

**ARTIFICIAL INTELLIGENCE BIAS AND NEURAL
MACHINE TRANSLATION: TRANSLATING
HEAVILY_LOADED IDEOLOGICAL
ENGLISH→ARABIC MESSAGES**

Dr. Safa'a A. Ahmed

Associate Professor of Interpreting and Translation
Faculty of Languages, MSA University, Egypt

Abstract

Bias, "an inclination, prejudice, preference or tendency towards or against a person, group, thing, idea or belief" (Murphy 2021), raises ethical questions whether in human or machine communication and it can have detrimental impacts on individuals and societies, e.g. criminal judgments. The deployment of AI systems in real-world settings has exposed various kinds of bias against certain social groups. The big number of research on bias in AI applications generally and the few studies on bias in neural machine translation (NMT) particularly have given the present study the momentum to further investigate the issue. It aims to examine bias in NMT through exploring the translation of some heavily-loaded ideological messages from English into Arabic. A multidisciplinary perspective deriving its tenets from translation studies, political sciences and computer science is utilized to explore bias in the translation of ideology by Google Translate API. It employs a qualitative approach using analysis, comparison and deduction as tools of research. It has reached some interesting findings, which came contrary to my initial expectations.

Keywords: Translation; Neural Machine Translation; Bias; Artificial Intelligence; Ideology in Translation

تحيز الذكاء الاصطناعي والترجمة الآلية العصبية:

ترجمة الرسائل المحملة بالأيديولوجيا بكثافة من الإنجليزية إلى العربية

أ.م.د. صفاء أحمد

أستاذ مشارك الترجمة الشفهية والتحريرية

كلية اللغات- جامعة أكتوبر للعلوم الحديثة والآداب

الملخص

التحيز، "الميل أو التحيز أو التفضيل أو الميل تجاه أو ضد شخص أو مجموعة أو شيء أو فكرة أو معتقد" (Murphy 2021)، يثير أسئلة أخلاقية سواء في التواصل البشري أو الآلي؛ ويمكن أن يكون له آثار ضارة على الأفراد والمجتمعات، كالأحكام الجنائية على سبيل المثال. ولعل انتشار أنظمة الذكاء الاصطناعي في بيئات العالم الحقيقي، لعله يكشف عن أنواع مختلفة من التحيز ضد فئات اجتماعية معينة. إن العدد الكبير من الأبحاث حول التحيز في تطبيقات الذكاء الاصطناعي عمومًا والدراسات القليلة حول التحيز في الترجمة الآلية العصبية (NMT) بشكل خاص، دفع هذه الدراسة إلى مواصلة البحث في هذه المشكلة. من ثم تهدف هذه الدراسة إلى البحث في التحيز في الترجمة الآلية العصبية من خلال فحص وتقييم ترجمة بعض الرسائل المحملة بالأيديولوجيا بكثافة من اللغة الإنجليزية إلى اللغة العربية. استخدم فيها منظور متعدد العلوم يستمد مبادئه من دراسات الترجمة والعلوم السياسية وعلم الكمبيوتر لاستكشاف موضوع التحيز في ترجمة الأيديولوجية بواسطة برنامج جوجل للترجمة Google Translate. وتستخدم لذلك منهجًا نوعيًا يعتمد على التحليل والمقارنة والاستنباط كأدوات للبحث. توصلت الدراسة إلى بعض النتائج المثيرة للاهتمام، لا سيما وأنها قد جاءت مخالفة لتوقعاتي الأولية عنها.

الكلمات المفتاحية: الترجمة؛ الترجمة الآلية العصبية؛ التحيز؛ الذكاء الاصطناعي؛ الأيديولوجيا في الترجمة

**ARTIFICIAL INTELLIGENCE BIAS AND NEURAL
MACHINE TRANSLATION: TRANSLATING
HEAVILY_LOADED IDEOLOGICAL
ENGLISH→ARABIC MESSAGES**

Dr. Safa'a A. Ahmed

Associate Professor of Interpreting and Translation
Faculty of Languages, MSA University, Egypt

1. INTRODUCTION

In 2015, Jacky Alciné, a black software developer, was taken aback when he saw the selfies he had had with a friend labelled by Google Photos as a 'gorillas' default folder; in 2017 Joy Buolamwini, a Ghanaian-American computer scientist working at MIT Media Lab was facially recognizable only when she put on a white mask and then let into the Lab (Bias in Artificial Intelligence 2021:1). These are only examples for bias in image recognition applications. Bias largely extends over artificial intelligence (AI) software.

Bias is "an inclination, prejudice, preference or tendency towards or against a person, group, thing, idea or belief" (Murphy 2021). It is a divergence from an ideal or a set/expected value (Shah et al. 2020; Glymour& Herington 2019). Bias is as old as the history of human beings. Therefore, machines fed with biased data or algorithm (a finite sequence of clear, well-defined instructions given to the machine to perform a specific task) by humans can manifest unfair, biased, decisions. AI-based systems "are widely employed nowadays to make decisions that have far-reaching impact on individuals and society" (Ntoutsis et.al. 2020:1). AI has invaded various fields from spam filtering, management of logistics, speech and image recognition, media collecting, translation, human resources software, planning traffic, banking accounts and transactions, to diagnosing diseases and surgical operations, to name but a few (Boegesius 2018:9). The European Union Agency for Fundamental Rights admits that incomplete or biased data in AI systems "can lead to inaccurate outcomes that infringe on people's fundamental rights, including discrimination" (Data Quality 2019:1). Be it conscious or unconscious, the most common bias types include racial, gender, religious, sexual orientation, educational background or socio-economic bias. It raises ethical questions whether in human or machine communication and can have detrimental impacts on individuals and societies e.g. criminal judgments. Jones et.al. (2020) say that AI

applications that can influence human decisions and lives raise ethical questions about fairness (as cited in Pagano 2022:1).

Murphy (2021) argues that biases "are often based on stereotypes, rather than knowledge or experience" and that it is "usually learned". So it may become common sense to conclude that if humans are biased, machines are going to be biased too since they learn from humans' data-sets. For instance, mass media may rely on biased AI systems to facilitate voting tasks, but the "systematic exposure to such bias may foster intolerance as well as ideological segregation" (Baly et.al. 2020:4982).

The history of machine translation (MT) has gone through three phases: early machine translation dictionaries in the mid-1950s, corpus-based computer-aided translation tools developed in the 1980s, and recently_ since 2016_ neural machine translation (NMT) with AI in the core. This study aims to examine bias in NMT through exploring the translation of some heavily-loaded ideological messages from English into Arabic. To this end, a multidisciplinary perspective deriving its tenets from translation studies, political sciences and computer science is utilized to explore bias in the translation of ideology by Google Translate (GT) API. It uses a qualitative approach employing analysis, comparison and deduction as tools of research.

The main contribution of the present paper lies in its attempt to shed the light on such issues related to machine bias: the global trend towards the inclusion of AI in various fields, the machine bias discovered in many applications, and necessity for paying attention to ideological translation (see for example Ahmed 2014, 2016a &b, 2019a &b) and how it can change or shape the identity of nations. Also it attempts to contribute to the rare studies on racial and religious bias in MT and to address gender bias from a new perspective. Moreover, it challenges the issue from a multi-disciplinary perspective which will hopefully add some value to the field of translation (especially as millions of people around the world use MT daily) taking political and computer sciences into consideration. However, I am aware of the limitations of the study, firstly data size (one speech only is used), secondly the space restrictions of the paper, thirdly the language combination (English/Arabic), and fourthly the direction of the translation (from English into Arabic). More investigations of different data samples or corpora and similar and other language combinations would definitely benefit the discipline and the profession.

Finally, the paper is divided into three sections: Related Works, Theoretical and Methodological Frameworks, and Data Discussion, in addition to this introduction and the conclusion.

2. RELATED WORKS

The UK Commission for Racial Equality found in 1988 that a British medical school was guilty of racial (against non-Europeans) and gender bias because the software system used was discriminatory, selecting certain applicants to interview; this incident attracted attention to both human and machine bias (Silberg and Manyika 2019:1).

A negative attitude, not based on facts, towards a certain person or group of persons or an idea is prejudice. When prejudice is interpreted into behaviour or action, it becomes discrimination. Prejudice can result in bias, an inclination towards or against a particular person, social group, idea, etc. Stereotyping, an unscientific over-generalized image or inference toward or against people or ideas, cultivates prejudice.

