# Artificial Intelligence, the Secret Industry: The Politics of Technology and Translation

**Prof. Dr. Safa'a A. Ahmed**
Professor of Translation, Linguistics
Faculty of Languages, MSA University

## Abstract

Artificial Intelligence (AI) industry is developing very fast but without the necessary oversight or regulation that guarantees the openness, rather than the secrecy, of the industry. Valuable, essential information is hidden from the international public at a time when the nature of AI applications and consequently risks are becoming more globalised. The development of advanced AI systems is quite often kept out of the academia and the civil society's reach under allegations like competition laws or issues of intellectual property. Meanwhile, the global propaganda made for artificial intelligence (AI) is much bigger than the scope of our imagination that it raises valid questions about the nature of this industry and the one or ones who control it. Therefore, the present study aims to investigate the secret industry of AI and the politics of technology and to explore the role of translation in this regard. For this purpose to be achieved, this study uses a multi-disciplinary approach, deriving its tenets from world politics, computer engineering and translation studies and employs basically content analysis and interpretation as research methods to analyse the data. It challenges our understanding of and conceptions about AI technology and translation, thus affecting the field of AI in general, national policies and security, international peace and security, users, and translation.

**Keywords:** AI Risks; Secret Industry; Politics of Technology; Globalisation; International Peace and Security; Translation

**Abbreviations**:

(This is a list of abbreviations repeatedly used in the main body of the research)

**AI**: Artificial Intelligence
**AGI**: Artificial General Intelligence
**ANI**: Artificial Narrow Intelligence
**ASI**: Artificial Super Intelligence
**IR**: Image Recognition
**MT**: Machine Translation
**NLP**: Natural Language Processing
**VR**: Voice Recognition

**Figures:**

# الذكاء الاصطناعي، الصناعة السرية: سياسات التكنولوجيا والترجمة

الأستاذة الدكتورة صفاء أحمد
أستاذة الترجمة، اللغويات
كلية اللغات، جامعةMSA

## ملخص

تشهد صناعة الذكاء الاصطناعي تطورًا سريعًا، ولكن دون الرقابة أو التنظيم اللازمين لضمان شفافية هذه الصناعة بدلًا من سريتها. وهنالك معلومات قيّمة وأساسية يتم إخفاؤها عن الرأي العام العالمي لا سيما في وقت تتزايد فيه عولمة الذكاء الاصطناعي وتطبيقاته، وبالتالي عولمة مخاطره أيضا. وغالبًا ما يُحجب تطوير أنظمة الذكاء الاصطناعي المتقدمة (Advanced AI) عن الأوساط الأكاديمية والمجتمع المدني، تحت مزاعم مثل قوانين المنافسة أوالملكية الفكرية. وفي الوقت نفسه، تتجاوز الدعاية العالمية للذكاء الاصطناعي حدود خيالنا، ما يثير تساؤلات مشروعة حول طبيعة هذه الصناعة ومن يتحكم بها. لذلك، تهدف هذه الدراسة إلى دراسة الصناعة السرية للذكاء الاصطناعي وسياسات التكنولوجيا، والكشف عن دور الترجمة في هذا الصدد. وهي تعتمد على نهج متعدد التخصصات، مستوحى من مبادئ السياسة الدولية وهندسة الحاسوب ودراسات الترجمة، وتستخدم تحليل المحتوى وتفسيره بشكل أساسي كأدوات للبحث وتحليل البيانات. من ثمّ، تتحدى الدراسة فهمَنا ومفاهيمنا حول تكنولوجيا الذكاء الاصطناعي والترجمة، مما له أكبر الأثر على مجال الذكاء الاصطناعي بشكل عام، وعلى السياسات والأمن الوطني والسلم والأمن الدوليين والمستخدمين والترجمة.

**الكلمات المفتاحية**: مخاطر الذكاء الاصطناعي؛ الصناعة السرية؛ سياسات التكنولوجيا؛ العولمة؛ السلم والأمن الدوليين؛ الترجمة

# Artificial Intelligence, the Secret Industry: The Politics of Technology and Translation

**Prof. Dr. Safa'a A. Ahmed**
Professor of Translation, Linguistics
Faculty of Languages, MSA University

## §1. Introduction:

> AI companies possess substantial non-public information about the capabilities and limitations of their systems, the adequacy of their protective measures, and the risk levels of different kinds of harm. However, they currently have only weak obligations to share some of this information with governments, and none with civil society. We do not think they can all be relied upon to share it voluntarily. (A Right to Warn 2024)

Artificial Intelligence (AI) industry is developing very fast but without the necessary oversight or regulation that guarantees the openness, rather than the secrecy, of the industry. Valuable, essential information is hidden from the international public at a time when the nature of AI applications and consequently risks are becoming more globalised. The development of advanced AI systems is quite often kept out of the academia and the civil society's reach under allegations like competition laws or issues of intellectual property. Meanwhile, the global propaganda made for AI is much bigger than the scope of our imagination that it raises valid questions about the nature of this industry and the one or ones who control it. The conception that the AI margin which we, as end-users or even developers, are allowed to move within would make us excel in this industry or outperform the owners seems not only naïve but also immature and unwise. Also, the idea that AI can do anything and will excel the human brain can imply misleading lies. A globalised defeat and surrender discourse has helped enhance a terrifying, weak de facto for many countries. Therefore, this study aims to explore the secret industry of AI and the politics of technology and to investigate the role of translation in this regard. For this purpose to be achieved, this study uses a multi-disciplinary approach, deriving its tenets from world politics, computer engineering and translation studies and employs basically content analysis and interpretation as research methods to analyse the data. It challenges our understanding of and conceptions about AI technology and translation, thus affecting the AI field in general, national policies and security, international peace and security, users, and translation.

For simplification, AI is a branch of computer science dealing with developing machines which can learn, make decisions, and perform tasks like humans. It is "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments" (Executive Order 2023). At the heart of the mind, as it were, of these machines lies the group of rules or instructions (called algorithms) given to them to perform a task. So, algorithms are procedures or sequences of rules, instructions, or programmes that guide training (Gurevich 2012) and set to solve a particular problem. "A set of rules defining how to perform a task or solve a problem. In an AI context, this usually refers to computer code defining how to process data" (Boucher 2020:VI). In my opinion, these algorithms are quite dangerous because they can be ethical or unethical, if I may use the term 'ethical' to refer to what should be done simply, depending on what the owners want. Another important factor in the machine decision-making and -taking processes is datasets, fed to the machine to provide it with experiences and information which act like past experiences and memory in the human mind. Datasets are an organised collection of data, defined by content, purpose, grouping or relatedness and used for analysis and modelling. Hence, we can imagine the situation if algorithms or datasets are unethical, for example. This industry requires transparency, at every level of developing and implementing applications, including algorithmic and databases transparency and accountability.

Algorithmic transparency, thus, is "about disclosing how algorithmic tools enable decision-making by public policy makers and regulators by providing information in an open, understandable, easily accessible, and free format", and algorithmic accountability means "the ability of those who design, build, procure, or implement the algorithm to be held responsible for their actions and impact according to policies and laws concerning algorithm use" (Stankovich et al. 2023:10). Most importantly, the industry owners themselves should be transparent and held accountable.

There are various categorisations of AI types. They can be categorised based on AI capability to learn and apply knowledge into: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI), see Figure 1. Based on functionality, i.e. how it applies learning capabilities to process data, respond to stimuli and interact with the surrounding environment, they can be categorised into Reactive Machine AI, Limited Memory AI, Theory of Mind and Self-Aware AI. The present paper adopts the first classification.

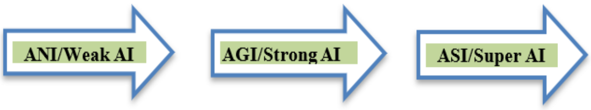

**Figure 1**: AI Phases

Artificial Narrow Intelligence (ANI) is a weak version of AI and is designed to perform specific tasks according to a particular set of inputs which lead to particular outputs. It refers to “the current paradigm of AI tools which exhibit intelligence only in specific niches such as playing chess or recognising cats" (Boucher 2020:VII). In this case, the machine cannot learn by itself. It can use machine cognition and reasoning, machine learning and neural network algorithms as in Natural Language Processing (NLP), and reactive AI or limited memory AI, see Figure 2:

Cognition and reasoning
Use of natural language processing (NLP)
AI
Reactive AI or Limited memory AI

**Figure 2**: Features of Artificial Narrow Intelligence (ANI).
*Source*:https://www.engati.com/glossary/artificial-narrow-intelligence

Examples for this type include Voice Recognition (VR) and Image Recognition (IR) applications, purchasing or search recommendations, and self-driving cars, etc.

Artificial General Intelligence (AGI), on the other hand, is a strong version of AI and is designed to learn, think and make many tasks similar to humans. It depends on abstract thinking, creativeness, background knowledge, comprehension of course and effect, common sense in making decisions and transferring learning, see Figure 3. It acts like assistants as smart as humans. Take for example generative AI systems like ChatGPT, which can write an essay or answer a question upon some prompts given to it; and here risks can include academic integrity, misinformation, lying, directing, validity, biases, etc. (cf. Ahmed 2024a).

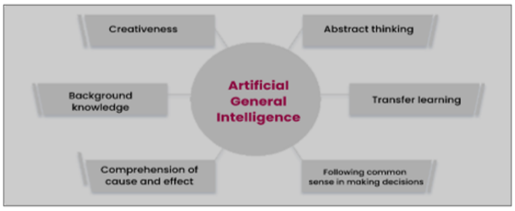
Creativeness
Abstract thinking
Background knowledge
Artificial General Intelligence
Transfer learning
Comprehension of cause and effect
Following common sense in making decisions

**Figure 3**: Features of General Artificial Intelligence (AGI)
Source:https://codiste-website.cdn.prismic.io/codiste-ebsite/Zg_PeRrFxhpPBVBf_CharacteristicsofAGI.svg?auto=compress,format

Artificial Super Intelligence (ASI) is a super AI and is designed to outperform the human knowledge and capabilities. The publicly available applications have not reached this stage yet. Though AI companies and experts warn against its potential dangers, full secrecy surrounds this industry in particular. There are no examples from the real world we can give here, but a very simple, primitive example is individualistic robots like Sohpia, Figure 4. This type of AI aspires to create a machine that surpasses what humans can do, as companies propaganda claims. Its imaginary potentials seem like a Hollywood fictionist movie, as in Figure 5.

**Figure 4: Sophia Robot** https://www.newsweek.com/sophia-robot-saudi-arabia-women-735503

**Figure 5**: Futuristic Imaginary AI *Source*:https://www.justthink.ai/artificial-general-intelligence/the-creative-potential-of-artificial-general-

Deep learning and the algorithms behind how the machine makes and takes decisions are still not understood totally in ANI, e.g. which applicant gets the job, which student is accepted in a school or a college, which bank customer gets the loan, etc. If this is the situation now, one may wonder, what about stronger or super versions of AI? Tommi Jaakkola, a professor at MIT, says "Whether it's an investment decision, a medical decision, or maybe a military decision, you don't want to just rely on a 'black box' (Knight 2017). Deep learning is "a particularly dark black box" and "Even the engineers who built these apps cannot fully explain their behavior", Knight argues. In deep learning, the machine generates its own algorithms according to data and desired outcome. We may know that inputs are fed to thousands of simulated neurones which are arranged along interconnected neural layers where they pass through the first layer and give a particular outcome, which in turn goes through the next layer and so on, maybe for hundreds of layers. Yet, we do not know what happens inside the box, how these neurones behave and how they result in the final machine decision.

Bletchley Declaration, an agreement signed by 29 countries on AI safety in 2023, mentions that all AI types pose risks and harms, including even ANI:

> Particular safety risks arise at the ‘frontier’ of AI, understood as being those highly capable general-purpose AI models, including foundation models, that could perform a wide variety of tasks - as well as relevant specific narrow AI that could exhibit capabilities that cause harm - which match or exceed the capabilities present in today’s most advanced models. (Bletchley Declaration 2023)

Risks will affect all, users and non-users. Hence, transparency and accountability are vital for everyone around the globe. Here comes the significance of this study. Secondly, at the theoretical frameworks, it explores the issue from a multidisciplinary perspective and thus offers other dimensions to understanding AI as a secret industry and the role of translation in this concern. Thirdly, at the level of documents and data collected. It draws our attention to the importance of being aware and reactive, instead of being just submissive users.

This paper is divided into 6 sections in addition to the implications and conclusion. §1 is an introduction. §2 reviews the literature on the topic. §3 presents the theoretical framework and the research method. §4 tackles how this secret industry is reflected in the documents of 3 of companies that control this industry. §5 investigates 3 documents issued by the US, UK and world governments. Meanwhile §6 discusses how translation can be used to enhance this secrecy. Finally, comes the conclusion.

## §2. Literature Review

### 2.1 Serious Warnings

On 4 June 2024, a group of Google DeepMind and OpenAI staff released an important open Letter, called 'A Right to Warn about Advanced Artificial Intelligence', that it lacks transparency, is unregulated, and can lead to 'human extinction':

> We also understand the serious risks posed by these technologies. These risks range from the further entrenchment of existing inequalities, to manipulation and misinformation, to the loss of control of autonomous AI systems potentially resulting in *human extinction*. (A Right to Warn 2024)

The employees explain that "AI companies themselves have acknowledged these risks, as have governments across the world and other AI experts". They mention that such risks should be mitigated and guided by professionals, policy makers and the public. But unfortunately, companies refuse 'effective oversight', and here arises the danger of the secrecy of the industry. They assure that those companies have serious 'non-public information', which they hide from public eyes and are not ready to share, although such a kind of information can harm peoples in the whole world:

> AI companies possess *substantial non-public information* about the capabilities and limitations of their systems, the adequacy of their protective measures, and the risk levels of different kinds of harm. However, they currently have *only weak obligations to share some*

*of this information with governments, and none with civil society.* (A Right to Warn 2024)

Companies have only 'weak obligations' to share ‘substantial’ information related to systems capabilities, limitations, risks and harms publicly. They do not discuss the adequacy of their protective measures with governments or even relevant expertise.

There are confidentiality rules that hinder such employees from speaking publicly about their concerns. They can just express their fears to their companies, which do not address these fears often under allegations of protecting trade secrets and intellectual property rights. Upon hiring, they were obliged to sign a disclosure statement and various confidentiality agreements which "block us from voicing our concerns" and consequently are not allowed to disclose any information against their companies (A Right to Warn 2024). They think that these companies are not trusted to share concerns 'voluntarily'.

Geoffrey Hinton, father of AI, quitted Google in May 2023 to speak freely about risks, and he signed the Letter, too. He expressed his "concerns over the flood of misinformation, the possibility for AI to upend the job market, and the 'existential risk' posed by the creation of a true digital intelligence… existential risk of what happens when these things get more intelligent than us" (Taylor and Hern 2023). He adds "I've come to the conclusion that the kind of intelligence we're developing is very different from the intelligence we have" (Taylor and Hern 2023). Kleinman and Vallance (2023) indicate that Yoshua Bengio, another Godfather of the industry who signed the Letter, suggests that "we need to take a step back" because of "unexpected acceleration". Alex Grant in his book 'The Dark Side of AI: Geoffrey Hinton's Warning' (2023) discusses Hinton's insights into machines autonomous decision-making, ascendancy and algorithmic bias and his call for action which should include a responsible approach by fostering transparency. Elon Musk, the famous businessman who owns Tesla, SpaceX, OpenAI, among others, told Fox News that Google co-founder Larry Page wanted ASI, which Musk describes as 'a digital god', as soon as possible (Taylor and Hern 2023).

### 2.2 Algorithms & the Black Box

Stankovich et al., in their paper 'Toward AI Meaningful Transparency and Accountability of AI Algorithms in Public Service Delivery’, argue that AI can cause harm to public policy delivery and human rights through bias and privacy issues and "frequently lack transparency and accountability" (2023:5). They describe its tools as 'black boxes' whose reasoning, i.e. 'algorithmic opacity', is difficult to understand by 'human beings' (p.5). They warn against the consequences if this technology 'goes awry' (p.13). Therefore, the authors recommend the following measures to foster transparency and accountability in public service delivery:

1-to suggest a human-rights approach to build and govern reliable systems;
2-to use simplicity, context and trust for algorithmic transparency;
3-to fill in the implementation gap;
4-to tailor algorithmic transparency to local cultures, economies and development contexts; and
5-to utilise a multi-stakeholder approach and partnerships between public and private sectors to enhance digital literacy. (p.13)

Describing AI as a 'speculative technology', Boucher (2020:19-20) identifies four transparency challenges. First, experts themselves cannot explain how algorithms work, i.e. how the machine takes decisions, let alone users. Second, some actors exploit imbalances in information access to serve their commercial and strategic interests, e.g. manipulating data to set prices for each individual consumer according to his willingness to pay. Third, users are not always certain about whom they are interacting with, a human or a machine. Finally, there is a lack in transparency about the expected developments. He wonders which data inputs shape algorithms, which features shape the machine decision and which changes in data inputs are made to change that decision (p.44). Boucher, therefore, suggests some measures to increase accessibility to data and algorithms, like:
1-to call for writing good commands by software engineers in order to help others understand the algorithm operation;
2-to make use of the available explainability mechanisms which explain how algorithms work;
3-to offer more initiatives to access data and algorithms through open sources and creative commons;
4-to create a competitive market for services which can increase users' control over their data and enhance more accountable practices;
5-to present more open AIs on platforms which allow third parties to access data; and
6-to improve trust and respond to uneven distribution of the benefits and risks of sharing data. (p.44)

The human contribution to machine learning is limited to writing initial codes or algorithms and sometimes supervising it during its learning process, Ali and Yu (20021:3) clarify. The machine behaves according to algorithms and input data. Flaws in design or implementation, like wrong generalisations or errors in defining context, result in its misbehaviour. Hence, the call for oversight becomes 'exceptionally compelling' (p.4). They declare that companies' confidential information is important but they go far to protect this confidentiality:

> Hence, *companies go to great lengths to secure the secrecy of their AI systems*. These include s*ecurity and access control mechanisms*, *confidentiality agreements, clauses in employment contracts*, and even *frequent changes in the algorithms powering their AI systems* to thwart unscrupulous parties. (p.6)

Companies maintain the secrecy of their industry through algorithms, their rules related to security and access to control systems, their confidential agreements, employment contracts, among others. Under such allegations, it is not surprising that "corporations openly oppose transparency and try to resist the forced disclosure of their AI components", they conclude (p.6). But what constitutes trade secrets is still controversial and needs more laws and international agreements.

No tangible actions have been taken for transparency and accountability and the industry retains many secrets. The Letter, which explains that AI represents an existential risk to humans, admits that companies' ‘strong financial incentives’ or drives enhance secrecy and may hinder efforts for transparency:

> AI companies have *strong financial incentives* to *avoid effective oversight*, and we do not believe bespoke structures of corporate governance are sufficient to change this. (A Right to Warn 2024)

The UK government confesses that humans may not be able to control the advanced systems. "As advanced AI systems become increasingly capable, autonomous, and goal-directed, there may be a risk that human overseers are no longer capable of effectively constraining the system’s behaviour" (Policy Paper Introducing 2024).

Thus, from this review of the literature, there is still a gap in our understanding of and knowledge about this ‘secret’ industry, transparency, accountability, algorithms, databases, and who owns this industry and their interests, in regard to all types, ANI, AGI and ASI.

## §3. Theoretical Framework & Methodology

The secrecy of AI technology, in my opinion, cannot be interpreted without reading the scene from a political perspective. The same applies to the globalised translation trend in the context of its development. Here is where three disciplines intersect together: world politics, computer engineering and translation studies.

### 3.1 Politics of AI Technology

In the recent years, AI technology has proved to be a great power in the hands of its owners. Allen and Massolo consider technology, in geopolitical terms, "a key driver of any transformation of power at the international level" and "could abruptly change balances of power in the international system" not only in the three traditional domains air, land and sea, but also in space and cyberspace (2019:7). Rizzo writes that it will change warfare in the coming decades into 'hyperbolic warfare' or 'hyper-war' is fuelled by AI and waged by machines; soldiers will not face a fair battle for the unparalleled speed of automated decision-making and for the concurrency of actions made by machines (2019:72,92). We are still trying to understand what is, or will probably be, going on:

> Artificial Intelligence, quantum technologies, robotics, autonomous weapons, and neural implants will all concur in transforming future warfare in ways we are only starting to understand. (Allen and Massolo 2019:8)

It is a new race for 'technological leadership' where the borders between the civil and military become unclear (p.8).

Also, Rugge describes AI as making available some 'emerging disruptive technologies which threaten international stability influencing the international balance of power and the 'rules of the game' (2019:16). He expresses his fears that there is the risk that:

> our adversaries will field a new disruptive military technology that provides them with an overwhelming military advantage they may use to our harm. In this sense, it is not the new technology per se that poses the greatest problem, but rather the asymmetric advantage that our adversaries receive from being the first to field it. (p.23)

This could mean paving the way for 'painful adjustments' to the international balance of power, to the 'rules of the game' and nuclear strategic stability, and to how wars are conducted. Indeed, whole nations become vulnerable in front of AI technology owners as such. It is, thus, a tool for power and control. This interprets the substantial secrecy pertaining to the industry.

### 3.2 Theories of Mind and Self-Awareness

Theories of Mind constitute the theoretical base for AI developments now and in the future, a phase that comes after reactive machines and limited memory machines. They come from different disciplines. A Theory of Mind aims to give machines the ability to and imitate human mental states, his beliefs, thoughts, desires, intentions, emotions, etc. Theorists try to understand how the human brain does all this. So, like the brain which consists of very complex neural networks and over 100 billion neurones, they have integrated machines with what they call minds which have neural networks and neurones in an attempt to create a machine able to understand and remember other entities' needs and emotions. The theories are still undergoing heavy research activities and at the moment theorists hope to improve machine-human interaction, as in ChatGPT and collaborative robots.

Another futuristic self-awareness theory is a stage beyond the Theory of Mind which aims to make a machine that understands emotions and has self-awareness or consciousness, i.e. to be aware of its own emotions, behaviour and needs and can evaluate the consequences of its behaviour. From the current data available to researchers and the public, this scenario seems theoretical. Research in this regard lacks transparency. Unfortunately, this type of research is not governed by ethics or international accountability rules.

At the heart of both Theories of Mind and Self-awareness Theory, lies the concept of algorithms. This brings us back again to the allegation that the lack of transparency and the secrecy of the industry are attributed to trade competition. It is an unacceptable, invalid allegation for two reasons. First, according to Article 1/2 and 39/2 of the Treaty on the Trade-related Aspects of Intellectual Property Protection (1994), 'trade secrets' should meet three requirements: that information must be secret, have a commercial value and be subjected to reasonable measures to keep it a secret. This should not apply here because the 'existential risks' are greater than any competition standards. Second, the couple of AI frontline companies appear to develop in a parallel way, as if someone harmonises their speed and the topics of their developments. They agree that future AGI and ASI look like nothing experienced by humanity before and that the developments should be gradual because societies should be prepared! This means that they probably have an idea about something specific rather than imagining a future scenario. Third, the individual(s) behind the scenes is the one who really comes on top of the globalised system, a system which the US President George W. Bush called a New World Order in the 1990s. Fourth, no one, even the few companies which seem to own this industry, has the right to dominate and control all humans in the 21st century.

### 3.3 Politics of Translation

Translation as a communication tool can be used as a means to achieve certain goals and agendas. Ahmed argues that colonialists and neocolonialists manipulated it as a soft power to colonise, globalise or westernise nations (2024b, 2020, 2019). It conveys ideologies and values and thus can shape people's knowledge about and attitudes towards certain sensitive issues which can direct beliefs and consequently behaviours. On the one hand, English, the language of power, exercises hegemony over the less powerful languages. On the other hand, translation into the less powerful languages may play the same centric role.

Tymoczko (2010:7) reveals that Eurocentric domination over translation is "an instrument of domination, oppression, and exploitation". Robinson (2002:31) demonstrates that translation in post-colonial contexts plays three overlapping roles, namely a 'colonisation channel' (for political dominance, economic exploitation, cultural hegemony, among others), a 'lightning-rod' (for revealing the imbalance in coloniser-colonised relationship after independence) and a 'decolonisation channel' (for correcting the deformed stereotype image about the colonised). In Bassnett and Trivedi's terms, translation is not an innocent or transparent activity, instead a 'manipulative activity' that "rarely, if ever, involves a relationship of equality between texts, authors or systems" (1999:2).

Therefore, realising this role, AI technology owners have encouraged the development of AI-generated translation tools to make sure that their values and ideas reach all users, freely in most cases. New language pairs are continuously added to tools to guarantee full access to each and every mind, if possible, around the globe.

### 3.4 Methodology

AI industry is developing very fast but without the necessary oversight or regulation that guarantees the openness rather than the secrecy of the industry. From this problem statement, the present study has set its aim to investigate the secret industry of AI and the politics of technology and to further explore the role of translation in this regard. It raises three research questions:

1-What is the nature of the secret side in AI industry?

2-What are the standpoints of big AI companies and governments?

3-What is the role of translation in AI industry?

To achieve the aims and answer the research questions, the study sets three objectives:

1-to collect data;

2-to analyse and reinterpret data in the light of the multidisciplinary theoretical framework to reveal the secrecy of the industry;

3-to explore the role of translation in this industry; and

4-to draw attention to the potential implications of such secrecy for various actors.

To this end, the study has chosen a qualitative methodology to suit the nature of the topic and its aims, using content analysis and study themes as tools of analysis in a multidisciplinary theoretical framework, as explained earlier.

The data collected in this paper was inspired by the Open AI staff Letter 'A Right to Warn about Advanced Artificial Intelligence' of 4 June 2024 for three reasons. Firstly, the staff who wrote the warning are not only AI experts, but they also worked in one of the big companies that is considered a major player in the field. Secondly, it referred to nine documents, which are crucial to the present paper. Thirdly, the Letter seems to have been under-researched. They mentioned documents issued by AI companies, governments and experts. They gave examples of the big companies Open AI, Anthropic and Google DeepMind. Documents for governments are the US government's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence on 30 October 2023; the UK government's Policy Paper Introducing the AI Safety Institute updated 17 January 2024 ad presented to Parliament by the Secretary of State for Science, Innovation and Technology, in November 2023; and Bletchley Declaration signed by various governments during AI Safety Summit, held on 1-2 November 2023. Meanwhile, they referred to 3 documents issued by experts, namely the Statement on AI Harms and Policy FAccT, the Encode Justice and the Future of Life Institute, and the Statement on AI Risk CAIS. Out of the nine critical documents, only the first six documents were selected to reflect companies' and leading governments’ visions and missions in this regard. Though the vision of experts is highly valuable, but their documents, mentioned above, are not included in the data for reasons of space and time only, a matter which manifests itself as a limitation of the study. The collected data as such has shaped the discussion part and divided it into three main sections; one explores secrecy in companies' documents, the second secrecy in governments' and the third translation role.

## §4. Secret Industry & Companies

This section analyses and interprets the thoughts of three of the leading AI companies, Open AI, Anthropic and Google DeepMind. I added italics to the data for emphasis and drawing the attention to certain wording.

### 4.1 OpenAI: Planning for ADI and Beyond

Established in 2015, OpenAI is an American AI research organisation known for its GPT language models, text-to-image applications and text-to-video systems. Its official website talks about Planning for AGI and

Beyond (Open AI 2023). Its 'announced' mission is ensuring that AGI benefits 'all of humanity':

> Our mission is to ensure that artificial general intelligence—AI systems that are generally *smarter than humans—benefits all of humanity*. (Open AI 2023)

The extension of some acclaimed 'benefits' to all humans implies that there are benefits and everyone should seek to have a share. It reflects domination and globalisation and further urges all humanity to apply AI systems. The claim that AI is ‘smarter' than humans is an exaggeration because the machine is just more productive and speedier in performing some tasks, i.e.: the machine can perform multiple mathematical operations in no time; and the machine can store and retrieve billions and billions of data, which the human brain cannot do. But it cannot be trusted to take decisions for humans in many sensitive situations It can be biased and mislead humans and it makes errors. Humans, who are created to instinctively interact and communicate with each other, may reject intruding machines to their personal lives at a certain point; when Covid-19 limited human interaction, most humans missed human-human interaction. And dependence on machines can negatively affect human skills, replace humans in the job market and thus threaten societal fabric, violate users' privacy and security, make grave mistakes that impact human life, etc. Over-dependence can cause many health risks and lead to technology addiction and death in the end (cf. Ahmed 2024a). So, the argument for 'benefits' needs reconsideration since the risks threaten human existence itself in various ways! Moreover, such a general appealing discourse deepens a sense of surrender to and non-critical thinking about the reality of the 'benefits'.

The choice of words in the company's propaganda is cunning. For example, it says:

> AGI has the potential to give *everyone incredible new capabilities*; we can imagine *a world* where *all of us* have access to *help with almost any cognitive task, providing a great force multiplier for human ingenuity and creativity*. (Open AI 2023)

The company talks about an AI that gives 'everyone' 'incredible' and 'new' capabilities. Can anyone resist such an offer that will make you a superman with incredible capabilities? Only the insane may. It promotes it as a 'great force multiplier' of our 'ingenuity' and 'creativity'. This mouth-watering marketing is very appealing to people and could block brains to even argue the possibility of the benefits to be merely a matter of lies!

Open AI company claims to work according to three principles. It 'wants' AGI to: 'empower humanity to maximally flourish in the universe', 'widely and fairly' share 'benefits, access, and governance’ with users, and 'successfully navigate massive risks'. The question how to move these principles from hopes and 'want to' into realities remains unanswered. That it will make humans 'maximally flourish in the universe' is just a general statement, an exaggeration at best and a lie at worst, and unverified promises. There is additionally no transparency about the nature of risks vs. the benefits.

Though OpenAI discloses that AGI would result in 'serious' risks of misuse, 'drastic accidents', 'accelerating an unsafe race', 'societal and economic disruptions', inter alia, it recommends to continue its development ‘forever’ whatever the consequences are or will be:

> Because the upside of AGI is so great, we do not believe it is *possible* or *desirable for society* to stop its development *forever*. (Open AI 2023)

In its opinion, the benefits are 'so great' to stop here. But what if benefits look like giving me a nice dish of food by one hand and stabbing me with a knife by the other. An unbelievable illogic! Look at inserting the word 'society' in this context to transfer the battle, as it were, from the battlefield of the company to the society's. AI is a product, like any other, developed by a multinational company seeking its own interests. Then why make it the battle of the society? Indeed, this serves to give it popularity and make users themselves an international public opinion tool to be directed for the defence of the project. The company believes that it is the decision of individuals to decide whether to use it or not:

> We believe in empowering *individuals* to make their own decisions and the *inherent power of diversity of ideas*. (Open AI 2023)

This strategy weakens the authority of governments and collective actions by experts and seemingly gives the individual this power. However, though it is the choice of the individual to use it or not, this is not a real choice because he is asked to ‘accept’ the company’s terms to use an application, or ‘unaccept’ which looks meaningless for a particular application since his privacy and security are violated by others. A child as an ‘individual’, for instance, is not an expert and does not know what is right to do and wrong to avoid; he should not be left alone to face multinational companies' overwhelming, sweeping, risky propaganda. The same applies to grownups. Another serious issue is the 'inherent power of diversity'. Diversity is a double sword, i.e. it is not in itself a guarantee for good 'power', it can be a destructive weapon. Diversity can mean accepting others and can mean division and disunity.

Whatever the consequences are, the survival of the project is more important to OpenAI, who ‘cannot predict’ where this project would lead to, as it assumes:

> Although we *cannot predict exactly what will happen*, and of course our current progress could hit a wall, we can articulate the principles we care about most. (Open AI 2023)

Ironically, it acknowledges that 'like any new field, most expert predictions have been wrong so far'. All it suggests amidst such an unpredictable AI future is to further deploy more systems to learn from experience, a 'gradual transition' to AGI world rather than a 'sudden one' since it "think[s] more usage of AI in the world will lead to good”. If there is something risky and unpredictable, how ‘will [it] lead to good”? One has the right to doubt. This clearly shows that it wants and insists on spreading more AI, instead of confronting risks, which are 'perhaps more impactful than everything else’, and despite the announcement that 'Success is far from guaranteed’. Spreading a harmful product by such a logic means spreading more risks, some of which, if not all, could be irreversible. It maintains that:

> We believe that democratized access will also lead to more and better research, decentralized power, more benefits, and a broader set of people contributing new ideas. (Open AI 2023)

It is a ‘belief’, not a truth; plus “Success is far from guaranteed”! This last phrase alone proves that the benefits are incomparable to the potential risks and harms of usage. It may be the ‘success’ of companies’ in achieving their interests and the ‘failure’ of ‘all’ humanity. Actually both 'democratized access' and 'decentralized power' refer to AI domination and control over powerless countries, who should concede their power and enjoy a Westernised version of 'democracy' instead, according to that irrational thinking!

### 4.2 Anthropic: Core Views on AI Safety

Anthropic is an American AI public-benefit corporation established in 2021. Its official website talks about Core Views on AI Safety (Anthropic 2023). Again, italics is added in the present paper for emphasis. Anthropic starts with comparing its impact to those of the industrial and scientific revolutions. But "we aren’t confident it will go well" and expect drastic impacts to emerge in the next decade.

Anthropic promotes AI systems to be "possibly equaling or exceeding human level performance at most intellectual tasks" in the next decade (Anthropic 2023). "We are most optimistic about a multi-faceted, empirically-driven approach to AI safety", it adds. It 'has a feeling' that machine capabilities will exceed 'our own capacities'; a matter which raises doubts about how to control this super powerful technology. The idea that "we aren't confident it will go well" cannot harmonise with developing a powerful technology as such and centralising it in the hands of few owners, thus endangering the future and even the existence of humanity. Hence, the secrecy about 'the thing we're working on' must come to an end and concerns in this regard must be addressed because human existence is more important than that 'thing'.

The magnitude of the change that future systems can bring about and whether they will act independently or generate information for humans represent an issue that 'remains to be determined' (Anthropic 2023). It admits that consequences of developing an AI that is smarter than humans, or more precisely, experts, can be terrible, 'dire':

> If we build an AI system that's significantly more competent than human experts but it pursues goals that conflict with our best interests, the consequences could be *dire*. (Anthropic 2023)

That is why it suggests to progress at a much slower pace and change to happen over 'centuries', not decades:

> While we might prefer it if AI progress slowed enough for this transition to be more manageable, taking place *over centuries rather than years or decades*, we have to prepare for the outcomes we anticipate and not the ones we hope for. (Anthropic 2023)

One may ask: what terrifying outcomes Anthropic is afraid to reach in years or decades and would rather prefer to reach in centuries? What kind of secret information it knows and goes beyond our imagination, anticipation and management? As if all peoples do not have the right to share such information threatening their existence.

Not all the potential risks of a rapidly-progressing industry are identified. Anthropic mentions just a few 'very disruptive' ones in employment job market, macroeconomics, and power structures domestically and internationally. The disruptions could be so 'catastrophic' and horrible that they could result in 'chaos' and more associated 'problems':

> These *disruptions* could be *catastrophic* in their own right, and they could also make it more difficult to build AI systems in *careful, thoughtful ways, leading to further chaos* and *even more problems with AI*. (Anthropic 2023)

The company agrees that progress has diverted from what was intended by its creators, a matter which has led to the emergence of problems like bias, unreliability, etc.:

> Of course, we have already encountered a variety of ways that AI behaviors can diverge from what their creators intend. This includes

> *toxicity, bias, unreliability, dishonesty*, and more recently *sycophancy and a stated desire for power*. (Anthropic 2023)

In fact, the reference in this statement to 'sycophancy' and 'a stated desire for power' proves and increases our fears regarding the global conflict for power, domination and control.

All these risks are engulfed into "in a multitude of hard-to-anticipate ways" because simply:

> *So far, no one knows how to train very powerful AI systems to be robustly helpful, honest, and harmless*. (Anthropic 2023)

A solution which, in my opinion, does not meet fears of human extinction, is suggested by Anthropic to comprise three ingredients "leading to predictable improvements in AI performance": training data, computation, and improved algorithms. Data could be biased or misleading; computation capabilities are owned by those who afford them; and algorithms are often a mystery. Yet, the existential risk far exceeds these ingredients and it necessitates transparency and accountability. Though Anthropic promises to "make externally legible commitments to only develop models beyond a certain capability threshold if safety standards can be met, and to allow an independent, external organization to evaluate both our model's capabilities and safety", this remains just promises and insufficient measures and procedures amidst much uncertainty where:

> One particularly important dimension of *uncertainty* is *how difficult it will be to develop advanced AI systems that are broadly safe and pose little risk to humans*. (Anthropic 2023)

This implies the risk of developing an AI that is neither 'broadly safe' nor 'pose little risk to humans'!

Anthropic discusses three possible scenarios, optimistic, intermediate and pessimistic scenarios. In the pessimistic scenario, still a possibility, AI cannot be controlled and must not be developed or deployed:

> AI safety is *an essentially unsolvable problem* – it's simply an empirical fact that *we cannot control or dictate values to a system that's broadly more intellectually capable than ourselves – and so we must not develop or deploy very advanced AI systems*. (Anthropic 2023)

Here, it explains that safety is an 'unsolvable problem', which they can neither 'control' nor 'dictate values' to because systems are smarter than humans. It 'hopes' to work towards "creating AI systems that are transparent and interpretable", etc. to guide policy makers and researchers. Only then when these 'hopes' come true, we can talk of a possible understanding of real risks.

### 4.3 Google DeepMind: Safety and Responsibility

DeepMind Technologies Limited, famous as Google DeepMind or DeepMind, is a British-American software company founded in London in 2010. It is a research laboratory subsidiary of Alphabet Inc. The UK Secretary of State for the Department of Science, Innovation and Technology asked DeepMind to share its approach to safety and responsibility for frontier AI in a policy paper, called AI Safety Summit: An Update on Our Approach to Safety and Responsibility, presented in AI Safety Summit 2023.

DeepMind identifies its aim clearly as "we aim to build AI responsibly to benefit humanity" (Google DeepMind 2023). It says that:

> We believe applying AI across *all sorts of domains* – including scientific disciplines, economic sectors, and to improve and develop new products and services– will *unlock new levels of human progress*. (Google DeepMind 2023)

The company make an irresistible appealing propaganda to apply it in ‘all’ sorts of domains, none excluded, claiming it will ‘unlock new levels of human progress’, i.e. the progress will exceed our imagination. Who may resist such a miraculous tool? Such a propaganda actually anaesthetising users. AI can support scientists in addressing some societal challenges better, like detecting breast cancer, achieving healthcare breakthroughs, limiting climate change effects, forecasting floods, etc., asserting further that "vast potential remains to supercharge scientific research and economic productivity, tackle global challenges like climate change and co-create new approaches to perennial policy priorities like education". It believes that AI paves the road to a new filed of computational biology and 'possible' transitions in disciplines such as 'energy, climate and education'. Definitely problems like energy and climate are vital for humanity to tackle, but look at the reference to 'education' in this regard. Education, in my opinion, means generating whole generations with special reformulated ideas and values, a result which far exceeds that which according to traditional mass media theories is produced by the repeated exposure of audience to certain messages.

Promoting it further, DeepMind maintains that AI releases human, yet-unleashed, potentials:

> Perhaps most exciting is the potential for AI to help release *human potential*: alongside helping solve problems that face us as a society.. (Google DeepMind 2023)

It is true that it offers 'assistive' applications. However, the size and return value gained due to using such applications should be assessed in comparison to the risks that have been observed and will possibly emerge in the future.

DeepMind states that researchers had noticed some risks, such as offensive cyber activities, deceiving people, manipulating them into performing harmful actions, developing dangerous weapons, developing high-risk models by 'people with malicious intentions', and causing harmful actions due to system failures. It talks about the necessity of achieving 'responsible AI' now. Noteworthy, it proposes some methods to face risks. It argues that "it is unlikely these methods [its process to face the risks] will be needed for today's models" (Google DeepMind 2023), a matter which unfortunately reflects that potential risks are more critical than we can imagine. The already grave, observed risks we experience now surpass what is coming! Hence, I may argue, DeepMind believes "advance preparation to mitigate future potential risks is important" while it mentions 'it is unlikely' to use the proposed methods at the moment. (Google DeepMind 2023)

The action plan proposed by DeepMind is based on two principles 'Be accountable to people' and 'Incorporate privacy design principles', which seem to me quite contradictory since the first means transparency while the second privacy, and so long as there are no clear-cut rules to remove the intervention between and vagueness of both principles. In other words, the problem is, among others, inherent in the use of algorithms in addition to datasets which companies subject to 'privacy' rules. 'Privacy' in a time when it threatens or may threaten human existence becomes meaningless vs. being 'accountable to people'.

DeepMind action plan suggests that AI will not be designed or deployed in areas where technology causes or may cause overall harm, in weapons and technologies which cause injury to people, in technologies which gather and uses information in violation to international norms, and in technologies which breach international law and human rights. Though these suggested actions seem not only proper and urgent, there is no guarantee or details of how to enforce them. For instance, in the first area where technology causes or may cause harm to people, DeepMind says that "Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints" (Google DeepMind 2023). This means that it can 'substantially' proceed designing and deploying harmful technologies at some 'unidentified' circumstances, and thus risks endure. Until then, such measures would remain appealing words, rather than deeds. "Responsible AI Council, specialized teams, evaluations, internal governance body, and collaborations with partners across

Google" will continue to be subject to being attested in what DeepMind would achieve in the next few years.

Google DeepMind states to be "an industry leader in transparency" through templates that document systems and to regard information security a key component in its models "that can be significantly misused, to ensure models with dangerous capabilities do not irreversibly proliferate" (Google DeepMind 2023). It makes available to customers information about its foundation models and 'other' (not all) tools for third-party developers including safe application and limitations. It also circulates information about the performance of systems and best practices. Indeed, it is a good practice to document and share information but what to document and share is more important.

## §5. Secret Industry & Governments

The following section addresses AI secret industry through the analysis and discussion of three documents issued by the US, the UK, and world governments, namely the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (2023), Policy Paper Introducing the AI Safety Institute (2024), and Bletchley Declaration (2023) respectively.

### 5.1 White House's Executive Order

The White House issued the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence ordered by the US president J. Biden on 30 October 2023. Like the previous documents mentioned in the present paper, the US government starts the Order with mouth-watering benefits that are promised not only to solve 'urgent challenges' but also to make the world more 'more prosperous, productive, innovative, and secure':

> Responsible AI use has the potential to help solve *urgent challenges* while making *our world more prosperous, productive, innovative, and secure*. (Executive Order 2023)

Look also at the use of the pronoun 'our' in 'our world' to imply that the president, as a representative of the US government, and the world are one, to imply that this is the project, or rather the agenda, of the addresses in the whole world so they have to defend and support the project, as if he was given a carte blanch to behave on behalf of the world.

A high sense of US superiority and 'leading' the world is clearly discerned in the Order, which claims the United States is 'compelled' to do so "for the sake of our security, economy, and society" (Executive Order2023). In other words, the Unites States alleges that it is the leader whom we should all thank for being heroically 'compelled' to accept to take responsibility. Biden sees his country superior since only the US superman is "capable of harnessing AI for justice, security, and opportunity for all". Notice the use of 'our' here, which literally refers to the US security, economy and society, unlike the previous 'our' in 'our world'.

In the same first paragraph of the Order, next to the sentence of benefits, Biden refers to potentially dangerous harms and risks that AI poses to nations, to societies, e.g. fraud, discrimination, bias, displacing and disempowering workers, national security, etc.:

> At the same time, irresponsible use could exacerbate *societal harms* such as *fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition*; and *pose risks to national security*. (Executive Order 2023)

This sentence actually reveals the ridiculous contradiction in AI discourse. On the one hand, it mentions wishes and on the other risks which should bombard any alleged benefits. Indeed, I can interpret such a contradiction only by assuming that benefits are merely lies that hide more dangerous agendas!

The idea of the reference in 'our' opens up another discussion about the concept of responsibility. When the Order refers to 'our' as a shared responsibility between builders and users, this legally means that the users, in this case the victims, accept the terms of that responsibility, whatever the terms are and whether the users have a say regarding these terms or not_ the latter do not even read the terms when they use an AI system in most cases. Once they choose the option 'accept' the terms, they have conceded their legal rights. The United States disclaims all responsibility and disclaims all liability. Biden assures that users are responsible as much as builders are:

> In the end, AI reflects *the principles of the people who build it, the people who use it*, and the data upon which it is built. ((Executive Order2023)

It is noteworthy to mention that although users seem to be given the choice between the two options to 'accept' or 'unaccept' the offer, they do not actually have that much choice at least since 'unaccept' a particular application does not mean that they and their data are not controlled by another.

To face risks, the Order suggests the following 8 guidelines and priorities:

1-AI must be safe and secure.

2-Responsible innovation, competition, and collaboration must be promoted.

3- Responsible development and use require a commitment to supporting American workers.

4-Policies must be consistent with US Administration’s dedication to advancing equity and civil rights.

5-The interests of Americans must be protected.

6-Americans’ privacy and civil liberties must be protected.

7-It is important to manage the risks from the Federal Government’s own use of AI and increase its internal capacity to regulate, govern, and support responsible use to deliver better results for Americans.

8-The Federal Government should lead the way to global societal, economic, and technological progress.

While the first and second principles seem to be general, the rest of the principles address the Americans, with their global role to lead the world in the last principle. USA, which suggests to lead and benefit the whole world, clearly discriminates the Americans and targets their interests only. Also, the first principle is too general and no specific concrete actions were mentioned. The current status quo proves that the phrase ‘safe and secure’ AI is merely void words since we have neither safe nor secure in the present or in the future, in my opinion. The second principle ‘Responsible innovation, competition, and collaboration must be promoted’ can thus be interpreted in the light of the US endeavours to benefit from others through putting the others’ innovation and collaboration under American control, again to benefit Americans, or more precisely AI owners.

Hence, a sentence like the following one seems a lie amid this grave contradiction. Biden argues that those principles will be promoted while:

> leading key global conversations and collaborations to ensure that *AI benefits the whole world*, rather than exacerbating *inequities, threatening human rights, and causing other harms*. (Executive Order2023)

He seeks the benefits of Americans, as mentioned in the principles, in an explicit racial discrimination against non-Americans, while claiming “AI benefits the whole world”. So, the talk about ‘inequities’ appear to be meaningless. As an AI discourse strategy, fear is utilised in ‘exacerbating inequities’, 'threatening human rights’ and ‘causing other harms’ in order to mobilise the international public opinion around it and one globe policy and to guarantee the American tight control of users. Humans instinctively gather around anyone who claims to defend them in danger; they may become paralysed, voluntarily or involuntarily, and unable to think.

### 5.2 UK Policy Paper

The Policy Paper Introducing the AI Safety Institute is presented to the UK Parliament by the Secretary of State for Science, Innovation and Technology in November 2023 and updated by government on 17 January 2024. The ministerial forward includes a big propaganda to the extent that the Paper refers to ChatGPT release as a ‘Sputnik moment for humanity’:

> The release of ChatGPT was *a Sputnik moment for humanity* – we were surprised by rapid and unexpected progress in a technology of our own creation. With accelerating investment into and public adoption of advanced AI, these systems are becoming *more powerful and consequential to our lives*. (Policy Paper 2024)

The UK government argues that it is becoming more ‘consequential’ to our lives and promises that it “could free people everywhere from tedious routine work and amplify our creative abilities”:

> *the potential to drive economic growth and productivity, boost health and wellbeing, improve public services, and increase security*. (Policy Paper 2024)

Who can resist a tool that frees him from laborious routine tasks, drives economic growth and increases productivity, improves public services and above all enhances security? These extraordinarily good intentions raise doubts about the hidden intentions and real agenda behind this industry.

This over-statement of benefits coincides with the risks associated with systems:

> But they [AI systems] could also further *concentrate unaccountable power into the hands of a few*, or be *maliciously used to undermine societal trust*, *erode public safety*, or *threaten international security*. (Policy Paper 2024)

Risks include the concentration of power in the hands of owners, malicious use and its effect on societal trust, erosion of public safety, and even a threat to international security. A rational decision-maker normally weighs benefits to risks in order to take a decision of yes or no. The question then is whether the benefits deserve to overlook the risks and go on adopting systems, or not. The contrast between benefits and risks cannot be simply ignored. How to boast economic growth with an ‘undermined societal trust' and people replaced by machines, in a society whose public safety is eroded, or in an insecure world? What kind of ‘boosting’ health and wellbeing, or ‘improving’ public services we can talk about? Ahmed talks about many health risks associated with over-indulgence with AI technology including addiction, stress, weak concentration and eyesight, backbone diseases, and even death (2024a). How can I understand the phrase 'threaten international security' and the opposite 'increase security’ at the same time? Moreover, the idea of 'unaccountable power' into ‘the hands of a few' refers to the control of a few owners over this industry, which is full of secrets about their ethics, intentions, goals, interests, agendas, etc.

The Policy Paper does not hide the experts’ concern that humans can lose control over advanced AI and its possible catastrophic consequences:

> Some experts are concerned that *humanity could lose control of advanced systems*, with *potentially catastrophic and permanent consequences*. (Policy paper 2024)

Ridiculously, the consequences are even irreversible, ‘permanent’. To lose control in this regard is like shooting one's self on the foot or committing suicide. The government admits that it is unable to make powerful systems safe now:

> *At present, our ability to develop powerful systems outpaces our ability to make them safe*. (Policy Paper 2024)

This implies that it cannot guarantee safe systems in the future. Therefore, insistence on developing something unsafe looks illogical, abnormal, and suicidal.

Anyway, to address these risks, including threats to international security, the mission of the AI Safety Institute as introduced in the Policy Paper is quite strange and needs analysis:

> Its mission is to *minimise surprise to the UK and humanity from rapid and unexpected advances in AI*. (Policy Paper 2024)

Advances are so rapid and so unexpected that the government needs to administer them to minimise ‘surprise’. The reference to ‘surprise’ denotes the huge concern about what is coming and reveals the government’s desire to pave the way to the acceptance of advances internationally. To this end, it aims to “develop the sociotechnical infrastructure needed to understand the risks of advanced AI and *enable its governance” and* “to move the discussion forward from the speculative and philosophical, further towards the scientific and empirical”. But how to move from philosophical to scientific discussions is not explained. The reference itself to philosophical and scientific discussions do not exceed sheer political rhetoric, i.e. no concrete, serious action plan or real contribution will be taken to face such dangerous challenges, and the government feels satisfied with and proud of such rhetoric as it mentions “This is our contribution to addressing a shared challenge posed to all of humanity”! The idea of a challenge, a danger, facing ‘all humanity’ stresses the strategy of fear, explained in the previous section, which attempts to mobilise the world around one globalised policy. It is all about accepting any future de facto situation created by the hazards and risks of systems.

Evaluating risks is necessary, if we suppose that all risks can be possibley evaluated at all, at least because "There may be fundamental limitations in the ability of evaluations to assess risks" and in other required capabilities (Policy Paper 2024). Suggested actions and solutions derive derision and look extremely disproportional to the present or potential challenges. The only action that has been taken so far is the insistence to adopt, develop and deploy it, to move forward, no backtrack.

### 5.3 World Governments & Bletchley Declaration

On 1 November 2023, 29 countries, including USA, UK, EU, and China, attended the AI Safety Summit in Bletchley Park, UK, and signed Bletchley Declaration. The Declaration starts with stressing the idea of 'enormous opportunities' AI can offer globally to human well-being, peace and prosperity as it:

> presents e*normous global opportunities*: it has *the potential to transform and enhance human wellbeing, peace and prosperity*.
> (Bletchley Declaration 2023)

It has the potential, the Declaration claims, to become a miracle for humans. The first sentence is, indeed, too good and general to be accepted as a logical giving for what follows. It is mouth-watering, anaesthetising nations and paving the way psychologically for the peoples to receive whatever said without resistance or even thinking. This goal manifests itself immediately in the second sentence:

> To realise this, we affirm that, *for the good of all*, *AI should be designed, developed, deployed, and used*, in a manner that is *safe*, in such a way as to be *human-centric*, *trustworthy* and *responsible*. (Bletchley Declaration 2023)

It repeats again 'for the good of all' so that no one may question such good intentions and consequent actions, i.e. the design, development, deployment and usage by 'all' since it is for the good of 'all'. This means the continuation of the same path no matter what the risks are! It is true that the Declaration uses the adjectives 'safe', 'human-centric', 'trustworthy' and 'responsible' but they are general and vague and no identified measures have been taken so far to make sure that they are performed in a 'safe', 'human-centric', 'trustworthy' or 'responsible' manner. On the contrary, the de facto situation now shows that it is 'unsafe' for its current and potential risks, 'machine-centric' seeking the replacement of humans by machines in many domains, 'untrustworthy' for the fake promises we get and for the secrecy engulfing the industry, including its true hidden agendas, and ‘irresponsible’ for the grave risks and problems it entails (cf. Ahmed 2024a).

The Declaration clearly aims to get the ‘full’ cooperation and support of every and each one on earth to ‘fully realise’ such void promises like promoting ‘inclusive’ economic growth where none is excluded, and protecting human rights and fundamental freedoms:

> We welcome the international community’s efforts so far to *cooperate* on AI to *promote inclusive economic growth*, *sustainable development* and *innovation*, to *protect human rights and fundamental freedoms*, and *to foster public trust and confidence in AI systems* to *fully realise their potential*. (Bletchley Declaration 2023)

The situation, conflicts and wars in the Middle East, for example, can easily reveal the naivety of the allegations of promoting fundamental human rights or economic growth. In this way, ‘full’ support means blind submission to agendas, the overt and covert ones, with full surrender. It equals the ‘full’ control over humanity. To this end, it targets ‘public trust and confidence’ through public anaesthesia. Note the reference to ‘innovation’ which implies the pursuance of development, regardless of consequences as I explained before, and the guarantee that all global research should be ‘controlled’ by USA, through what is called ‘global governance’ systems.

On the other hand, the risks indicated in the Declaration vary from violating human rights and safety to bias and privacy as follows:

> the protection of *human rights, transparency and explainability, fairness, accountability, regulation, safety, appropriate human oversight, ethics, bias mitigation, privacy and data protection* needs to be addressed. (Bletchley Declaration 2023)

The use of the words ‘transparency’, ‘regulation’ and ‘oversight’ brings to our minds immediately the great secrecy involved in this industry. In addition, the above-mentioned list of risks is not inclusive. The ‘surprise’, that is awaiting humanity due to ‘rapid and unexpected advances’ as raised by the UK government in its Policy Paper (2023), is similar to the Declaration’s ‘unforeseen risks’ in:

> We also note the potential for *unforeseen risks* stemming from the capability to *manipulate content or generate deceptive content*. (Bletchley Declaration 2023)

It admits that governments are incapable of controlling content; content can be manipulated. In fact, if countries, one wonders, are in such a no-man’s situation, why should they develop or deploy it before being capable of the control, then? Why the rush? Actually, ’substantial’ risks, that are 'existential' as the 'A Right to Warn' Letter (2024) describes them, are expected to arise due to usage, be it 'responsible' use or 'misuse':

> *Substantial risks* may arise from potential intentional misuse or unintended issues of control relating to alignment with human intent.

It says they cannot 'fully understand' potentials, therefore it becomes 'hard to predict' them!

The solution or action plan proposed in the Declaration does not exceed the act of 'encouraging':

> We *encourage* all *relevant* actors to provide *context-appropriate transparency and accountability* on their *plans to measure, monitor and mitigate potentially harmful capabilities and the associated effects*.. (Bletchley Declaration 2023)

But what are the specific steps to be taken to measure, monitor or mitigate 'harmful' risks, and how to implement them? Or how to guarantee 'transparency' or 'accountability' and to what extent? Would USA be held as accountable and transparent as the rest of the countries? Is USA really incapable of controlling AI as much as the others? These are valid questions that need full, clear answers, a matter which is hard at the moment due to the humongous size of secrets involved. So, general statements plus no mechanism for achieving them seem useless in this regard. And the only logical interpretation is that they hide a dangerous plan. Before I leave this part, two important words mentioned here have attracted my attention. First, why only 'relevant' actors are addressed here rather than all countries in the world as we expect in case of an international danger as such that threatens the international community? And who are those 'relevant' actors? Second, the use of 'appropriate' in 'context-appropriate transparency and accountability', in addition to being unidentified, opens the door for various interpretations that hinder true transparency and accountability according to each country' conditions and interpretation. In the end, neither the needed transparency nor accountability is attained.

Finally, the Declaration refers to an agenda to 'inform' action: to 'identify' and 'understand' frontier AI safety risks, and to build risk-based policies in the 29 signatory countries which implies 'increased transparency' by the actors developing frontier AI. Though risks are potentially grave and transparency is necessary, yet it is deeds, not words; detailed, specific actions, not general, vague actions.

## §6. Secret Industry & Translation

For thousands of years, man has recognised the need for and importance of translation for communication with others who speak a language different from his. He has used it to make commercial, political, diplomatic, or social relations, among others. He needed translation in peace and in war.

Colonising powers in the Middle Ages and up to modern times understood this fact, too, and used translation as a tool to help them usurp the treasures and knowledge of the colonised and further create a superior, fake image about themselves and a contrary, wrong and weak stereotyped image about the colonised peoples (Ahmed 2019, 2020).

The West aimed desperately to develop a machine that increases the volume of translations done from their own perspective, thus guarantee to pursue the images they want to create about themselves and others. Georges Artsrouni developed an automatic bilingual dictionary using paper tape in the 1930s. In 1954 the Georgetown–IBM made the first known public demonstration of a machine translation (MT) system. The rest is history in regard to the development of computers and the internet (cf. Ahmed 2022:329-31). Artificial intelligence has changed the rules of the game totally with easy, free access to machine translation systems available at the hands of millions of people around the globe. Though not 100% accurate, the quality of translation is still acceptable enough to be used, particularly for the huge amount of productivity and less human effort it entails.

If we argue that any AI system is based on the algorithms that operate the machine, it follows that algorithms reflect the ethics, knowledge and agendas of the very few people who control or write the algorithms; the same applies to datasets. In other words, if those people are biased, then the algorithms will be written in such a way that reflects the bias. If they have a certain agenda, algorithms will operate to fulfill that agenda, etc. If they want to create a fake stereotyped image, like those of the colonised or colonisers, about someone, something or some group, then algorithms are there. Simply, algorithms can shape the ‘product’ of any system in a manner consistent with the industry owners’ interests.

For example, I asked AI through the application “iask’ on 20 June 2025: What is Palestine?’ And it answered:

> Palestine is *a region in West Asia with a long and complex history*, *encompassing parts of modern Israel and the Palestinian territories of the Gaza Strip and the West Bank*. The term "Palestine" has been used, sometimes *controversially*, to refer to *this area for over three millennia*.. (iask 2025a)

While it answered when asked ‘What is Israel?’ as follows:

> Israel, officially the *State of Israel*, is *a country located in West Asia, situated at the eastern end of the Mediterranean Sea*. It shares borders with Lebanon to the north, Syria to the northeast, Jordan to the east, Egypt to the southwest, and the Mediterranean Sea to the west. Israel also *controls* the *Palestinian territories* of *the West Bank and the Gaza Strip*. (iask 2025b)

Comparing the two answers, we find out that this system refers to Palestine as just ‘a region’ in West Asia that consists of the West Bank and Gaza Strip ‘territories’, i.e. there is no reference that it is or was a state occupied by Israelis decades ago. It even alleges that those ‘Palestinian territories’ are ‘parts of modern Israel’. Meanwhile, it refers to Israel as a ‘State’, a ‘country’ with definite neighbours, Lebanon, Syria, Jordan and Egypt. Look at referring to the history of Palestine as long but ‘complex’ and to ‘Palestine’ as a ‘controversial term’ for ‘three millennia’ to establish in our minds the idea that it is a controversial issue rather than a flagrant aggression by an Israeli occupation, in a manner that deforms reality. Also note the description of Israel as ‘modern’ to stress a widely-circulated, fake image that it is a civilised, democratic state respecting human rights, etc. Indeed, the adjective should be interpreted, instead, in the light of its modern history since its establishment in 1948. Moreover, in such a small paragraph, AI uses the term ‘control’ in place of occupation to distract attention from occupying and devouring the whole Palestinian state by Israelis. That is just a simple example for bias against Palestinians and how it can reshape the awareness of people and the international public opinion.

We can imagine the effect of translating such biased thinking into various languages. MT systems support translation between hundreds of languages. For instance, Google Translate supports over 133 languages. Facebook can translate between any pair of 100 languages. Microsoft Translator makes available translation between 179 languages. Different low-resource languages have been added. In short, translation can be reached in systems without resort to humans_ I am talking here about accessibility to the product as quantity, not quality. In this case, translation acts as a dangerous soft power in the hands of owners (Ahmed 2019). It shapes minds towards one globalised viewpoint, the owners’, to achieve some hidden agendas.

AI problems such as bias can appear in source messages; in this case translation is biased as a result of the source’s bias. Or otherwise, bias can stem from algorithms or datasets in machine translation systems. Stories on social media platforms indicate that some Arab users approached

YouTube or Facebook, for example, to translate content regarding the recent Arab-Israeli conflict into or from Arabic and they discovered a clear bias against Arabs in MT. The point is that when such platforms realise that a bias is discovered, they can change it. Yet, this does not mean that bias is eradicated from the platform forever, for no one guarantees that bias would not be restored again at any time when things calm down. For instance, Google carried out 3,234 updates to its search algorithms in 2018 and more than 4,500 changes in 2020 (Ali and Yu 2021: 32-33).

In addition to circulating certain ideologies, values and ideas to enhance a particular viewpoint, AI can counteract and deliberately hide others. Reality is thus endangered and users become trapped in a bubble of misinformation and wrong knowledge. On the long run, this dim situation will get darker and more difficult to address if it pursues its current track.

Meanwhile, the idea of replacing the human translator by a machine stands as a challenge to human supervision of the translation process and product, a matter which not only challenges accuracy but also influences translation education, training and job market so severely that a potential over-dependence on machines may replace humans and threaten social fabric. Mureșan thinks that AI will inevitably influence future occupations in a complex way and issues such as automation, creation of new jobs, decision assistance, transformation of current occupations and professional adaptation are some of the important aspects to consider (2023:83). Therefore, there is an urgent need for the transparency of the industry in general. Ali and Yu assure that "the call for transparency rests on the need to look inside AI technology, in order to try to fully understand its logic and regulate its behaviour" (2021:5).

## §7. Implications

The analysis of data shows that all the investigated documents start with rhetoric hopes and wishes that no sane human may refuse, like "Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure" (Executive Order 2023). The UK argues that AI "presents enormous global opportunities: it has the potential to transform and enhance human wellbeing, peace and prosperity" (Bletchley Declaration 2023). This seems a diplomatic or a political strategy of international relations that announces appealing goals that may hide other secret harmful ones, which in turn flouts and violates the Gircean maxims of communication. Grice (1975) identifies four rules to achieve effective communication: be informative (maxim of quantity), be true with adequate evidence (maxim of quality), be relevant (maxim of relevance) and be perspicuous (maxim of manner). If we follow the proverb 'Deeds, not words", this means simply no importance should be given to words. In other words, maybe the rules of the game have changed. Such a strategy aims to mobilise a unified international public opinion that supports and surrenders totally to any actions taken in this concern.

The analysis also shows that documents refer to risks that can reach the levels of threatening humans, e.g. "irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security" (Executive Order 2023) up to 'human extinction' (A Right to Warn 2024). However, the documents stress the importance and even necessity of continuing research and going in the same direction. This is a clear contradiction and an unexplainable illogic, for how can they target human welfare and suggest a consequently probable human destruction in the same time? My interpretation is that such companies and governments warn humans so as not to be legally sued, exactly like when they ask users to 'accept' terms if they want to use or download an application. So, it is users' choice whatever the consequences. In other words, users are also responsible for all potential harms and risks exactly as owners and builders.

A severe contradiction, thus, between wishful aspirations, in this case they do not exceed lies, and catastrophic risks become evident in the data. For example, "Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure" and in the same paragraph "At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security" (Executive Order 2023). Regardless of the two vague, unidentified adjectives 'responsible' and 'irresponsible' of AI, no concrete actions have been taken to avoid such risks. Long lists of proposed actions have not reverted the dangerous course of actions. However, there is an unjustified push and rush towards continuing the path despite the grave consequences. The contradiction in the data, in my opinion, can be interpreted only in the light of some hypnosis and anaesthesia of the victims (in this case users), a matter which results in total control at best, and their entire destruction at worst, though both best and worst scenarios imply the destruction of the victims, directly or indirectly. In all the cases, the danger facing humanity is greater than maintaining the secrecy of the industry for reasons like competition between companies or economic incentives.

A deliberate strategy of fear can be discerned. First, AI builders and experts themselves claim that they cannot predict or control the future potentials. Second, there is neither enough transparency nor accountability. The secrets of the industry increase the fear from what is coming; the image is so vague and unclear to be able to make conceptions. Third, the risks we have already experienced are quite frightening, take for example the use of AI by Israel in its wars against neighbouring countries (Iran, Palestine, Iraq, Syria, Lebanon, Yemen, Jordan, inter alia).  Fourth, it is hard to differentiate between industry lies, facts, or exaggerations since the globalised propaganda is overwhelming solely the international scene and other contrary voices are fought. Therefore, the discourse on the 'inevitability of change' to AI technology is illusionary and unacceptable and represents surrender and submission. The end results are users' paralysis of minds, surrender to the project (overt and covert agendas), and being controlled!

In this context, translation becomes a dangerous soft power tool to help fulfill certain agendas, especially the control one. AI-generated translation tools are easily used, accessible, free in most cases, and appealing in regard to productivity, globally. The almost two hundred language pairs available today, in addition to those low-source languages added continuously guarantee that the translation of any messages reach out everyone on the globe. Hence, translation can spread certain values and reshape users' information, knowledge and awareness. It is capable of creating particular realities. Even if we overlook the translation problems associated with AI-generated translation tools, like biases, errors, or inaccuracies, translation is still an ideological tool (Ahmed 2014).

