IEEE
- James, A.; and Whelan, A. (2022). ‘Ethical’ Artificial Intelligence in the welfare state: Discourse and discrepancy in Australian social services. *Critical Social Policy, Vol. 42(1): 22–42* <https://doi.org/10.1177/0261018320985463>



- Jobin, A.; Ienca, M.; and Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. Health Ethics & Policy Lab
- JSAI (2017). *The Japanese Society for Artificial Intelligence Ethical Guidelines*. Available at: <http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf> (accessed 14 January 2022).
- Kindylidi, I.; and Cabral, T.S. (2021). Sustainability of AI: The Case of Provision of Information to Consumers. Sustainability, 13, 12064. <https://doi.org/10.3390/su132112064>
- Kodiyar, A. A. (2019). An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool
- Krijger, J. (2021). Enter the metrics: critical theory and organizational operationalization of AI ethics. AI & SOCIETY <https://doi.org/10.1007/s00146-021-01256-3>
- Leidner, J. L., and Plachouras, V. (2017). Ethical by Design: Ethics Best Practices for Natural Language Processing. Proceedings of the First Workshop on Ethics in Natural Language Processing, pages 30–40, Association for Computational Linguistics
- Leins K., Lau J.H., Baldwin T. (2020): Give Me Convenience and Give Her Death: Who Should Decide What Uses of NLP are Appropriate, and on What Basis? Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 2908–2913 July 5 - 10, Association for Computational Linguistics
- Li, Y., Thomas, T., & Liu, D. (2021): From semantics to pragmatics: where IS can lead in Natural Language Processing (NLP) research, European Journal of Information Systems, 30:5, 569-590
- Makridakis, S. (2017). The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository. Retrieved from <http://hdl.handle.net/11728/9254>
- McStay, A. (2019). Emotional AI and EdTech: Serving the Public Good? Learning, Media and Technology, Vol 45 Issue 3, pp. 270-283. <https://doi.org/10.1080/17439884.2020.1686016>
- Microsoft. (n.d.). Responsible AI. <https://www.microsoft.com/en-us/ai/responsible-ai>
- Mökander, J., and Floridi, L. (2021). Ethics-Based Auditing to Develop Trustworthy AI. Minds and Machines
- Nadimpalli, M. (2017): Artificial Intelligence Risks and Benefits. International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 6, June 2017
- Nakata, T., Araki, T., Tsuchiya, S., Nakao, Y., Naseer, A., Arahori, J., and Yamamoto, T. (2021). Initiatives for AI Ethics: Formulation of Fujitsu Group AI Commitment. Findings of the Association for Computational Linguistics: ACL-IJCNLP, pages 3895–3901
- Naudé, W. and Dimitri, N. (2020). The race for an artificial general intelligence: implications for public policy. AI & SOCIETY Vol.35 pages 367–379
- Omary, J. (2018): Jobs and University Skills: Artificial Intelligence Effects on Employment in the Future. Memorial University of Newfoundland
- Oosterwyk, G., Brown, I., & Geeling, S. (2019). A Synthesis of Literature Review Guidelines from Information Systems Journals. In *Proceedings of 4th International Conference on the Internet, Cybersecurity and Information Systems*. 12, 250-260.
- Petropoulos, G. (2018): The impact of artificial intelligence on employment, <http://bruegel.org/wp-content/uploads/2018/07/Impact-of-AIPetrouopoulos.pdf>, Online (accessed 26.4.2020)
- Pichai, S. (2018). AI at Google: our principles. Google. Available at: <https://blog.google/topics/ai/ai-principles/> (accessed 05 January 2022).

- Quest, L., Charrie, A., and Roy, S. (2018). THE RISKS AND BENEFITS OF USING AI TO DETECT CRIME. RISK JOURNAL Harvard Business Publishing
- Schrader, D. E., and Ghosh, D. (2018). Proactively Protecting Against the Singularity: Ethical Decision Making in AI. IEEE Computer and Reliability Societies
- Shahriari, K. and Shahriari, M. (2017). IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems, In Humanitarian Technology Conference (IHTC), IEEE Canada International (pp. 197-201). IEEE
- Statistics South Africa. (2019). Quarterly Employment Statistics. Statistical Release - P0277, December, 2019
- Taddeo M. And Floridi, L. (2018). How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control. ResearchGate VOL 361 ISSUE 6404
- Teng, C. H. (2020). How to Approach Ethical AI Implementation? A practical guide to implementing ethical Artificial Intelligence (AI), *AIBotics*
- Thomaz, F., Efremova, N., Mazzi, F., Clark, G., MacDonald, E., Hadi, R., Bell, J. and Stephen, A. (2021). Ethics for AI in Business. Oxford Future of Marketing Initiative. Available at SSRN: <https://ssrn.com/abstract=3871867> or <http://dx.doi.org/10.2139/ssrn.3871867>
- Trocin, C.; Mikalef, P.; Papamitsiou, Z.; & Conboy, K. (2021). Responsible AI for Digital Health: a Synthesis and a Research Agenda. Information Systems Frontiers <https://doi.org/10.1007/s10796-021-10146-4>
- Walsh, T., Levy, N., Bell, G., Elliott, A., Maclaurin, J., Mareels, I., & Wood, F. (2019). The Effective and ethical development of Artificial Intelligence (p. 250). ACOLA. [https://acola.org/wp-content/uploads/2019/07/hs4\\_artificial-intelligence-report.pdf](https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf)