JUNE 2024

COLUMBIA CLIMATE SCHOOL
CENTER FOR SUSTAINABLE DEVELOPMENT

FII INSTITUTE
Future Investment Initiative Institute
Impact on Humanity

ARTIFICIAL INTELLIGENCE & THE FUTURE OF WORK

Closing the AI Gap between High-Income Countries and Emerging Markets and Developing Economies

Prepared by

Yanis Ben Amor

Executive Director, Center for Sustainable Development, Columbia University

A new initiative launched by Columbia University and the Future Investment Initiative Institute proposes to focus on the "AI & the Future of Work" in the context of the growing importance and impact of Artificial Intelligence (AI), and Large Language Models (LLMS) such as ChatGPT, among other topics. Our project is a timely and important initiative that seeks to explore the potential of emerging technologies to transform the way we learn, teach and work while making sure it does not exacerbate the digital divide between the "Global North" and the "Global South".

Our project sets out to work with all relevant stakeholders to imagine how AI will complement and disrupt employment in the future, and create all the relevant curricula that allow universities to train the next generation of AI-enhanced workers who will not be obsolete on the job market in the next 10-25 years. The integration of AI and LLMs into the education system is an area of growing interest and innovation, as it promises to revolutionize the traditional classroom model and create new opportunities for learners and educators alike.

This first brief explores the structural differences between High Incomes Countries and Emerging and Developing Economies, which will affect universities on either side of the divide not only in their ability to use new AI tools, but also to participate in their development.

I. *Imagining the professions of the future*

In 1993, the World Wide Web became available to the broader public and led to dramatic changes in communication, education, collaboration and access to amenities, to list just a few. The Year 2023 will be remembered as the "*Year 1*" of the Artificial Intelligence (AI) revolution that is taking Humanity by storm, and dividing it into two camps: techno-utopians and techno-pessimists. Both sides agree that within the next few years, many of the current professions will have radically changed with the addition or inclusion of AI. AI tools will significantly increase productivity in certain sectors, and job performance will be improved drastically for workers who are able to operate with AI tools as co-pilots. This will lead to unprecedented productivity gains. Yet techno-pessimists imagine a future where many, if not most, jobs will be taken over by AI, leading to massive layoffs, exacerbating inequalities even further than now, with significantly greater gaps between the haves and have-nots. Both visions are conceptually possible, as demonstrated by economic theory. The outcome – Utopian versus Dystopian – will depend on public policies, social ethics, and creativity in facing the new challenges and opportunities.

At this early stage of what could be a seismic change in the way mankind will experience work within one decade, and certainly within a quarter century, there are still opportunities to have a major impact on the direction that society will take. One important step is to rapidly involve universities and schools for the integration of AI in curricula and teaching methods. The wrong approach would be to expect that AI will reach a "final stage" of development (in a year, in five

years) and therefore wait for that stage before embarking on the mission to rethink Education and educational approaches and tools so they can be adapted to the "ultimate state" AI will take. Experts predict that this will likely never happen as, by design, AI constantly learns from iterative processes and in some instances, has the ability to even improve itself independently of human involvement. There will not be a fixed goalpost that can be reached once AI reaches its final stage, because the goalpost will constantly be moving.

It is therefore critical to lay down, in the early years of the AI revolution, the foundations of an initiative that will work with key stakeholders, from private and public sector to government and education institutions, to imagine how various professions are likely to evolve by 2050 (based on the current understanding of what AI can do and how it can evolve in the coming 25 years), and then to design appropriate curricula for each field/profession in order to update the way universities currently prepare their students for future careers, and importantly, prevent sending on to the job market future workers who will soon be obsolete, and who will urgently need significant retraining to take up AI-related skills that they were not prepared for. Because of the constantly evolving nature of AI, such curricula will need continuous updating over the years, and methods to help students to prepare for enormous flux in their future occupations. The new curricula should ensure that students always have access to the latest developments of the technology. Similarly, professors and teachers will also need to stay up to date with ongoing rapid changes in AI, and allow for frequent updates to curricula and teaching methods.