## Conclusion

From the beginning, this study has made it clear that it aims to investigate the secret industry of Artificial Intelligence and the politics of technology and to explore the role of translation. It has raised three questions about the nature of the secret side in this industry, the standpoints of big companies and governments, and the role of translation in this industry. To answer, it has designed a research methodology based basically on content analysis and interpretation and approached data through a multi-disciplinary perspective deriving its tenets from world politics, computer engineering and translation studies. It was not my intention to call for stop the use, the development, or the deployment of AI systems. It has challenged our understanding of this technology and its secret dimensions as well as the role of translation in this regard, with a view to deconstruct the scene which is full of harms and grave existential risks and maybe reconstruct again but without dangerous risks. The governments of the world cannot "fully understand" potentials and confess that they are "hard to predict" them (Bletchley 2023). Anthropic (2023) discloses that frontier AI progress should be slowed down to be "more manageable, taking place over centuries rather than years". The AI Safety Institute in the UK reflects experts' fears from "potentially catastrophic and permanent consequences" and stresses the need to "minimise the surprise" to humanity from "rapid and unexpected advances".

The study has come to the conclusion that AI global discourse follows a constant strategy that starts with exaggerated propaganda full of rhetoric hopes and wishes, then mentions risks which could threaten human existence, and ends with some measures, which in fact do not rise to the level of threats and do not resolve them. A simple comparison between benefits and potential risks shows a clear contradiction in the messages and further raises doubts about the unjustified rush for using, developing and deploying AI applications globally. Another strategy of fear was utilised in this discourse to enhance AI domination and control and to help in the process of hypnotising and anesthetising humanity.

The lack of transparency and accountability has become unacceptable. The secrets associated with algorithms and datasets in many cases should be revealed. All the real and potential benefits and risks must be thoroughly shared with experts as well as the public since the latter is a user and is and will be affected. The recent Israeli wars have disclosed a change in the rules of the game. A drone targeted Yahya Al-Sinwar, Hamas Chief, while he was sitting on a sofa in a destroyed building in Gaza, see Figure 6. 'Questions like 'Who owns this industry?' become not only valid but also necessary for preserving human life on the planet and unveiling AI industry owners' real hidden motives, interests, ethics, and agendas. As I explained before, I do not argue against AI just for the sake of it. On the contrary, we should have a full understanding of the benefits to make use of and maximise and the risks to avoid or address. This understanding requires transparency. Nations should develop their own systems that serve their interests and all humanity's. The implications are serious for the industry generally, for countries'

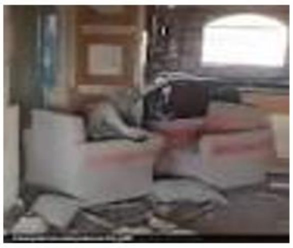

**Figure 6**: Drone Footage of Yahya Sinwar's Last Moments https://www.ndtv.com/world-news/drone-footage-of-hamas-chief-yahya-sinwars-last-moments-released-by-israel-6815660

national policies and security, for the international peace and security, for end-users who often adopt technological advances without realising the risks, for translation industry, and indeed for all humanity. Therefore, further research is urgently needed in these under-investigated areas.

## References

Ahmed, Safa'a Ahmed. (2024 a). The Politics of Integrating Artificial Intelligence into Higher Education: Benefits < > Risks. *Occasional Papers in the Development of English Education*, CDELT, Vol. 88 (October), pp. 457-90.

--- . (2024b). Globalisation vs. Islamic Universality and the Politics of Translation. *Occasional Papers in the Development of English Education*, CDELT, Vol. 86 (April), 203-247. DOI: 10.21608/OPDE.2024.362821. https://opde.journals.ekb.eg/article_362821_8b4cfc04bbb0a4ab01f8ef67b412f64a.pdf

---. (2022). Technology and Artificial Intelligence in Simultaneous Interpreting: A Multidisciplinary Approach. *Occasional Papers in the Development of English Education*, CDELT, 78(1) April, 325-353. DOI: 10.21608/OPDE.2022.249945. ISSN 1110- 2721. https://opde.journals.ekb.eg/article_249945.html

--- . (2020). Translation and the Western Ideology of Domination during the 11th -15th Centuries: An Arab Perspective. *Philology*, 37(74), June, 7-34. DOI: 10.21608/gsal.2020.131414. ISSN: 1678-4242. Online ISSN: 2812-4952. https://alsun.journals.ekb.eg/article_131414.html?lang=en

--- . (2019). 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. DOI: 10.21608/opde.2019.132682.
https://opde.journals.ekb.eg/article_132682.html

… . (2014). Ideological Translation and Mass Communication: A Modernisation or a Conflict Enterprise? A Case Study of Al-Jazeera vs. Al-Arabiya. *English Language and Literature Studies, ELLS*, XI (1), (December), 181-249.

Alì, Gabriele Spina; Yu, Ronald. (2021). Artificial Intelligence between Transparency and Secrecy: From the EC Whitepaper to the AIA and Beyond. *European Journal of Law and Technology*, Vol 12 No.3, 1-25.

Allen, R. John; Massolo, Giampiero. (2019). Introduction. In *The Global Race for Technology Superiority*, by Fabio Rugge, pp.7-12.

Anthropic. (2023). Core Views on AI Safety: When, Why, What, and How. Updated 24 February 2023. https://www.anthropic.com/news/core-views-on-ai-safety

A Right to Warn about Advanced Artificial Intelligence [public letter]. (4 June 2024). https://righttowarn.ai/

Armellini, Alvise (November 25, 2024).G7 seeks unity on ICC arrest warrant for Netanyahu. *The Reuters Daily Briefing newsletter*, Reuters. https://www.reuters.com/world/g7-seeks-unity-icc-arrest-warrant-netanyahu-2024-11-25/

Bassnett, Susan and Harish Trivedi. (1999). Introduction: of Colonies, Cannibals and Vernaculars. In Susan Bassnett and Harish Trivedi. In Susan Bassnett and Harish Trivedi (eds.), *Post-Colonial Translation*, London and New York: Routledge, 1-18.

Bletchley Declaration [Policy Paper]. (1 November 2023). AI Safety Summit, 1-2 November 2023, published by Sunak Conservative government. https://www.gov.uk/government/ publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

Boucher, Philip. (2020). *Artificial intelligence: How does it work, why does it matter, and what can we do about it?* European Parliamentary Research Service (EPRS).

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. (2023). The White House, 30 OCTOBER 2023. https://www.whitehouse.gov/briefing-room/presidentialactions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Google DeepMind. (2023). AI Safety Summit: An Update on Our Approach to Safety and Responsibility [Policy paper]. 27 OCTOBER 2023**.** HTTP://DEEPMIND.GOOGLE/PUBLIC-POLICY/AI-SUMMIT-POLICIES/

Grant, Alex. (2023). *The Dark Side of AI: Geoffrey Hinton's Warning*. Independently published.

Is Israel Acting like the ICC is 'only for Africa '. *Al-Jazeera*. https://www.aljazeera.com/news/2024/5/29/is-israel-acting-like-the-icc-is-only-for-africa-and-thugs-like-putin

Grice, Paul (1975). "Logic and Conversation." Pp. 41–58 in *Syntax and Semantics 3: Speech Acts*, edited by P. Cole and J. J. Morgan. New York, NY: Academic Press.

Gurevich, Yuri. (2012). What is an Algorithms? [conference proceedings]. SOFSEM 2012: Theory and Practice of Computer Science - 38th Conference on *Current Trends in Theory and Practice of Computer Science*, Špindlerův Mlýn, Czech Republic, January 21-27, 2012. Technical Report MSR-TR-2011-116, July 2011, Microsoft.

iask. (2025a). What is Palestine? Retrieved on 20 June 2025 from https://iask.ai/q/What-is-Palestine-kivkgv0

iask. (2025b). What is Israel? Retrieved on 20 June 2025 from https://iask.ai/q/What-is-Israel-od7h7bo

Kleinman, Zoe; Vallance, Chris. (2 May 2023). AI 'Godfather' Geoffrey Hinton Warns of Dangers as He Quits Google. *BBC News*.

Knight, Will. (2017). The Dark Secret at the Heart of AI. MIT Technology Review, 11 April 2017.

Mureșan, Mircea. (2023). Impact of Artificial Intelligence on Education [RAIS Conference Proceedings]. *Research Association for Interdisciplinary Studies RAIS*, June 8-9, 81-85.

OpenAI. (2023). Planning for AGI and Beyond. (2023). Updated 24 February 2023. https://openai.com/index/planning-for-agi-and-beyond/

Policy Paper Introducing the AI Safety Institute. (2024). A policy paper presented to Parliament by the Secretary of State for Science, Innovation and Technology in November 2023 and updated by UK government on 17 January 2024. https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute.

Rizzo, Gabriele (2019). Disruptive Technologies in Military Affairs. In *The Global Race for Technology Superiority* by Fabio Rugge (ed.), pp.55-92.

Robinson, Douglas. (2002). *Western Translation History from Herodotus to Nietzsche* (2nd edn). London and New York: Routledge.

Rugge, Fabio (ed.). (2019). *The Global Race for Technology Superiority: Discover the Security Implications*. Introduction by John R. Allen and Giampiero Massolo. Milano, Italy: Brookings, ISPI.

Rugge, Fabio. (2019). Disruptive Technology and International Stability. In *The Global Race for Technology Superiority* by Fabio Rugge (ed.), pp.13-54.

Stankovich, Mariam; Behrens, Erica and Burchell, Julia. Toward AI Meaningful Transparency and Accountability of AI Algorithms in Public Service Delivery. (2023). *CDA Insights*, DAI's Center for Digital Acceleration

Taylor, Josh; Hern, Alex. (2023). 'God Father of AI' Geoffrey Hinton Quits Google and Warns over the Dangers of Misinformation. *The Guardian*, 2 May 2023.

Tymoczko, Maria. (2010). *Enlarging Translation, Empowering Translators*.
London and New York: Routledge.