Artificial Intelligence for Climate Change:

Design and Legitimacy Concerns for Policy Advice and Public Administration

Sunil Tankha

Erasmus University Rotterdam

tankha@iss.nl

*****DRAFT: NOT FOR CIRCULATION OR CITATION*****

1. INTRODUCTION

ChatGPT's release to the public in September 2022 immediately sparked various prospects and concerns that had hitherto been discussed more in abstract and theoretical terms than in practical and immediate ones. While some embraced it with the enthusiasm of a toddler receiving a new toy which could be prompted to create outputs of such genius as a haiku about pandas, others warned about the potential disasters lurking within this powerful technology. Artificial Intelligence (AI) became a buzzword, dominating countless discussions and conversations. Though language purists would insist that what was released was neither artificial nor intelligent, AI is undeniably a powerful computing tool with the potential to drive profound transformation (or disruptive change as those in the business like to hype it) in various domains. Indeed, given the flurry of public concerns shortly after ChatGPT was released, it seems that the sophistication and potential of the technology caught even the insiders by surprise.

This paper will enumerate and assess the near-term capability of AI to contribute to addressing the challenge of climate change adaptation in developing countries and to catalog several policy concerns and potential approaches to these concerns. Drawing on two decades of experience in research and teaching policy development and public administration, this paper identifies a promising use-case for AI in less developed countries (LDCs): to make up for the capacity deficit that the LDCs face, especially at the subnational and local levels, in tackling the challenges of climate change adaptation and climate resilience.

Due to the prevalent misconceptions about AI outside of the tech field, this paper begins with a brief discussion of the technologies and platforms underlying what is commonly referred to as AI, catering to a non-technical readership. Subsequently, the paper presents the use-case argument for AI based on capacity development and administrative challenges (Section 3). Sections 4 and 5 discuss the use cases for generative AI in terms of prediction and prescription while section 6 delves into the technical and policy challenges associated with the use of AI. Section 7 concludes with some observations on next steps for research, policy and technological development.

2. WHAT IS ARTIFICIAL INTELLIGENCE?

As an idea, AI has been around for around seventy years. John McCarthy (1956) was the first to use the term but already Alan Turing (1950) had set the ball rolling with a paper on "Computing Machinery & Intelligence" in which he proposed that the test of an intelligent machine would be whether it could fool a human into believing it was interacting with another human instead of with a machine. Since then, the goal of AI has been conceived as to create machines that mimic human intelligence in terms of learning, reasoning, problem-solving, perception and decision-making and to eventually perform tasks that normally have required human intelligence and to do so more quickly, more accurately and at a larger scale than humans can. They are, as McCarthy (1969) said, "programs written to solve a class of problems which give humans intellectual difficulty."

Playing chess was one of the first applications of AI programming. Scientists in the Soviet Union and the United States developed the first programs already in the 1950s. These programs were rudimentary and easily defeated by humans, but one group of experts predicted that computer programs would defeat the best chess players within 10 years, i.e., by 1967 (Simon & Newell 1958). It would, however, take a further three decades and only in 1997 did a computer program—IBM's Deep Blue—finally defeat the then reigning champion Garry Kasparov in a six-game series played under standard chess tournament rules (though Kasparov did later accuse the programmers of cheating). Given the slow pace of development, even experts thought that machines would be unable to overcome humans in Go, a more mathematically challenging game, but Google's DeepMind did that in 2017. Moore's law on the exponential increase in computer processing power was largely responsible for increasing machine capabilities in this sphere. It was not intelligence but brute computing force which won the contest. The computer that beat Kasparov could evaluate 200 million moves per second, but it required constant human intervention and tweaking between games. A contemporary observer noted that in the six games, in only one specific move in game 2 did the machine play like a human (Krauthammer 1997). Google's DeepMind too used enormous processing power—almost 2000 CPUs, 300 GPUs and an unknown number of Google's proprietary Tensor Processing Units, but it also used a different programming approach that did not rely solely on brute computing force.

To obtain the level of sophistication (as opposed to brute force) of analysis that would approximate human capabilities, machines needed access to massive amounts of data in addition to processing power, which essentially limited their development. Underlying today's AI systems are machine learning (ML) algorithms—the sets of instructions or rules that define how a computer program should solve a specific problem—that learn from data without explicit programming on how to do so. These machines learn not so much by following rules of logical behavior as much as they do by identifying patterns in data. The models they use are first trained on historical data and then the machines use those starting points to work on new data. While the basic algorithmic architecture of AI was developed fairly early, it was the explosion of information on the internet, basically the digitalization of almost all human knowledge, that ultimately permitted ML algorithms access to the volumes of data that were required to make AI feasible. In 2001, there were 2 zettabytes of data in circulation, (i.e. a sextillion bytes or one trillion gigabytes) Now, we are close to 100 zettabytes. Without this data, ML could not be effective.

The final advance necessary to the realization of AI was made in 2017 with the development of Large Language Models (LLM) based on transformer neural networks (see Vaswani 2017). These are a type of AI model which are trained on massive amounts of text data and utilize deep learning techniques to process and generate natural language and generate original content. This was an important step in accelerating AI development because not only did it allow human-machine interactions at unprecedented levels (and success at Turing's test) but also because the learning mechanisms embedded in LLM are transferrable to other AI domains such as image recognition, thus facilitating rapid progress in these contiguous arenas. The analytic capabilities of LLM-based AI, like ChatGPT, are immense. ChatGPT uses 5 million tokens of information and tracks 175 billion parameters. Because of this enormous number of parameters and the ability of the machine to keep track of them, such generative programs can develop insights based on statistical patterns that would elude humans. Moreover, because there are so many parameters, humans will not be able to really understand or unpack *how* AI arrives at certain decisions. It will, truly, be a black-box.

Commercially, AI is already a huge market—according to one estimate a US$ 150 billion market—and clearly projected to grow at a faster clip in the coming years. The main applications of ML and AI in the corporate world are in "predicting" and "prescribing". While predicting or forecasting was always part of the corporate agenda, what ML and AI promise is more accuracy, more speed, and lower costs. Some researchers have already documented how ML has substantially reduced the cost of making predictions (Panetta 2018).

In doing so, AI could re-invent decision making as well as business models. By processing incredible amounts of data and finding non-obvious patterns, AI can increase customization and personalization to a much greater degree (Baldwin 2019) and provide insights that humans would ordinarily not be able to see (McAfee 2014 & Brynjolfsson), thus improving the forecasting of future trends and outcomes to reduce market and investment risks (Davenport and Harris 2007). AI can also run large amounts of simulations to develop and run scenarios and thus improve decision-making under uncertainty (Pentland 2014). And of course, one of the biggest advantages of machines is their ability to perform well-defined routine tasks more accurately than humans, thus improving efficiency and reducing costs.

In summary, while computing power advancements enabled machines to outperform humans in controlled environments, such as strategy games, they still struggled to replace human experience and intuition in non-codified environments. The development of LLM algorithms, coupled with the availability of vast amounts of electronic data, has enabled computers to recognize statistical patterns in knowledge domains where mathematical codification is lacking, such as writing and communication. Consequently, AI has become increasingly adept at pattern recognition and deriving meaning from unstructured data, allowing it to imitate and sometimes surpass human cognitive functions that rely on knowledge and experience. Furthermore, AI can directly communicate with the public, summarizing complex ideas at different levels of complexity and human understanding. While the "intelligence" in AI does not refer to the ability to generate conceptually aware original and creative thought, but rather to assimilate and analyze massive amounts of disparate information sources to generate credible analysis, argument, and action, language models such as ChatGPT have demonstrated remarkable proficiency, now often winning the imitation game.

3. AI AND CLIMATE ADAPTATION: THE USE CASES

In the context of global warming, the two main areas of action are mitigation and adaptation. Mitigation implies the reduction of greenhouse gas (GHG) emissions to minimize the increase in mean global temperatures while adaptation implies increasing the capacity of a community to respond to climate-induced shocks to their wellbeing. Mitigation depends upon the reduction of fossil fuel use through disincentives and through investments in carbon neutral energy sources. Adaptation or adaptive capacity is a function of the resources available to a community for dealing with shocks, either in anticipation of or in reaction to them as well as of the decision-making processes by which a community employs these resources. Such resources include productive, infrastructural, and financial assets as well as knowledge and practices. In the context of climate adaptation, vulnerability and resilience have been related to the amount of disturbance that a system can withstand and return to a stable state following a perturbation (Gunderson 2000), or the ability of a system to recover from an adverse event. The IPCC (2007) treats it as "the ability of a social or ecological system to absorb disturbances while returning to the same basic structure and ways of functioning, the capacity for self-organization and the capacity to adapt to stress and change appropriately." In this view, adaptation can also mean settling into a new state.

The mitigation goals facilitated by AI are capital intensive and though developing countries could benefit from them, they are more relevant to advanced industrialized countries. AI can contribute to climate change mitigation by transforming electricity and transportation, largely through optimization routines which analyze production (e.g., smart grids) and use (e.g., smart buildings) patterns to reduce overall energy consumption and GHG emissions. AI can also facilitate a greater integration of renewables into the energy mix by providing more accurate forecasts of renewable energy generation based upon weather patterns and historical data. The modelling required for such initiatives is both simpler to code mathematically and compatible with general business process improvement modelling and can thus be deployed rapidly since they respond to concerns with potentially high levels of economic payback in terms of cost savings or commercial opportunities.

For climate change adaptation, however, this paper will argue that the potential of AI is more pronounced in the LDCs and could be truly transformative in these cases. The reason for that is that while wealthier countries already have sophisticated warning systems and disaster recovery capacities because of better infrastructure and more resources, the LDCs are in more precarious situations because they have less physical infrastructure and fewer financial and administrative capacities to handle disasters.

For climate change adaptation, however, the use cases are more complex in terms of cognitive requirements and decision making under uncertainty than they are for mitigation. In terms of adaptation, the potential of AI can be divided into the same two main categories as in business applications: Predication and Prescription. There are discussed in the following two sections.

4. PREDICTION

Better predictive capacity would give administrators more advance and accurate warnings of crisis and therefore more time to prepare to reduce impacts and increase resilience. Based on large databases and ML, early warning systems (EWS) can provide better and earlier predictions of adverse climate and weather-related phenomena such as hurricane trajectories, droughts and (crop) disease AI systems can also enhance resilience through an integrated analysis of shocks, vulnerabilities, resources, capacities, and options. Having more time to react to extreme weather events such as floods and landslides would reduce loss of life and infrastructure damage. It would also enable them to remove and safe movable assets.

Given that the most vulnerable in developing countries generally reside in rural areas and are engaged in small-scale agriculture, ML can help them increase and climate proof their yields by optimizing water and fertilizer use as well as predicting disease and pest outbreaks.

The potential for leveraging large amounts of interconnected data to improve climate change mitigation and adaptation has been recognized for quite some time, but progress in developing and deploying these resources has been poor. An enormous amount of climate-relevant data which has been collected, but data science and big data has not been leveraged to furthering planetary understanding (Faghmous and Kumar 2014). Huntington et.al. (2019) argue that AI could facilitate better climate analysis EWS by building on climate connections which have already been discovered by scientists but again this has not occurred. As a result, larger scale and more detailed predictions have not been realized. At the micro-level, there have been important developments in crop management. Goap et.al. (2018) present an interesting use-case of Internet-of-Things based smart irrigation management, basically an autonomous sprinkler system. These too have not yet been deployed at scale.

This should not be surprising. Use applications are prioritized not by economic and social value but by monetization strategies. Even though climate disruption related losses are in the hundreds of billions—just in 2021, the EU member states suffered over USD 50 billion in losses—researchers (e.g., Care & Weber 2023) find that the finance literature generally neglects climate finance. If 'show me the money' is the operative mantra, then data science and AI is not likely to prioritize applications related to climate disruption, especially in the least developed countries which have limited disposable financial resources. This is frustrating to many observers. An analysis of the Sustainable Development Goals (SDGs) by Vinuesa et.al. (2020) based on consensus-based expert elicitation found that AI would have more positive impacts on the environment and on climate issues and more negative impacts on society, and yet AI is already being deployed at much larger scales in areas which affect the latter arena, precisely because that's where the money is. Clearly this requires more urgent attention from policymakers, especially in terms of the US$ 100 billion which wealthier countries have pledged to (but have not yet) distribute to developing countries for dealing with climate change.

5. PRESCRIPTION: ADDRESSING THE CAPACITY GAP THROUGH GENERATIVE AI

While better predictions with more lead time would give administrators more time to prepare (though in many cases very little more time), they would not actually recommend or prescribe how administrators should deal with the situation. What to do is an issue which has been the subject of a long period of study by scholars of public policy and administration. In this section, I will illustrate how policy making and administration is fraught with incomplete analysis (a problem which is even more pressing in capacity constrained LDCs) and how generative AI can improve the quality of administrative decision-making.

In 1910, John Dewey, an American philosopher of the pragmatist school, published an influential text which was still circulating some one hundred years later when I decided to use it as the first assigned reading in my graduate-level Policy Analysis and Design course. Dewey's book was titled *How We Think* and in one of its chapters he outlined a five-phase process of reflective thought which seemed to me particularly useful to begin a conversation with new policy scholars about rational policy analysis and design processes. According to Dewey's narrative, the human mind moves from a state of troubled and perplexed curiosity about a problem to intellectualizing about it, developing hypotheses about causes and solutions, reasoning about it and finally testing the hypotheses by action. Here 'reasoning' referred to developing the complex chain of cause and effect which leads to root causes, observed effects and desired outcomes.

This was an effective caricature about how we (ought to) think and solve problems and lent itself to the utilitarian neo-classical utility maximization model of decision making which described decision making as an exercise in evaluating a set of alternatives by calculating the probability distribution of its outcomes and selecting the best alternative from among them. In 1957, however, Herbert Simon introduced the concept of bounded rationality whereby he claimed that humans do not make perfectly rational decisions because of “the limits of human cognitive capacity for discovering alternatives, computing their consequences (and) making comparisons among them”. Instead, Simon proposed that humans generate alternatives as needed and do not seek to maximize utilities but merely indulge in a process of satisfying. Shortly thereafter Charles Lindblom (1959) applied this thinking to administrative science where he proposed that administrators actually follow a process of “muddling through” in which an administrator responsible for formulating policy does not use that complete, elegant, unimpeachable but in the end unpracticable approach of fully analyzing all of the relevant data with all of the relevant theories to generate all of the possible alternatives and to select among them the one best alternative to maximize his value function through a comprehensive analysis of all of the potential outcomes. Instead of this rational-comprehensive or root approach as he called it, Lindblom said administrators rely on limited comparisons, incrementalism, and “satisficing” (as opposed to March's “satisfying”). This approach was developed further by Cohen and March (1972) into their Garbage Can Model in which problems and solutions are only loosely coupled and decision making happens under conditions of "organizational anarchy" characterized by problematic preferences, unclear technology, and fluid participation. Thus, it turns out that administrators do policy much in the same way as amateurs play chess—with feeling and intuition and limited explicit strategic calculation.

To combat against this organizational anarchy, policy scholars have repeatedly developed and recommended systematizations of administrative decision processes into discrete steps where specifying objectives, analyzing tradeoffs, developing alternatives, prioritizing, etc. are tackled sequentially and impartially. In other words, while policy analysts needed to act fast (or intuitively muddle through) because of politically defined timelines (see Lidman & Sommers 2005)), they also needed to think slow to develop deliberated and rational solution that would durably solve a problem (See Kahneman 2011 for details of fast and slow thinking). Indeed, in one popular book for policy analysts, Bardach (2000) offered a structured approach to policy analysis primarily as “a reminder of important tasks and choices that otherwise might skip your mind", essentially critiquing the muddling-through approaches administrators follow.

If developing and processing alternatives through a process of elimination and dominance is the essence of good decision making, then issues of staffing and information processing abilities become critical. Cognitive and informational limitations are an important component of bad and limited decision making. Good analysis often requires slowing down the decision process to gather data and analyze the cause-effect chains through theoretical and empirical knowledge. If good process is merely a guide which is more often replaced by intuition supplemented by quick analysis conducted by existing analytical resources, one can easily connect the dots to conclude that LDCs have a serious capacity gap not only because of training but also experience gaps. After all learning is best done by doing, and doing depends on opportunities to do (past and present) which themselves are a function of available financial resources.

Based on our previous studies on institutional and organizational capacities for climate change adaptation (Tankha et.al. 2020, Tankha, et.al. 2019, Ranabhat et.al. 2018, Tankha & Rauken 2014), we find that while knowledge of climate change causes and potential impacts at a general scale is high, the capacity to decide what to do in specific conditions, especially at the local level, is extremely limited. The national and subnational climate change adaptation plans that have been developed in LDCs are unhelpful. Most of these were prepared by donor-paid consultants and while there were formal and informal stakeholder participation processes in place, actual engagement in these process was superficial and limited. As a result, administrators in these countries either are not familiar with the details of these plans or if they are familiar, they find them of little practical use.