To prevent a widening of the digital gap between the "Global North" and the "Global South", an initiative focusing on universities to prepare for AI & the Future of work will need to include partners and stakeholders from Emerging and Developing Economies (EMDEs) and participants from international organizations (UN, IMF, World Bank, regional development banks) that work with the EMDEs. First, strong representation of the EMDEs will ensure that the specific geographical, historical, economic and cultural contexts of each geographic zone are taken into account when designing new teaching solutions. Such global participation will also help to ensure that AI tools are effective globally, not merely in the regions in which they were first designed.

By operating in an inclusive manner, by inviting all relevant partners and stakeholders to the drawing board, and operating within a worldwide network of partners and universities, such an initiative will prevent a detrimental shortcoming of being "sector-specific" or "context-specific" and will have cultural adequacy at the heart of its processes. The UN Sustainable Development Solutions Network (SDSN), with its 2,000 partner institutions, will provide a global platform for the program from the start.

The potential divide between universities in High Income Countries (HICs) and the EMDEs will require a global concerted effort to bridge the AI and Digital Divides. It is crucial to prioritize capacity-building initiatives, knowledge sharing, and international collaborations that empower universities in EMDEs. Partnerships between universities, governments, and international

organizations will facilitate access to resources, promote knowledge exchange, and support skill development in AI-related fields. By fostering inclusive AI ecosystems and providing support to universities in the EMDEs, the program will help to ensure that the benefits of AI are accessible to all, and that the development of new curricula and teaching methods are accessible to everyone and relevant for everyone.

Unfortunately, on our current trajectory, AI, and existing Large Language Models such as ChatGPT, have the potential to exacerbate the existing divide between teaching methods and curricula at universities in HICs and the EMDEs, for two main reasons: first, differences in resources; and second, differences inherent in training models for Artificial Intelligence and the lack of strong models in most languages other than in English.

This brief report explores these issues, most of which will be addressed by the new initiative.

II. Differences in Resources and Infrastructure.

While AI technologies offer numerous opportunities for enhancing education, research, and innovation, their widespread adoption and effective utilization require substantial resources, infrastructure, and expertise. This discrepancy in resources between HICs and EMDEs could rapidly widen the AI gap.

Computational Power

The technology that drives ChatGPT relies on thousands of specialized computer chips. While it is impossible to accurately quantify AI's energy use (because companies like OpenAI - parent company of ChatGPT - do not disclose how many specialized chips they need to run their software), it was estimated that ChatGPT is probably using more than half a million kilowatt-hours of electricity to respond to its 200 million requests a day (compared to 29 kilowatt-hours daily for the average US household). It is predicted that by 2027, AI servers could use between 85 and 134 terawatt hours annually, which is similar to what Argentina, the Netherlands and Sweden each use in a year, or about 0.5 percent of the world's current electricity use. The EMDEs are bound to face a major gap in investments in AI given these staggering costs. How the EMDEs should use their scarce resources to develop and use AI tools will be one important area of inquiry on the project.

Additionally, the availability of computational resources plays a crucial role. Training AI models like ChatGPT requires vast computational power and storage capabilities. A recent report describes computational power as "a predominant factor driving the industry today" in the private sector and highlighting that companies have spent "more than 80% of their total capital on computational power resources"[1] An International Data Corporation (IDC) white paper titled

[1] Guido Appenzeller, Matt Bornstein, and Martin Casado, *Navigating the High Cost of AI Compute*, Andreessen Horowitz, April 27, 2023, https://a16z.com/navigating-the-high-cost-of-ai-compute

"2021-2022 Global Computing Index Assessment" examined the correlation between a country's Gross Domestic Product (GDP), the digital economy, and computing power. On average, for every 1-point increase in the Computing Index of the 15 key countries under scrutiny, the national digital economy and GDP increase by 3.5‰ and 1.8‰, respectively. When the Computing Index surpasses 40 points, the impact on GDP increases by 1.5 times, and this impact grows to 3 times when the Computing Index exceeds 60. The Assessment results show that the United States and China rank as the top two countries, with 77 points and 70 points respectively. The scores of the rising countries are between 40 and 55 points, including Japan, Germany, the United Kingdom, France. Those with scores lower than 40 points are startup countries, including India, South Africa, Russia and Brazil. While not on the list of countries under scrutiny, it is intuitive that most EMDEs would not fare well on the Global Computing Assessment Index, therefore stunting their opportunities for independent development of new AI tools and models both locally relevant and context-appropriate.