Many attempts have been made to categorize human bias especially in cognitive sociology in the 1980s (cf. Allport 1979). Some scholars divide bias into conscious and unconscious; under each, tens of types can be found. Yet, the categorisation approach is criticised because it assumes the inevitability of bias; it ignores the idea of tolerance and results in a mechanical mode of recognition (Billig 1985:79). Therefore, Billig suggests a rhetorical approach combining between categorisation and particularisation (ibid.). Actually, the augment of categorisation falls outside the scope of the present paper and it will be sufficient to mention that bias can be based on race, gender, religion, sex, age, or social, economic or educational background, etc. (Nelson 2019).

In machines, bias exists too. Boegesius demonstrates that machines "may appear to be rational, neutral and unbiased but, unfortunately, AI and algorithmic decision-making can also lead to unfair and illegal discrimination" (2018:9). The following sub-sections deal with related works on bias in AI and then on bias in MT.

2.1 Bias in AI

Kaul explains that machine bias results from human bias because AI systems are built by humans who can be prejudiced towards some people. Her study explores human biases evident in AI systems and she suggests few solutions to make those systems more objective and less subjective (2021:9887).

The existence of bias in machine learning (ML) systems was studied by many scholars. Farkas and Németh (2021:1) mention that many cases where algorithms result in race, gender, religion and age bias and consequently decisions against some social groups, have been discovered in the last few years, e.g. gender bias in hiring algorithms (Chen, Hannak and Wilson 2018; Schwarm 2018; Dastin 2018), ageist and racist advertisement (Angwin et al. 2017; Barocas and Selbst 2016), and

regional dialects bias in natural language processing (NLP) training corpus (Jurgens, Tsvetkov and Jurafsky 2017). Hovy and Prabhunoye (2021) delineate five sources for bias in NLP: data, models, research design, annotation process and input representations and they suggest some possible counter actions, as we shall see later.

AI can intensify the biases and racial inequalities in society, where biased systems produce inaccurate predictions and harmful decisions against certain social groups; such AI applications "have already affected public systems such as criminal justice, healthcare, financial systems and housing" (Livingston2020:1). Livingston accounts for the primary causes for developing, deploying and using racially biased systems and proposes actions to improve them, like increasing racial diversity among software designers and impact assessment. Zajko says that AI has increased the dangers of intensified social inequalities especially with the increasing number of government agencies which adopt these systems (2022:1). He suggests that sociologists should have a valuable role since AI applications "will continue to become even more central to the reproduction of inequality in the future" (p.11).

Fuchs (2018) studies the effects of human-like bias in social media chatbots, image recognition tools and criminal justice software. He used three datasets from Digital Library, IEEE Xplore and Scopus. He reaches the conclusion that social media chatbots, by nature, are often exposed to discriminatory input and lead the machine to learn dangerous 'human-like biases' (p.4). Similar learned bias towards dark skin and females was noted in face recognition and criminal justice applications showing 'harmful' effects on human lives. He calls for improving both the accuracy and fairness of these systems (p.12). The Centre for Data Ethics and Innovation reviews the impacts of bias on decision-making, risk management, and the opportunities to enhance fairness, in recruitment, financial services, policing and local government (Review into Bias 2020:6); it concludes that using a fairer, more trusted data is a challenge and it claims that UK can help countries since it is a lead in this concern. Garrido-Muñoz et.al. investigate how bias has been detected and corrected in many networks and proposed a strategy to handle it (2021:1).

AI needs to incorporate ethical issues in its various phases. Osoba and Welser (2017:1) refer to the risks associated with uncritical reliance on AI "especially when they implicitly or explicitly mediate access to services and opportunities" (e.g. financial services, employment, etc.). They add that algorithmic decisions are not automatically fair just because of the nature of their complex processes (ibid.). In a survey of AI challenges and

solutions, Ntoutsu et.al. (2020) assure that AI design, training and deployment should embed ethics and legal principles for the sake of social good. Sengupta and Srivastava state that racially-biased training datasets lead to poor model explainability, affect user experience and decisions, increase biases, and raise many ethical questions (2022:1). Blodget et. al. (2021) indicate that ML systems can develop various problems due to unethical decisions in different contexts. Liu et.al. (2022) say that language models can be politically biased due to the training dataset and can cause critical problems in real-world settings; they present two metrics in their study to measure bias and suggest a solution to mitigate bias without using training datasets or any extra data. Eliminating bias from machine systems improves the decision-making process (Haenlein and Kaplan 2019).

Tomalin et.al. (2021:1) talk about three strategies to debias NMT systems: downsampling, upsampling, and counterfactual augmentation; but findings show a translation performance far worse than any baseline system and a worse degree in gender bias. They conclude that the technique of domain adaptation can debias well-trained systems. Silberg and Manyika (2019:1) propose some pragmatic ways to remove human and machine bias; however, they admit that it is still early for "promises of quick fixes or silver-bullet solutions". Mitchell et.al. (2019) argue that identifying and mitigating bias and unfairness are not easy tasks because cultures define unfairness differently. Therefore, the user's experience and background influence unfairness criterion (as cited in Pagano et.al.2022:1). Reviewing systematically 40 eligible articles published in 2017 through 2022, Pagano et.al. found out that the development approach to mitigate bias is left entirely to the developer, who is not necessarily aware of the problem, and that despite the recommended solutions, further research is needed (ibid.).

2.2 Bias in MT

A decade ago, Chomsky (2011) was not optimistic about MT achievements and he mentioned that it only shows methods for how to approach data which is unanalyzed yet. Nevertheless, MT has improved particularly in the last few years, but not without criticism, like bias.

A MT model becomes biased when its behaviours cause problems upon deployment (Bender et. al. 2021). First account on gender bias in MT dates back to Frank et al. (2004). They studied linguistic incompetence in English-German translation where the machine could not often recover syntactic and semantic information to translate gender properly. Schiebinger (2014) goes one step further beyond linguistic

analyses to understanding gender bias. Her article on Google Translate's 'masculine default behavior' drew attention to the system inequalities.

Many scholars have studied the issue of gender bias from various perspectives. Hovy et al. (2020), for instance, explore how speakers' gender and age are conveyed in MT. Because models are under-exposed to women and children writings; their findings show that translation systems tend to make authors sound male and older. Also, Kurpicz-Briki (2020:1) found gender and origin biases in German and French word embeddings. They state that real-world bias and stereotypes borrowed from the 18th century are present in current word embeddings, and they suggest that gender bias differs from one culture to another.

Interesting studies reveal unexpected results. Prates et.al. (2020) list job positions according to the US Bureau of Labor Statistics and register sentences in constructions like "He/She is a Journalist", where this job is replaced by another one in twelve different gender neutral languages like Chinese. The sentences were translated into English using GT. They calculated the frequency of female, male and gender-neutral pronouns in the MT output. GT was discovered to yield male defaults more frequently than female, even if taken from a demographic distribution perspective.

Gender bias seems to be appealing to recent research on NMT bias. Savoldi et.al. (2021: 845) maintain to present 'the first comprehensive review' of this issue; they summarize previous studies, discuss some mitigating strategies, and address potential directions in the future. Hence comes their categorisation of bias into *under-representation* (low representation of women through male references where female should be the right translation, or through using male references for neutral forms in neutral gender languages) and *stereotyping* (negative generalisations of social groups, for example, assigning less prestigious occupations like (teacher) to females vs. (lecturer) to males) (p.846). Měchura discusses bias in translations due to ambiguities in gender, forms of address and word meaning (2022:2). Scholars agree that gender bias is quite common in automated translation (e.g. Escudé-Font and Costa-jussà 2019; Costa-jussà and de Jorge 2020; Saunders and Byrne 2020; Stafanovičs, Toms and Pinnis 2020).

NMT exposes 'problematic biases', according to Wang et.al. (2022:2576). They uncover a new bias in translating persons' names and the necessity of identifying the gender of these nouns. They argue that the most famous software tools cannot handle this point appropriately, and that proper names can influence the translation of words with ambiguous sentiment and of proper nouns alluding to race. They propose a data

augmentation method to debias the system. The authors refer to other biases like racial, professional, and individual biases as well, but they focus on two types, gender and sentiment (see also Kiritchenko and Mohammad 2018 for sentiment analysis), and the resulting serious distress to users and reputational or legal consequences for vendors, a matter which may cause a 'widespread offence' (p.2576).

Although Google, for instance, has repeatedly attempted to remove gender bias from its NMT system, there are still "predominantly masculine options" and "perpetuating or exaggerating (of) sexist stereotypes", Medel (2021:1) states. Training software to infer semantic gender from pronouns and word terminations, and gender-tagging could help (p.6). Similarly, Farkas and Nemeth (2021) made a case study of gender bias in GT where it translated sentences with names of occupations from Hungarian (a language system with gender-neutral pronouns) into English. They wanted to have a fair measure for bias by comparing the product to a product of a non-biased software. They discovered cases of bias against both genders generally and females particularly (p.1).

From this review of related works, it becomes clear that previous literature on machine bias has often focused on bias in AI with a special stress on racial and gender bias and has rarely focused on NMT bias despite some references to gender bias in MT.

3. THEORETICAL & METHODOLOGICAL FRAMEWORKS

3.1 Theoretical Framework

Grant, Gode and Amstrong argue that traditional translation theories have focused on comparing the source text to target text, using faithfulness and fidelity as assessment criteria, yet ideology has an 'invisible hand' and should be taken into account (2017:2).

The term 'ideology' was coined by the French philosopher Destutt de Tracy in 1795 to define ideas and improve public debate. It is an ambiguous term but Tracy regards it as the science of the formation of ideas. For simplicity, it denotes a set of dominant ideas influencing all spheres of human life (Zajda 2014:2).

Translation itself is sometimes looked at as a political act, where ideology comes in its core. Tahir-Gurcaglar considers translation a political act, be it a process or product (2013:113). Hatim and Mason say that ideology is a set of ideas and beliefs of an individual, group of people, etc. which takes the form of language (1997:218); hence comes the role of ideology in translation. Pérez points out that it is difficult to explain the ideological differences between the source and target texts

whether they are attributed to the translator's intentional mediation or unconscious mediation (2003).

When a translator approaches an ideological source text, he either transfers the message as it is or '*mediates*' giving ideology another dimension depending on the extent of mediation. Hatim and Mason define mediation as the amount of the translator's intervention in the process of translation where he feeds his beliefs and knowledge (1997:147).

In international relations, states do not always manage their relations through co-operation, good intentions, international law as idealist political scholars think. History and practice_ of wars and conflicts_ tell us that they are sometimes run via power and conflicts, as realists suggest. Sovereign states have their own interests which can clash/ conflict with other countries' and they have to protect them by any and all means (diplomatically, politically, economically, militarily, etc.). Realism is the most dominant theory of international relations since WWII. It has passed through three phases: *classical realism* (based on the belief that people are selfish and aggressive by nature and so are states), *neo-realism* (the international system needs a sovereign power to enforce international law, therefore states do not trust each other and they act the way they like to maximize their power and security (Wolff 2015:1,9)), and *neo-classical realism* (states act according to the opportunities and constraints of the international system, so their co-operation or conflict depends on the situation). Most criticisms of the Realism theory come from liberals, e.g. Robert Jervis, for instance, claims that the Realism theory will fail to account for conflict or cooperation in the next years (1982; as cited in *ibid.*). First, the theory is criticised for being unable to explain co-operative and peaceful relations as in the end of the Cold War; this point can be easily refuted in the light of the increased number of wars worldwide particularly in the last three decades. Bush announced that a New World Order era was being formed with one dominant power (USA) in 1990 in the aftermath of the beginning of the collapse of ex-USSR; in fact, a one-polar, imbalanced system as such will never attain or maintain international peace and security. Second, the theory considers states as actors and disregards institutions like the UN or NGOs. It seems that those institutions can further enhance international conflicts, for example the UN is recently used as a platform for USA and the NATO in the Russian-Ukrainian conflict to promote war against Russia. Third, it is criticised since there was no 'major war' until some time ago. The world appears to be currently on the edge of a Third World War between

Ukraine, USA and the NATO on the one hand and Russia and its allies on the other. Actually, the Realism theory, though grim as it may look, seems to explain international relations proper.

Ahmed (2014) argues that a multidisciplinary perspective developed from various theories and disciplines adds a great value to approaching and analysing the complexity entailed in the translation of ideology. In this context, a perspective deriving its underpinnings from cultural translation studies (cf. Hatim and Mason 1997; Ahmed 2014), political sciences (to analyse ideology and the political contexts of texts), and computer science (borrowing concepts like AI, NMT, etc.) would be paramount in evaluating bias in the NMT product.

3.2 Method of Research

The deployment of AI systems in real-world settings has exposed various kinds of bias against certain social groups. The big number of research papers on bias in AI applications in general and the few studies on bias in NMT in particular has given the present study the momentum to further investigate bias in NMT. It aims to explore bias in the translation of ideology (in heavily-loaded ideological messages) by Google Translate (GT) API from English into Arabic. It raises four Research Questions (QRs):

RQ#1: How does NMT translate English→Arabic messages heavily loaded with ideological racial bias?

RQ#2: How does NMT translate English→Arabic messages heavily loaded with ideological gender bias?

RQ#3: How does NMT translate English→Arabic messages heavily loaded with ideological religious bias?

RQ#4: How far is the current form of GT biased?

In order to answer these questions, the study sets the following objectives:

- To collect English sample data which is heavily loaded with ideological messages;
- To translate them via NMT, precisely GT;
- To analyse bias (racial, gender or religious) in the source and target texts; and
- To compare the two texts in order to assess GT's accuracy in transferring biases and to assess GT's bias in itself.

In so doing, the study adopts a multidisciplinary perspective based on cultural translation studies, political sciences and computer science in order to explore some original English heavy-loaded ideological messages and assess their Arabic machine translation. It uses a qualitative

methodology whose research tools are mainly analysis, comparison and deduction.

The sample data is carefully selected from a variety of heavily-loaded, ideological messages/texts in English. Certain sensitive topics evoke ideological problems when translated by humans between two different languages and cultures (cf. Ahmed 2019a & b, 2016a&b, 2014). Rabie calls such texts 'civilisational' because they can change and shape the nation's identity; in other words these texts are important so long as they can articulate people's beliefs and ideas about some critical issues (1986:247; as cited in Abdel-Fattah 2006:35-36; trans.). I would like to call them 'ideological texts' instead, since they are loaded with the source speaker/writer's ideology, i.e. ideas and beliefs. Therefore, data is extracted from Barak Obama's Remarks on the Middle East and North Africa on 19 May 2011 (source text) and its Arabic NMT (target text). Google Translate is chosen as a MT tool in the present study for its availability (free), popularity and quality. Google launched the application in 2006 to translate texts or websites from one language into another. In 2016, it made use of developments in neural networks and switched to Google Neural Machine Translation. Now it supports 133 languages and over 500 million users daily.

Only three types of bias (racial, gender and religious) are examined because they represent crucial, common biases in societies and because of the time and space limitations of the research. From the three types, the division of the discussion section was inspired.

4. DATA DISCUSSION

As mentioned before, machine translation (MT) has gone through three phases, the latest is neural machine translation (NMT). The birth of NMT dates back to Nal Kalchbrenner and Phil Blunsom who proposed an end-to-end MT encompassing early developments in neural networks in 2013. Google introduced NMT in 2016, based on natural language understanding (NLU), natural language processing (NLP), MT, translation memory, statistics-based MT and deep learning. Better computer and mobile capabilities, increased training data, and advanced machine deep learning have made outstanding breakthroughs in the quality of the service provided (Zong 2018:4).

The following sub-sections discuss Google's NMT of racial, gender and religious biases in heavily-loaded ideological messages. The sample data is extracted from President Obama's speech Remarks on the Middle East and North Africa on 19 May 2011. That was a time of turmoil in the

region. Tunisians started revolting against their president Zine El-Abidine bin Aly, who ruled for 27 years; Egyptians demanded Hosni Mubarak to step down; Ali Abdullah Saleh was at risk too in Yemen; Bashar al-Assad was in no man's situation with violence erupting in Syria; the Bahraini regime managed to stop protestors; the UN imposed a no-fly zone over Libya, etc. The speech delineates a US policy towards what it called 'the Arab Spring' revolutions, based on its famous concept '*creative chaos*', which it had been promoting openly since 2005 by former US Secretary of State Condoleezza Rice. 'Creative chaos' is an ideological term referring to a political state where some unknown or known people make a deliberate chaos to change the situation for their interest. This ideology justifies the US unlimited support for chaos in the Middle East; though change 'will not be easy' and 'will take time', USA will do anything to this end as "It will require the broad engagement of America, Europe and all free nations", Rice mentions (2003).

4.1 NMT & Racial Bias

Bias is clear in describing the going turmoil in the region as 'Arab Spring' revolutions from the beginning of Obama speech. He describes the change as 'extraordinary' (غير عادي) and for the sake of 'basic human rights' (للمطالبة بحقوقهم الأساسية) from which these countries are deprived, according to Obama. That's why they revolted, for a noble cause:

[1] "...we have witnessed **an extraordinary change** taking place...
the people have risen up to demand their basic human rights."

NMT: "شهدنا تغييرًا غير عادي يحدث في منطقة الشرق الأوسط وشمال إفريقيا... انتفض الناس للمطالبة بحقوقهم الأساسية."

Google's NMT has managed to translate the change and to show the supposed aspiration of the people to their rights through the choice of words. The target text reflects the original message's ideology accurately. The idea is further strengthened through using expressions like 'story of self-determination' (قصة تقرير المصير) to make the change appealing to these countries, as in example [2]:

[2] "That **story of self-determination** began six months ago...It's the same kind of **humiliation** that takes place every day in many parts of the world– the relentless tyranny of governments that deny their citizens **dignity... a longing for freedom ... the frustration felt throughout the country... but in too many places their people did not** (win their independence)."

NMT: "بدأت قصة تقرير المصير تلك قبل ستة أشهر... إنه نفس النوع من الإذلال الذي يحدث كل يوم في أجزاء كثيرة من العالم - الاستبداد الذي لا يلين للحكومات التي تنكر

كرامة مواطنيها...توق إلى الحرية... الإحباط الذي يشعر به جميع أنحاء البلاد... لكن شعوبها لم تنال استقلالها في العديد من الأماكن."

Obama tells us about a new chapter in a story following chapters of 'humiliation' (الإذلال), citing Bouazizi's story which ignited the events in Tunisia on 17 December 2010. The contrast between the two situations, 'self-determination' and 'humiliation', is vivid in the target message too, a matter which makes Google translation sound accurate in conveying Obama's belief or ideology. The choice of the word 'dignity' (كرامة) is sensitive in provoking the Arab peoples' anger and hatred feelings against their governments. The same applies to the rest of the grim image of the situation in these countries, where people feel 'frustration' (الإحباط) everywhere. Obama, as a representative of a colonising power accused of occupying other countries (Vietnam, Afghanistan, Iraq, Syria, etc.), is talking ironically about the Arab countries superficial 'independence' but in essence 'people did not (win independence)' (لم تنال استقلالها), it argues_ there is a minor grammatical mistake in (تنال) for it should be (تتل) instead. The NTM transferred the speaker's ideas accurately.

Stereotyping, positive or negative, can result in bias through delineating an incorrect image about an individual, a social group, an institution, etc. The US policies, orientations and attitudes have helped stereotype an image of the Arab counties as weak, poor (with the exception of oil countries; but no development strategy depends on what is extracted from the ground alone as Obama explains), undemocratic, developing or underdeveloped, and unfree. This is a racial bias against the Arabs. The stress at the beginning of the speech is on presenting a negative image about people and governments in the Middle East and North Africa in order to pave the way to and accept change, keeping in mind the emotional nature of the Arab people. In example [3], he says that in Tunisia as in most of these countries:

[3] "a **citizen** like that young vendor **had nowhere to turn – no honest judiciary to hear his case; no independent media to give him voice; no credible political party to represent his views; no free and fair election where he could choose his leader.**"

NMT: "لم يكن لدى مواطن مثل هذا البائع الشاب مكان يلجأ إليه - لا يوجد قضاء نزيه - لسماع قضيته ؛ لا توجد وسائل إعلام مستقلة لإعطائه صوتاً ؛ لا يوجد حزب سياسي موثوق به لتمثيل آرائه ؛ لا انتخابات حرة ونزيهة حيث يمكنه اختيار زعيمه."

The human being, or 'citizen' since citizenship requires rights, comes at the core of events in the delineated image. The helpless citizen has 'nowhere to turn (لم يكن لدى مواطن..مكان يلجأ إليه)', 'no honest judiciary' (لا وسائل إعلام مستقلة)', 'no credible (يوجد قضاء نزيه)

political parties' (لا يوجد حزب سياسي موثوق به), 'no free, fair election' (انتخابات حرة ونزيهة), what else he has not? Google Translate conveys the original message again accurately.

Bias is deeply rooted in the US attempt to legitimise the war in Iraq in response to the events of September 11, 2001, when President Bush divided the world into either 'friends' or 'enemies'/ 'foes'. In Example [4], Obama speaks about US principles, another term for ideologies, of 'friend' (الأصدقاء) or 'foe' (الأعداء):

[4] "we need to speak honestly about **the principles that we believe in, with friend and foe alike**"

NMT: "نحتاج إلى التحدث بصدق عن المبادئ التي نؤمن بها، مع الأصدقاء والأعداء على حد سواء"

Why should the countries which adopt the US ideologies, 'the principles that we believe in' (المبادئ التي نؤمن بها), be friends and others foes? Who said that the US model (political and economic reform, modernisation, cultural identity, among others) is the right one for all countries by default? Why ignore each country's own principles, values, identity, etc.? The translation has succeeded in transferring the original message with its original, deeply-rooted bias in this example. Consequently, USA is keen to 'advance our (its) values' and support 'friends' in that 'historic opportunity' and 'this moment of promise'. In the next example:

[5] "we will keep our commitments to **friends and partners...** Our message is simple: **If you take the risks that reform entails, you will have the full support of the United States.**"

NMT: "وسنفي بالتزاماتنا تجاه الأصدقاء والشركاء... رسالتنا بسيطة: إذا جازفت بالمخاطر التي ينطوي عليها الإصلاح، فستحصل على الدعم الكامل من الولايات المتحدة."

USA offers its support as it 'will keep our commitments' (وسنفي بالتزاماتنا) to 'friends and partners' (الأصدقاء والشركاء); this an accurate Google translation. Obama states the message very simply then: if these countries implement the US vision, mission and agenda in the region and accept to 'take the risks that reform entails' (إذا جازفت بالمخاطر التي ينطوي عليها الإصلاح), or more precisely chaos, they have 'full support' (الدعم الكامل). The Arabic text conveys exactly the intended message in an excellent syntactic and semantic language reflecting the same ideology of the source. Funny enough, Obama admits that "for years the United States pursues our interests at their (Arab peoples) expense"_ what will be added to the new strategy is the change to 'self-determination of individuals'. The change which the States had supported has resulted in chaos, toppling some Arab presidents and helped other sects reach power. The US change could not and will never stand in front of the peoples' will and refusal of those imposed regimes, and that's literally

what happened in Egypt and the other countries alike. The US change could not achieve peace or security.

According to the United Nations Charter, Article 2/7, nothing authorizes the UN to intervene in matters which lie within the domestic jurisdiction of member states. Yet the last part of the Article ruins this basic right by the UN Security Council enforcement measures which confer to the Council the 'primary', unlimited authority in maintaining international peace and security. Any sovereign country should choose its own path to stability and security within the boundaries of international law and agreements. All what USA has been doing are blatant cases of intervention in the Arab countries' domestic jurisdiction. USA insists on intervention, though Obama admits that experience in Iraq taught them that 'imposing' change by 'force' is absolutely costly and difficult:

[6] "the United States joined **an international coalition to intervene...** and we have learned from our experience in Iraq just how costly and difficult it is to try to **impose regime change by force**"

NMT: انضمت الولايات المتحدة إلى تحالف دولي للتدخل.. وقد تعلمنا من تجربتنا في العراق كم هو مكلف وصعب محاولة فرض تغيير النظام من خلال القوة"

USA has been leading 'an international coalition to intervene' (تحالف دولي) (من خلال 'by force' (فرض تغيير النظام) to 'impose regime change' (للتدخل) although the opposite is explicitly stated but the message may imply force_ in these countries, imagining that the 'international coalition' would give them legitimacy. It is true that most colonialism had ended long time ago in the classical sense of the word, but a new form of neo-colonialism has started ever since, using soft power at times and force at others. Take for example Syria, if it does not abide by the American principles/ideologies, it will be defied from 'within':

[7] "Otherwise, President Assad and his regime will continue to be challenged from **within** and will continue to be **isolated abroad**."

NMT: خلاف ذلك، سيستمر الرئيس الأسد ونظامه في مواجهة التحديات من الداخل وسيظل معزولاً في الخارج."

The word 'from within' (من الداخل) shows the intervention in domestic jurisdiction of states. Also international isolation, 'isolated abroad' (معزولاً) (في الخارج), gives us an example of how far USA is ready to go to impose its ideologies on the Middle East and North Africa. The alliance of 'friends' against others in such cases violates states' sovereignty and

rights, and international law alike. All these implications can be easily detected from the NMT too.