Similar to the capacities in the private sector, Universities in HICs often possess powerful computing infrastructure and can afford high-performance hardware or cloud-based services for AI training. Universities in HICs allocate substantial funding in supercomputers, clusters and high-performance computing (HPC) facilities. They often have state-of-the art data centers equipped with powerful servers, Graphics Processing Units (GPUs – key to optimizing deep learning) and specialized hardware for data analysis and machine learning. They also often collaborate with industry and government agencies, which facilitates access to shared resources and expertise. In contrast, universities in the EMDEs will generally have much less computational capacity, hindering their capacity to train and deploy advanced AI models, and therefore forcing them to rely on models developed in other countries. Most universities in the EMDEs operate with tightly constrained budgets, making it very difficult to invest in high-end computing infrastructure. Limited Internet bandwidth can also be a major barrier, and outdated servers and limited access to GPUs hinder computational capabilities. While there are very talented researchers in the EMDEs, there may be a shortage of experts in managing HPC systems. International collaborations allow researchers in the EMDEs to access remote HPC resources and cloud services (such as Amazon Web Services or Google Cloud) democratize access to computational power.

Availability of Data

Another primary factor contributing to the divide is the availability of data. AI models like ChatGPT rely on large amounts of diverse and high-quality data for training, which have been fed into their algorithms. Developed countries, with their advanced technological infrastructure and access to abundant data sources, are in general more likely to possess the necessary data for training sophisticated AI models. In contrast, most developers and universities in EMDEs may have limited access to comprehensive datasets, inhibiting their ability to train AI models effectively. Less volume and less accurate training data will generally lead to less accurate models.

The table below shows the number of Google search results numbers (equivalent to a number of pages available for a specific topic) in the United States (US), United Kingdom (UK), Brazil, Kenya and Tunisia. This illustrates the differences in training data availability relative to specific countries:

	US	UK	BRAZIL	KENYA	TUNISIA
HEALTH	19,350,000,000	8,180,000,000	1,100,000,000	627,000,000	287,000,000
EDUCATION	16,380,000,000	3,710,000,000	1,190,000,000	1,070,000,000	202,000,000
AI	8,980,000,000	2,970,000,000	875,000,000	325,000,000	138,000,000
DEVELOPMENT	7,940,000,000	3,930,000,000	842,000,000	523,000,000	173,000,000

Table 1. *Number of Google search results number for specific terms in the US, UK, Brazil, Kenya and Tunisia.*

Based on this table, researchers in Tunisia aiming to develop AI models using data published about Education in Tunisia (in the English language) have access to approximately 80 times less web pages (i.e., data) than researchers in the US aiming to develop AI models using data published about Education in the US. This also has an impact on the quality of the data as not all sources on the web are of equivalent quality.

Expertise and Talent Readiness

Another critical aspect is the availability of expertise and talent. Developing and implementing AI systems require highly skilled professionals with expertise in areas such as machine learning, data science, and AI ethics. Developed countries tend to have a more established ecosystem for AI research and a larger pool of skilled professionals. They often attract top talent, have well-funded research programs, and offer extensive opportunities for collaboration. In contrast, universities in developing countries may face challenges in attracting and retaining AI experts, limiting their ability to build and deploy sophisticated AI systems.

A report by the Harvard Business Review[2] in 2019 ranked 60 countries in four categories relative to tech skills: Cutting-Edge (rank 1-15), Competitive (rank 16-30), Emerging (rank 31-45), and Lagging (rank 46-60) It showed that Argentina, the Czech Republic and Austria topped the list in tech skills, while Kenya, Pakistan and Nigeria placed in the bottom three. The report found stark skill contrasts between developed and developing markets, with North America and Europe far outpacing Latin America, Asia, and the Middle East, on average. Interestingly, the U.S. failed to reach the top quartile of countries. In 2019, most of the world lived in countries that were lagging in critical tech skills. Countries with developing economies — and with less to invest in

[2] https://hbr.org/2019/05/ranking-countries-and-industries-by-tech-data-and-business-skills

education — saw the largest skill deficiencies, with 90% ranking in the lagging or emerging categories.

Back in 2018, the World Economic Forum already predicted that by 2022, AI and machine learning skills would be among the world's top 10 most in-demand skills[3]. Considering the issues highlighted in this brief that LMICs face regarding access to computational power, data or expertise, it corroborates the findings of a recent International Monetary Fund (IMF) report[4] that found that 40% of jobs in emerging markets could see negative impact from AI because of a digital divide brought by AI.

Financial Resources

Financial constraints summarize the divide between HICs and LMICs. As described in the previous sections, AI research and development require significant investments in infrastructure, hardware, software, and ongoing maintenance. Developed countries with greater financial resources can allocate substantial funds to AI initiatives in universities, enabling them to pursue cutting-edge research, establish specialized AI centers, and offer competitive salaries to AI professionals. In contrast, universities in developing countries may struggle to secure sufficient funding for AI-related projects, limiting their ability to participate fully in the AI revolution.

III. *Differences inherent to current AI training models*

The training of ChatGPT's algorithm using mostly English language datasets poses significant challenges for speakers of other languages, including Chinese, Spanish and Arabic. While ChatGPT's capabilities in generating human-like responses and providing information are impressive, its effectiveness and accuracy was shown to be more limited when interacting with non-English speakers[5].

The availability of relevant data for training ChatGPT in languages other than English may be limited compared to English. AI models like ChatGPT require large volumes of high-quality, diverse data for training to ensure optimal performance. While there is a wealth of English-language data available on the internet, the same cannot be said for all languages, including Chinese and Arabic (estimated to be spoken by 1.35 billion and 422 million people worldwide respectively).

[3] https://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

[4] https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379

[5] https://www.nature.com/articles/d41586-023-00680-3

Language	Share of websites featuring the language
English	52.1%
Spanish	5.5%
German	4.8%
Russian	4.5%
Japanese	4.3%
French	4.3%
Portuguese	3.1%
Italian	2.3%
Chinese	1.3%
Arabic	0.6%

Table 3. *Share of websites featuring content in each language, compared with the share of the global population that speaks each language (January 2024).*
Adapted from Statista: https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/#:~:text=Common%20languages%20used%20for%20web%20content%202024%2C%20by%20share%20of%20websites&text=As%20of%20January%202024%2C%20English,language%20followed%2C%20with%204.3%20percent.

The scarcity of Arabic or Chinese training data can hinder the model's ability to accurately understand and generate responses in these languages (not to mention the diversity of dialects related to these two languages).

To illustrate this point, let us reuse the methodology applied to generate Table 1. The results of Table 1 reported searches done for specific key terms in English for the US, UK, Brazil, Kenya and Tunisia. But for Brazil and Tunisia for example, English is not a national language. Search results in other languages are expected to be vastly different.

Taking the example of Tunisia, with its two languages (Arabic and French), Table 2 reviews the same search terms again, and comparing results for English, Arabic and French.

ENGLISH	TUNISIA	ARABIC	تونس	FRENCH	TUNISIE
HEALTH	287,000,000	صحة	40,400,000	SANTÉ	48,300,000
EDUCATION	202,000,000	تعليم	35,400,000	ÉDUCATION	42,500,000
AI	138,000,000	الذكاء الاصطناعي	5,640,000	INTELLIGENCE ARTIFICIELLE	2,220,000
DEVELOPMENT	173,000,00	تطوير	29,400,000	DÉVELOPMENT	29,600,000

Table 2. *Number of Google search results number for specific terms in Tunisia when using English, Arabic or French terms.*

It is interesting, but not surprising, to note that the number of search results for all key words were systematically higher in the English language, than in either of the two official spoken languages of Tunisia. This in turn will have a direct impact on the quality of the training materials for less represented languages such as Arabic and for specific context of LMICs, such as Tunisia.