The idea of dividing people into different ethnicities and sects exemplifies by itself a core reason for bias; it is an extension of the coloniser's rule of thumb 'Divide and Rule'. Over about 1400 years, the Arab countries with all their sects and ethnicities lived together without asking to which sect or ethnic group anyone belongs. When such issues were provoked by the coloniser, only then conflicts and wars emerged, e.g. Muslims and Shiites in Lebanon, Iraq, Yemen, etc. Another example is Syria, which stands for a 'mosaic' of sects and races, so one can easily imagine what divisions can do when such issues are raised.

Obama in example [8] talks about a change to 'promising' democracy based on 'multiethnic' and 'multi-sectarian' division concept:

[8] "In Iraq, we see the promise of a **multiethnic, multi-sectarian democracy.**"

NMT: "في العراق، نرى الوعد بديمقراطية متعددة الأعراق والطوائف."

The trap of having representatives from different ethnic and racial groups in governments destroys any progress in taking decisions on any issue and thus causing paralysis of such multi-ethnic and multi-sectarian governments_ Iraqi, Lebanese and Libyan governments are brilliant illustrations in this regard. The translation of 'multiethnic' as (متعددة الأعراق) and 'multi-sectarian' as (و(متعددة) الطوائف) transfers the message precisely. These beliefs have been part of the US National Security Strategy for decades, despite the announced claims that "we want greater freedom and opportunity for the people of the region" as Rice (2003) maintains.

The plan to change the region depends on a number of tools. Obama makes a clear connection between young people, technology, education and information. The US administration would go beyond elites and try to reach 'young people' (الشباب) 'who will shape the future' (الذين سيشكلون) (المستقبل):

[10] "so that **we reach the people who will shape the future** – particularly **young people.** .. expand exchanges in **education...** use the **technology** to connect with – and listen to – the voices of the people."

NMT: "حتى نصل إلى الأشخاص الذين سيشكلون المستقبل - وخاصة الشباب. توسيع التبادلات في التعليم.. وسنستخدم التكنولوجيا للتواصل مع أصوات الناس والاستماع إليها."

He is well aware of the importance of the using young people as a soft power to reach change; young people are enthusiastic, more open to new, particularly US and Western, ideas, and ready to accept technological

updates. Therefore, it is not a surprise that one of the revolution leaders in Egypt in 2011 was a Google executive:

[11] "It's no coincidence that **one of the leaders of Tahrir Square was an executive for Google.**"

NMT: "ليس من قبيل المصادفة أن أحد قادة ميدان التحرير كان مديرًا Google".
"تنفيذيًا لشركة"

Here arises the role of technology_ especially social media_ in the region protests or revolutions. Education as part of the plan cannot be overlooked and it actually raises doubts about the criteria used in selecting those young people to be exchanged to have education in USA and Europe and the content they receive as well. The translation in [10] and [11] are accurate, but there is one remark regarding the English version of 'Google' amid the Arabic translation and it is unexplainable, indeed.

In the 21st century, the ideology 'information is power' (تعد المعلومات قوة) prevails. It relates closely to the concept of 'technology':

[12] "In the 21st century, **information is power**"

NMT: "في القرن الحادي والعشرين ، تعد المعلومات قوة ."

About four decades ago, we used to suffer from the scarcity of information. Now we suffer from the abundance of information. Uncontrolled, wrong and even misleading information is as bad as the scarcity of information. The availability of only 'right' information is power. So the belief 'information is power' needs reconsideration. Google Translate correctly conveys the message.

4.2 NMT & Gender Bias

In 2018, Google attempted to develop specific gender translations for some words, phrases and sentences in few languages, yet masculine default translations continued as feminist scholars and translators argue. Two years later, its NMT has improved its gender accuracy (Medel2021:1-2). Saunders, Sallis and Byrne clarify that NMT finds difficulties in grammatical gender resulting from the human referents gender, a matter which may produce gender biases (2020:1). Saunders and Byrne attribute the gender bias in NMT to the training datasets trained with few sentences referring to women; hence debiasing is a challenge (2020: 7724). In the following part, we will examine how the NMT dealt with ideological messages loaded with heavy gender references.

The US discourse directed to the Middle East and North Africa when it comes to women is obviously one that encourages the 'empowerment' of women:

[13] "History shows that countries are **more prosperous and more peaceful when women are empowered.**"

NMT: "يُظهر التاريخ أن البلدان تصبح أكثر ازدهارًا وسلامًا عندما يتم تمكين المرأة."

The US ideology in this respect states that countries will become 'more prosperous and peaceful' (أكثر ازدهارًا وسلامًا) _ the use of the adjectives is deliberate because what else nations can aspire to rather than prosperity and peace? _ 'when women are empowered' (عندما يتم تمكين المرأة). The translation transfers the original message accurately. To empower women means giving them power and rights. This should be done not from the US perspective, but from the Arab and Muslim countries', for the latter have their own values which can contradict the former's. That's why US principles must not be imposed on other nations.

Some principles, of course, apply to all humans, like:

[14] "What is true for religious minorities is also true when it comes to **the rights of women.**"

NMT: "ما ينطبق على الأقليات الدينية صحيح أيضًا عندما يتعلق الأمر بحقوق المرأة."

In this example, Obama compares the status of the 'rights of women' (حقوق المرأة) in the region to that of religious minorities. He then assures the idea that 'universal rights' (الحقوق العالمية) 'apply to women' (تنطبق على) (النساء والرجال) and men, not men only, see example [15]:

[15] "And that's why we will continue to insist that **universal rights apply to women as well as men.**"

NMT: "أنا سنواصل الإصرار على أن الحقوق العالمية تنطبق على النساء والرجال على حد سواء."

The translation is correct _ with the exception of translating "as well as" into (على حد سواء) which means 'alike'; the Arabic text hence gives an extra meaning of equality not mentioned in this sentence, and it would have been better to delete that addition at the end of the sentence. Normally, since women represent more than half of the region population, it is illogical to ignore their potential and role in societies:

[16] "The region will never reach its **full potential** when **more than half of its population is prevented from achieving their full potential.**"

NMT: "لن تصل المنطقة إلى إمكاناتها الكاملة أبدًا عندما يُمنع أكثر من نصف سكانها من تحقيق إمكاناتهم الكاملة."

The clever choice of words like 'full potential' (إمكاناتها الكاملة) and sentences like 'more than half the population is prevented' يُمنع أكثر من

(نصف سكانها) achieves an appellative function of the original text, i.e. it appeals to the text receiver and urges him to act in a certain way. Therefore, gender bias against females is not present in this discourse, on the contrary it boasts the idea of 'equality' (المساواة) between men and women:

[17] "**equality for men and women... that every man and woman is endowed with certain inalienable rights.**"

NMT: "والمساواة بين الرجل والمرأة... وأن كل رجل وامرأة له حقوق معينة غير قابلة للتصرف"

Obama describes these rights as 'inalienable' (غير قابلة للتصرف). Hence, according to his argument, everyone should not make any concessions in this regard. The NMT transfers these values to Arabic accurately.

From this discussion, it appears that Google Translate transferred the ideological messages of the source text with amazing accuracy and quality. But this time there is no bias (gender) in the original, instead the US discourse emphasizes a strongly-claimed belief regarding equality between men and women on the one hand and the empowerment of women on the other. The neural machine translation conveyed the ideologies, be them biased or not, accurately. Furthermore, I have not noticed any bias in translating gender referents generally and female referents specifically. It is noteworthy to mention that USA, however, is accused of racial and gender bias against black women, and so are many AI systems, as explained in section 2. Related Works.

4.3 NMT & Religious Bias

In the 21st century, in a blatant defiance to international law and agreements, and despite all calls for fundamental human rights and all the claims of the UN sustainable development goals (SDGs), occupation still exists and it is even hailed by USA and its allies, let alone being carried out under the auspices of an international organization like the UN. The Israeli occupation of Palestine is an example where the US bias towards Israel is unlimited. Obama admits that the Arab-Israeli conflict is an 'occupation' (الاحتلال) and a 'humiliation' (إذلال) for the Palestinians:

[18] "For Palestinians, it has meant suffering the **humiliation of occupation, and never living in a nation of their own.**"

NMT: "بالنسبة للفلسطينيين، فقد عنى ذلك معاناة إذلال الاحتلال وعدم العيش في وطن خاص بهم."

As part of a carrot and stick policy, Obama warns Palestinians that they must reach agreement with Israel, no matter if it is unfair, otherwise they will 'never' (عدم) live in a state 'of their own' (خاص بهم). By and large, the

translation is accurate with the exception of the translation of 'nation' as 'وطن' (homeland) instead of 'أمة' (nation); 'homeland' indicates the land while the 'nation' indicates the people.

The bias here is not only political or economic, it is religious in the first place since Jerusalem represents a sacred place (the cradle of the three religions, Islam, Christianity and Judaism) for all the Arabs and Muslims around the world. The bias represents a religious ideology from the very fact that Israel is the only country established on a (Jewish) religion, as stated in [19]:

[19] "Israel as a **Jewish state** and the **homeland for the Jewish people**, and the state of **Palestine** as the homeland for **the Palestinian people**"

NMT: "إسرائيل كدولة يهودية ووطن للشعب اليهودي، ودولة فلسطين كوطن له. الشعب الفلسطيني."

The religious bias in the US ideology is quite explicit here, where Obama refers to Israel as 'a Jewish State' (دولة يهودية) and 'homeland for the Jewish people' (وطن للشعب اليهودي) _ religion is involved _ vs. Palestine as a state or homeland for 'the Palestinian people' (الشعب الفلسطيني), just 'Palestinian', a nationality rather than a 'Muslim' or a 'Christian' state for instance. The neural translation is correct with the exception of adding ' له' (for it.) in (له. الشعب الفلسطيني). The addition of the preposition, the pronoun and the full stop is an unexplainable mistake, if deleted, the rest of the Arabic sentence would sound perfect.

The US bias towards Israel is not a hidden issue. The American administrations announce openly their 'steadfast' relationship with Israel and how they defend Israel's security:

[20] "**standing up for Israel's security**"

NMT: "الدفاع عن أمن إسرائيل"

The choice of the phrasal verb 'stand up', accurately translated into (الدفاع), to show the American orientation toward Israel reflects its too biased position concerning the Arab-Israeli conflict. In the next paragraph this ideology simply emerges as an evident example for bias:

[21] "As for Israel, **our friendship is rooted deeply in a shared history and shared values. Our commitment to Israel's security is unshakeable.** And we will **stand against attempts** to single it out for **criticism** in international forums."

NMT: "بالنسبة لإسرائيل، فإن صداقتنا متجذرة بعمق في تاريخ مشترك وقيم مشتركة. التزامنا بأمن إسرائيل لا يتزعزع. وسنقف ضد محاولات تمييزه بالنقد في المحافل الدولية."

All the bold words in example [21] and their accurate Arabic translations are highly loaded with ideological messages that echoes the religious

bias: 'our friendship' (صداقتنا), 'rooted deeply' (متجذرة بعمق), 'shared history' (التزامنا), 'shared values' (قيم مشتركة), 'our commitment' (التزامنا), 'Israel's security' (أمن إسرائيل), 'unshakable' (لا يتزعزع), 'stand against' (سنقف ضد), 'attempts' (محاولات), and 'criticism' (النقد). However, the translation of 'distinguishes it with criticism' (تميزه بالنقد) fails to convey the message smoothly for structural and semantic reasons; it would have been better if translated into 'لإقصائه عن طريق'. Indeed, USA is too biased to become a fair mediator in the Middle East peace process. This justifies the Palestinian losses in land, humans, infrastructure, etc. since the launch of the Madrid Peace Process in 1993; while Israel has been expanding its settlements and occupying new territories.

Now moving to the most sensitive topics in the conflict: Jerusalem and refugees. Obama describes the 'future of Jerusalem' (مستقبل القدس) and the return of Palestinian refugees to Palestine as 'wrenching and emotional' (مؤلمتين وعاطفيتين) because the first issue represents a religious ideology for Muslims and the second represents every refugee's fundamental human right to return back to his own country whenever he so desires.

[22] "two **wrenching and emotional issues** will remain: the **future of Jerusalem**, and the fate of Palestinian refugees."

NMT: " قضيتين مؤلمتين وعاطفيتين ستبقان: مستقبل القدس ومصير اللاجئين الفلسطينيين"

Here he implies the division of Jerusalem into an eastern part for the Palestinians and a western one for the Israelis and the non-return of refugees. The two issues signify cases of religious bias (ignoring history and violating international law) and racial bias (against the Palestinians). Ironically, even the illegal division of an occupied land between the occupier and the occupied proposed by Obama in 2011 was devastated by President Trump's shameful Statement on Jerusalem on 6 December 2017, whereby he announced 'Jerusalem is Israel's capital', acknowledging occupation. Also the issue of refugees has not been resolved yet. The target text transferred the message accurately.

One wonders how a biased situation as such may lead to a 'lasting, comprehensive, and just' peace in the region, as the US calls for more than three decades. In the next example, Obama talks about a 'more urgent' (أكثر إلحاحًا) drive for the two concerned parties to reach a 'lasting peace' (سلام دائم). Of course ending the occupation can be a drive for the Palestinians. But to make concessions of everything including the land would not 'end the conflict' (ينهي الصراع), as he so desires. He hopes to impose peace through reaching an agreement whereby Palestinians 'resolve all claims' (ويحل جميع المطالبات); in fact, 'resolve' here means 'end'

or 'make concessions about' all Palestinian claims rather than 'restore rights', 'end occupation' or 'have a state proper'.

[23] "the drive for a **lasting peace** that **ends the conflict** and **resolves all claims** is **more urgent** than ever."

NMT: أصبح السعي لتحقيق سلام دائم ينهي الصراع ويحل جميع المطالبات أكثر إلحاحًا من أي وقت مضى."

Google managed to convey the essence of these multiply-biased messages accurately. He then moves to a threatening message for the Palestinians, the occupied people, in the following example:

[24] "For the Palestinians, **efforts to delegitimize Israel will end in failure**... And Palestinians **will never realize their independence** by **denying the right of Israel to exist**."

NMT: بالنسبة للفلسطينيين، فإن جهود نزع الشرعية عن إسرائيل ستنتهي بالفشل... ولن يحقق الفلسطينيون استقلالهم أبدًا من خلال إنكار حق إسرائيل في الوجود

The message in the original English text and its Arabic translation goes like: Palestinians should simply stop their rights and claims against Israel in the United Nations and everywhere because Israel is a legitimate state_ according to the message_ and any 'efforts to delegitimize Israel will end in failure' (جهود نزع الشرعية عن إسرائيل ستنتهي بالفشل). More importantly, Palestinians 'will never realize their independence' (ولن يحقق الفلسطينيون استقلالهم أبدًا) so long as they 'deny the right of Israel to exist' (إنكار حق إسرائيل في الوجود). An occupier, which the US stands by unconditionally, is given the right to occupy a land and if the occupied just thinks of its rights, it will be punished.

The bias towards Israel and the bias against Palestine are emphasised in the following fragile, naïve security argument that Israel has 'the right to self-defense' (الحق في الدفاع عن نفسها) and 'must be able to defend itself' (ويجب أن تكون إسرائيل قادرة على الدفاع عن نفسها):

[25] "As for security, **every state** has the **right to self-defense**, and **Israel must be able to defend itself – by itself – against any threat**."

NMT: "أما بالنسبة للأمن ، فلكل دولة الحق في الدفاع عن نفسها ، ويجب أن تكون إسرائيل قادرة على الدفاع عن نفسها - بنفسها - ضد أي تهديد."

The argument refers to Israel's right and 'every state has the right to self-defense' (لكل دولة الحق في الدفاع عن نفسها), but how about Palestine? Isn't it one of those 'every state'? Doesn't it have the right to self-defense, too? We will return to these questions in example [26]. Noteworthy, the choice of words in the phrase 'by itself' (بنفسها) must be connected to the

words 'any' (أي) and 'threat' (تهديد); this is an attempt to legitimise Israel's massive, especially nuclear, militarisation on the one hand and to defend itself against any possible threat from the Arab countries or Iran, which _ if and when it happens_ USA and its 'friends' will not stand. Again, the NMT transfers these ideologically-biased messages accurately.

Coming back to the questions raised in the previous paragraph, the answer might be evident through the status quo on land after 30 years of negotiations and 70 years after dividing Palestine in the UN Security Council Resolution 181(1947) and the establishment of Israel in 1948. The Division of Palestine Resolution gave Israel more than 55% of Palestinian territories. Before the 6-day war of 1967 Israel had already occupied 78% of Palestine; the rest of the 22% was occupied in that war. Therefore all that Obama was offering to the Palestinians is actually a 'Palestinian security responsibility' (المسؤولية الأمنية الفلسطينية):

[26] "with the assumption of **Palestinian security responsibility in a sovereign, non-militarized state**"

NMT: "مع تولي المسؤولية الأمنية الفلسطينية في دولة ذات سيادة وغير عسكرية"

What does it mean to have a 'security responsibility' (المسؤولية الأمنية) in a 'sovereign state' (دولة ذات سيادة)? I do not really understand what it means in political sciences or politics. Either you have a sovereign state with all its rights and duties domestically and rights and obligations internationally, or not. And how is it going to be a 'sovereign' state and it is 'non-militarized' (غير عسكرية)? How can Palestine defend itself against the massively militarised, nuclear occupying power? Google's neural translation successfully transfers the source message with its all implications. The establishment of an occupying state based on a religion is an embodiment of religious bias, hence any justification for its actions would look absolutely biased, illegal, illegitimate and unethical.