This additionally raises another issue: what are the sources of these pages written in English about “Health” or “Education” in the context of Tunisia, when English is not a locally spoken language? These are likely written by organizations which are not based in the country, and can therefore lack context around the topic reported on. They are also likely to be the product of international research organizations or international Non-Governmental Organizations (NGOs) doing work in the country and reporting their findings in English on their webpages. This therefore might miss the perspective of the private sector, or local government organizations for example, possibly leading to one-sided, often non locally contextualized, perspectives.

These issues related to lack of data from diversified languages in turn lead to issues in language understanding and context. ChatGPT's current training models mean it may struggle to comprehend nuances, idiomatic expressions, cultural references, and specific linguistic structures present in other languages such as Arabic or Chinese. This limitation can lead to misinterpretations, incorrect responses, or a lack of understanding when ChatGPT interacts with Arabic or Chinese-speaking users, or other non-English speakers.

Another challenge is the potential bias in the training data. AI models like ChatGPT learn from the data they are trained on, and if those data contain biases, the model will likely exhibit biases as well. Since ChatGPT's training has primarily been focused on English, biases present in English-language data can carry over into its responses. This bias can be particularly problematic when interacting with Arabic or Chinese-speaking users, as it may perpetuate cultural, social, or linguistic biases that are specific to English-speaking contexts. This is often misunderstood as platforms being purposely biased to the Western world, in reality, the output of platforms such as ChatGPT is only a reflection of pre-existing opinions and facts online in English. Developing a strategy to include Arabic or Chinese data will ensure the Arabic and Chinese-speaking world’s perspective is captured within the never-ending dataset being fed into AI-powered platforms.

Conclusions

The rapid spread and advancement of Artificial intelligence is shaping up to bring revolutionary progress in almost every field, just as the arrival of the Internet and the World Wide Web did in the early 90s. But this progress is likely to be high uneven across and within societies. One possible outcome is that the High-income countries, with higher levels of funding, infrastructure, and access to data and talents will reap the greater benefits and enjoy tremendous improvements in productivity, overall job performance and increased Gross Domestic Product, widening the gap with the Emerging and Developing Economies, which are burdened by a relative lack of key resources. Yet AI could also lead to "leapfrogging" by the EMDEs, making their progress and "catching-up" even faster than it would otherwise be. In all societies, the rapid changes ahead will also shift the income distribution and introduce new inequalities. This digital divide brought by Artificial Intelligence will could lead to negative impacts for a large part of the workforce.

Through a network of collaborating universities and partners worldwide, particularly in emerging and developing countries, "AI & the Future of Work" holds great promise for transforming the way we learn, teach, work, and use public policies. This project aims to explore the potential of these emerging technologies and their implications for education and employment, and to develop new strategies, tools, and policies that can help learners, educators, and governments to adapt to the rapidly changing landscape without furthering the divides between the rich and poor within our societies and between nations. The "AI & the Future of Work" initiative will promote a worldwide intensive analysis of the role of AI in key fields such as education, health, ethics and employment, and then directly assist universities in selected emerging and developing countries to update their curricula and education strategies to include new tools and syllabi to ensure that the EMDEs can be leaders in the new AI-driven age.

Acronyms

AI: Artificial Intelligence
EDME: Emerging and Developing Economies
GDP: Gross Domestic Product
GPU: Graphics Processing Units
HIC: High Income Country
HPC: High Performance Computing
IDC: International Data Corporation
IMF: International Monetary Fund
LLM: Large Language Model
NGO: Non-Governmental Organization
UK: United Kingdom
US: United States