CONCLUSION

This study has made its aim clear from the beginning, namely to examine bias in neural machine translation (NMT) through exploring the translation of some heavily-loaded ideological messages from English into Arabic. It used a multidisciplinary perspective based on translation studies (the translation of ideology), political sciences (the neo-classical theory of Realism in international relations) and computer science (through concepts like artificial intelligence, neural machine translation and deep machine learning) to explore bias in Google's NMT of ideology in the sample data, extracted from Obama Remarks on the Middle East

and North Africa, 19 May 2011. It raised four research questions and employed a qualitative methodology relying on analysis, comparison and deduction as tools of research to answer them.

The first question asked: "How does NMT translate English→Arabic messages heavily loaded with ideological racial bias?" Racial bias, which can result from negative or positive stereotyping, is a prominent feature in Obama's Remarks. The study found out that the original messages are heavily-loaded with racial bias against the Arabs in the Middle East and North Africa and that Google transferred them accurately into Arabic.

The second question was: "How does NMT translate English→Arabic messages heavily loaded with ideological gender bias?" The study found out that the political discourse directed from USA to the Arabs has no gender bias at all. On the contrary, it urges the Arabs to 'empower' women and give them rights 'equal' to men's. Although this discourse sounds appealing to many people and feminists particularly, however, the study recommends that change in this respect should not be imposed on these countries because they have their own values, customs and traditions, and beliefs which should be respected. It also found out that this discourse was accurately transferred into the target text.

Coming to the third question "How does NMT translate English→Arabic messages heavily loaded with ideological religious bias?", the study discovered explicit, blatant cases of religious bias in the data. Obama's discourse not only seemed biased, but also naïve, contradictory and unable to give justification for his arguments. In his attempt to defend Israel unconditionally, the speech violates the various sources of international law and the least ethical rights of sovereign states. The translation accurately transferred all the implications of these religious and racial biases into Arabic.

By and large, Google's NMT was an accurate, not necessarily 'proper', translation. The findings actually contradicted my initial expectations that Google would have problems of bias and maybe quality, let alone accuracy. The product is surprisingly amazing. Definitely the Google translation we used to get before 2016 has tremendously improved. With the exception of very minor mistakes that do not affect the core message of the source text as in examples [11, 15, 18, 19 and 21], Google Translate is not biased according to the analysis of data: it transferred accurately racial and religious biases with no traces of gender bias. And this answers the fourth question of the study "How far is the current form of Google Translate biased?" Further research is recommended to assess NMT behaviour regarding bias in other sample data.

Appendix Abbreviations

AI: Artificial Intelligence

GT: Google Translate

ML: Machine Learning

MT: Machine Translation

NLP: Natural Language Processing

NLU: Natural Language Understanding

NMT Neural Machine Translation

REFERENCES

- Abdel-Fattah, Seif El-Din. (2006). Ro'aya fi afkar al-doctor Hamed Rabie. In Nadia M. Moustafa and Seif Abdel-Fattah (eds.), *Al-lugha wal hawayah wa hewar alhadharat*. Cairo: Faculty of Economics and Political Sciences, Cairo University. pp. 35-68.
- Ahmed, Safa'a Ahmed. (2019a). Translation as a Soft Power to Westernise Local Identities: An Arab Perspective. *Occasional Papers in the Development of English Education, CDELT*, 68(1) October, 385-402.
- Ahmed, Safa'a Ahmed. (2019b). Postcolonialism and Foreignisation as a Translation Strategy: A New Perspective to the Arab Family Values in Translated English Films. *Occasional Papers in the Development of English Education, CDELT*, 66 (A) April, 527-570.
- Ahmed, Safa'a Ahmed. (2016a). Translation and Shaping the Arab Identity in a Post-colonial Globalized World: A Multi-disciplinary Approach. *English Language and Literature Studies, ELLS*, X(1) (December), 403-456.
- Ahmed, Safa'a Ahmed. (2016b). The Other in Subtitled Documentary Films and Making History: A Multidisciplinary Perspective to Arabic/English Translation. *Occasional Papers in the Development of English Language Education, CDELT*, 62 (December), 297-327.
- Ahmed, Safa'a Ahmed. (2014). Ideological Translation and Mass Communication: A Modernization or a Conflict Enterprise? A Case Study of Al-Jazeera vs. Al Arabiya. *English Language and Literature Studies, ELLS*, XI (1), (December), 181-249.
- Allport, Gordon. (1979). *The Nature of Prejudice*. Cambridge: Addison Wesley.
- Angwin, J., Scheiber, N., & Tobin, A. (2017). Facebook Job ads Raise Concerns about Age Discrimination. *The New York Times*, 20 December.
- Baly, Ramy; Da San Martino, Giovanni; Glass, James; Nakov, Preslay. (2020). We Can Detect Your Bias: Predicting the Political Ideology of News Articles. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 16-20 November, pp.4982–4991.
- Barocas, S. & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.15779/Z38BG31>.
- Bender, Emily M. et.al. (2021). On the Dangers of Stochastic Parrots: Can Language Models be Too Big? In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT '21)*, pp.610–623.
- Bias in Artificial Intelligence. (2021). *Society Improve Diagnosis Medicine, Improve Dx Newsletter*, Vol.8(3), May.
- Billig, Michael. (1985). Prejudice, categorization and particularization: From a perceptual to a rhetorical approach. *The European Journal of Social Psychology*, 79-103.

- Blodgett, S.; Barocas, S.; Daumé III, H. & Wallach, H. (2021). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. <https://arxiv.org/abs/2005.14050>.
- Boegesius, Frederik Z. (2018). Discrimination, AI and Algorithmic Decision-Making. Strasbourg: Council of Europe.
- Review into Bias in Algorithmic Decision-Making. (2020). Centre for Data Ethics and Innovation. November 2020.
- Chen, L.; Ma, R.; Hannak, A. & Wilson, C. (2018). Investigating the impact of gender on rank in resume search engines. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing, pp.1–14.
- Chomsky, N. (2011). The Golden Age: A Look at the Original Roots of Artificial Intelligence, Cognitive Science, and Neuroscience [an interview with N. Chomsky at MIT150 Symposia]. Brains, Minds and Machines Symposium, 3 May 2011. <https://cbmm.mit.edu/video/keynote-panel-golden-age-look-original-roots-artificial-intelligence-cognitive-science-and>
- Clackson, Alexander. (2011). Conflict and Cooperation in International Relations. [part of a Master's programme]. E-Learning Relations, 1 February. <https://www.e-ir.info/2011/02/01/conflict-andcooperation-in-international-relations/>
- Costa-jussà, Marta, R & de Jorje, Andrià. (2020). Fine Tuning Neural Machine Translation on Gender-Balanced Datasets. In Proceedings of the Second Workshop on Gender Bias in Natural Language Processing, pp.26–34.
- Dastin, J. (2018). Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women. Reuters, 10 October/ <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.
- Data Quality & AI_Mitigating Bias and Error to Protect Fundamental Rights. (2019). FRA Focus. European Union Agency for fundamental Rights (FRA). 11 June. : fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf.
- Escudé-Font, Joel & Costa-jussà, Marta R.. (2019). Equalizing Gender Bias in Neural Machine Translation with Word Embeddings Techniques. In Proceedings of the First Workshop on Gender Bias in Natural Language Processing, Association for Computational Linguistics, pp.147–154.
- Farkas, Anna & Strobel, Maria. (2021). How to Measure Gender Bias in Machine Translation: Real-world Oriented Machine Translators, Multiple Reference Points. *Social Sciences & Humanities Open* 5 (2022) 100239, 1-11.

- Farkas, Anna & N´emeth, Renata. (2022). How to Measure Gender Bias in Machine Translation: Real-World Oriented Machine Translators, Multiple Reference Points. *Social Sciences & Humanities Open* 5(2022), 100239, 1-11.
- Frank, Anke; Hoffmann, Christiane & Strobel, Maria. (2004). *Gender Issues in Machine Translation*. University of Bremen.
- Fuchs, Daniel J. (2018). "The Dangers of Human-Like Bias in Machine-Learning Algorithms." *Missouri S&T's Peer to Peer*, 2 (1) May. <https://scholarsmine.mst.edu/peer2peer/vol2/iss1/1>.
- Garrido-Muñoz, Ismael et.al. (2021). A Survey on Bias in Deep NLP. *Applied Sciences* 11April 2021, 3184, 1-26.
- Glymour, Bruce & and Herington, Jonathan. (2019). Measuring the Biases that Matter: The Ethical and Casual Foundations for Measures of Fairness in Algorithms. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19*, Association for Computing Machinery, pp.269–278.
- Grant, L. S., Gode, B. S., & Amstron, M. G. (2017). Ideology of Translation Concept Approach on Determining a Decision by the Translator. *Linguistics and Culture Review*, 1(1), 1-12. <https://doi.org/10.37028/lingcure.v1n1.1>.
- Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, Vol. 61(4), 5-14.
- Hatim, B. & Mason, I. (1997). *The Translator as Communicator*. London & New York: Routledge.
- Hovy, Dirk; Bianchi, Federico & Fornaciari, Tommaso. (2020). "You Sound Just Like Your Father": Commercial Machine Translation Systems Include Stylistic Biases. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp.1686–1690,
- Hovy, Dirk; Prabhumoye, Shrimai. (2021). Five Sources of Bias in Natural Language Processing. *Lang Linguist Compass*. 2021;e12432, 1-19.
- Jones, David. (2020). Characterising the Digital Twin: A Systematic Literature Review. *CIRP Journal of Manufacturing Science and Technology*, Vol.29, 36–52.
- Jervis, Robert. (1982). "Security Regimes", *International Organization*, Vol. 36(2).
- Jurgens, D., Tsvetkov, Y., & Jurafsky, D. (2017). Incorporating Dialectal Variability for Socially Equitable Language Identification. In R. Barzilay, & M.-Y. Kan (eds.), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vol. 2, pp.51–57.
- Kaul, Mira. (2021). Understanding Human Bias in Artificial Intelligence: Suggesting a Future for Being More Objective, and Less Subjective. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, Vol. 10(7) July, 9886-9890.

- Kiritchenko, Svetlana & Mohammad, Saif. (2018). Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, June, pp.43-53. <https://aclanthology.org/S18-2005>.
- Kurpicz-Briki, Mascha. (2020). Cultural Differences in Bias? Origin and Gender Bias in Pre-Trained German and French Word Embeddings. CEUR-ws.org/vol.1-2624, 1-11.
- Liu, Riubu et.al. (2021). Quantifying and Alleviating Political Bias in Language Models. *Artificial Intelligence*, 304 (2022) 103654, 1-16.
- Livingston, Morgan. (2020). Preventing Racial Bias in Federal AI Morgan Livingston –Policy Memo: Prevent Racial Bias in AI. *Journal of Science Policy & Governance*, Vol. 16(2), May, 1-7.
- Měchura, Michal. (2022). We Need to Talk about Bias in Machine Translation. Fairslator Whitepaper, Version 1.0, published 28 February 2022.
- Medel, Maria Lopez. (2021). Gender Bias in Machine Translation: An Analysis of Google Translate in English and Spanish. *Academia Letters*, July, 2288. <https://doi.org/10.20935/AL2288>.
- Mitchell, Margaret et.al. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 220–229, 2019.
- Murphy, Nichole. (2021). Types of Bias. CPD College Online, 10 November 2021. <https://cpdonline.co.uk/knowledge-base/safeguarding/types-of-bias/>
- Nelson, G. S. (2019). Bias in Artificial Intelligence. *North Carolina Medical Journal*, Vol. 80(4), 220-222.
- Ntoutsis, Eirini et.al. (2020). Bias in Data-Driven Artificial Intelligence Systems_An Introductory Survey. *WIREs Data Mining Knowledge Discovery*, December, 1-14.
- Osoba, Osanda & Welser, William. (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica, Cliff: Rand Corporation.
- Pagano, Tiago Palma et.al. (2022). Bias and Fairness in Machine Learning Models: A Systematic Literature Review. *ResearchGate*, February, 1-19.
- Pérez, M Caldaza. (2003/2014). Introduction. In M. Calzada Pérez (eds.), *Apropos of Ideology*. London& New York: Routledge. pp. 1-22.
- Prates, Marcelo; Avelar, Pedro & Lamb, Luis C. (2020) Assessing Gender Bias in Machine Translation: A Case Study with Google Translate. *Neural Computing and Applications*, May, 1-34.
- Rabie, Hamed. (1986). *Al-Elaqa Al-Ettisalia baina al-Mafhoum al-Qawmi Wal-Tatawor Al-Igtimaei* (2nd edn). Al-Lugha Al-Arabia wal-Waei Al-Qawmi Symposium. Beirut: Centre for Arab Unity Studies.

- Review into Bias in Algorithmic Decision-Making. (2020). Centre for Data Ethics and Innovation, 27 November 2020. <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making>
- Rice, Condoleeza (2003). Transforming the Middle East. The Washington Post, 7 August 2003. <https://www.washingtonpost.com/archive/opinions/2003/08/07/transforming-the-middle-east/2a267aac-4136-45ad-972f-106ac91e5acd/>
- Saunders, Danielle & Byrne, Bill. (2020). Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, July 5-10, pp. 7724–7736.
- Saunders, Danielle; Sallis, Rosie & Byrne, Bill. (2020). Neural Machine Translation Doesn't Translate Gender Coreference Right Unless You Make It. arXiv:210.05332v2[cs.CL]10 December 2020, 1-10. <https://doi.org/10.48550/arXiv.2010.05332>.
- Savoldi, Beatrice et.al. (2021). Gender Bias in Machine Translation. Transactions of the Association for Computational Linguistics, Vol. 9, 845–874.
- Schiebinger, Londa. (2014). Scientific Research Must Take Gender into Account. Nature, Vol. 507(7490), 6 Mar 2014.
- Schwarm, A. (2018). Amazon, Machine Learning, and Gender Bias. LinkedIn, 31 October <https://www.linkedin.com/pulse/amazon-machine-learning-gender-bias-alex-schwarm/>.
- Sengupta, Kinshuk & Srivastava, Praveen Ranjan. (2022). Causal Effect of Racial Bias in Data and Machine Learning Algorithms on User Persuasiveness & Discriminatory Decision Making: An Empirical Study. 2 February 2022, Cornwell University. <https://arxiv.org/abs/2202.00471>
- Shah, Deven; Schwartz, Hansen & Hovy, Dirk. (2020). Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp.5248–5264.
- Silberg, Jake and Manyika, James. (2019). Notes from the AI Frontiers: Tackling bias in AI (and in Humans). Remarks from the Multidisciplinary Symposium on Ethics in AI, hosted by DeepMind Ethics and Society, June.
- Stafanovičs, Artūrs; Bergmanis, Toms & Pinnis, Mārcis. (2020). Mitigating Gender Bias in Machine Translation with Target Gender Annotations. In Proceedings of the Fifth Conference on Machine Translation, Association for Computational Linguistics, pp.629–638.
- Tahir-Gürçağlar, Ş. (2013). The translation Bureau Revisited: Translation as Symbol. *Apropos of Ideology*, 113-130.

- Tomalin, Marcus. (2021). The Practical Ethics of Bias Reduction in Machine Translation: Why Domain Adaptation is Better than Data Debiasing. *Ethics and Information Technology*, 6 March 2021, 1-15. <https://doi.org/10.1007/s10676-021-09583-1>
- Wang, Jun; Rubinstein, B.I.P & Cohn, Trevor. (2022). Measuring and Mitigating Name Biases in Neural Machine Translation. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 22-27 May 2022, Vol.1: Long Papers*, pp.2576 – 2590.
- Wolff, Stefan (2015) 'The Study of Conflict in Political Science and International Relations'. 1-17. www.stefanwolff.com/the+study+of+conflict+in+political+science+and+international+relations/
- Zajda, J. (2014). Ideology. In D. Phillips (ed.), *Encyclopedia of Educational Theory and Philosophy*. Thousand Oaks: Sage. Jan 2014. <file:///C:/Users/FUJITSU/Downloads/IDEOLOGYZajda.pdf>.
- Zajko, Mike. (2022). Artificial Intelligence, Algorithms, and Social Inequality: Sociological Contributions to Contemporary Debates. *Sociology Compass*, 2022, 16:e12962, 1-16.
- Zong, Zhaorong. (2018). Research on the Relations between Machine Translation and Human Translation. *Journal of Physics: Conf. Series* 1087 (2018) 062046. 1-7.