This capacity gap has long been recognized and indeed capacity building for development became one of the international aid agencies' preferred approaches for development assistance since the 1990s, but the motivations behind and designs of these programs have led to their failure. One was the preoccupation with corruption. Aid programs focused on building something were always accompanied with the risk that some contractor or middle man might use substandard material or in some other way siphon off money. Capacity building programs had the advantage that unlike bad cement, bad training does not leave evidence which can be used to chastise the administrators of these programs. Most of these programs also suffered from bad design, trying to do too much in too little time with overpaid trainers from donor countries and a modular approach with little long-term engagement with the trainees. Pedagogically, they were ineffective. In any case, the amounts allocated were entirely too little to reach the numbers needed and resulted in training programs more diluted (and even less effective) than homeopathic medicine.

This is precisely where generative AI can have a fundamental and transformative impact and what we envision is a generative AI powered platform which is developed and widely distributed in LDCs specifically oriented to dealing with climate change challenges. Such a platform would incorporate constant real-time data from satellites and other sensors and incorporate both predictive and prescriptive functions.

The natural language capabilities of the latest iterations of generative AI allow the machine to understand human intent and to not just generate responses but also to do so at different levels of complexity and clarity. Such programs can explain even complex concepts, such as quantum physics for example, tailored to the level of a middle school student or to that of a post-graduate physics student. Given this capability, local level administrators in countries such as Nepal or Uganda could ask the platform questions and receive responses that are scaled to their level of education and familiarity with the relevant concepts.

Because these programs have interactive capabilities, they could also be structured to ask for clarifying information from their human interlocutor or even to obtain the necessary information autonomously from other data sources. In essence this would mean that these front-line workers and administrators would have access to a handy expert on call.

LLMs also have the ability to "teach" the user about its logic and thinking. The ability to compartmentalize the instructions into discrete steps or in other cases to provide a more comprehensive context and approach way not initially be clear to the user. When faced with an output that might seem surprising, one can even attempt to get the platform to "explain" its reasoning by requesting step-by-step responses.

Before the emergence of LLM-based AI, the classical model of machine-augmented decision-making systems was a simpler linear model where humans wrote the processing rules and machines incorporated the data and did the calculations and provided an output which would then be implemented by mainly human interventions (Figure 1).

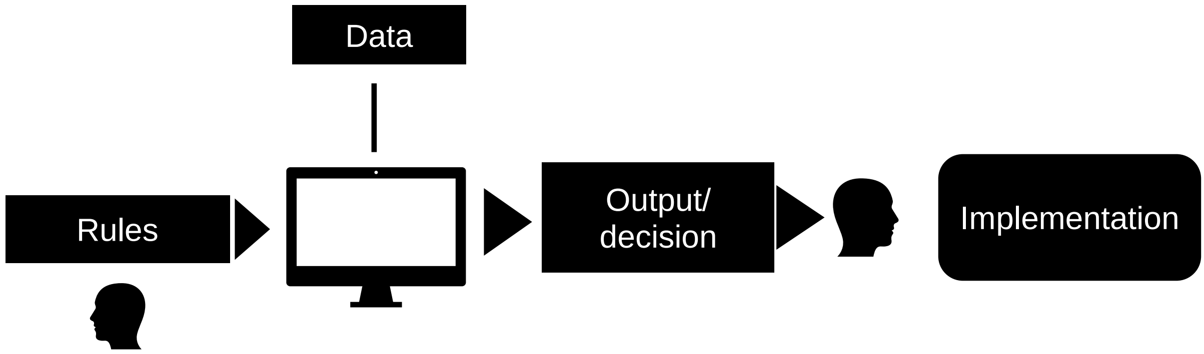

Figure 1: System Architecture for Traditional Computer-assisted Decision Making

After the emergence of LLM-based AI, the model is more complex. In this approach, humans write the learning algorithms and the learning objectives and provide the training data (which is tagged or identified to different degrees) and the machine itself writes the rules to which data will be applied to provide the output. At this stage, there are two variations to the model: in one variation, there is a human intermediary to implement, in the other the machine itself through advanced robotics can "implement" the decision output (Figure 2) So, in effect, the machines will now write the rules to serve human defined objectives rather than merely follow human rule-inputs.

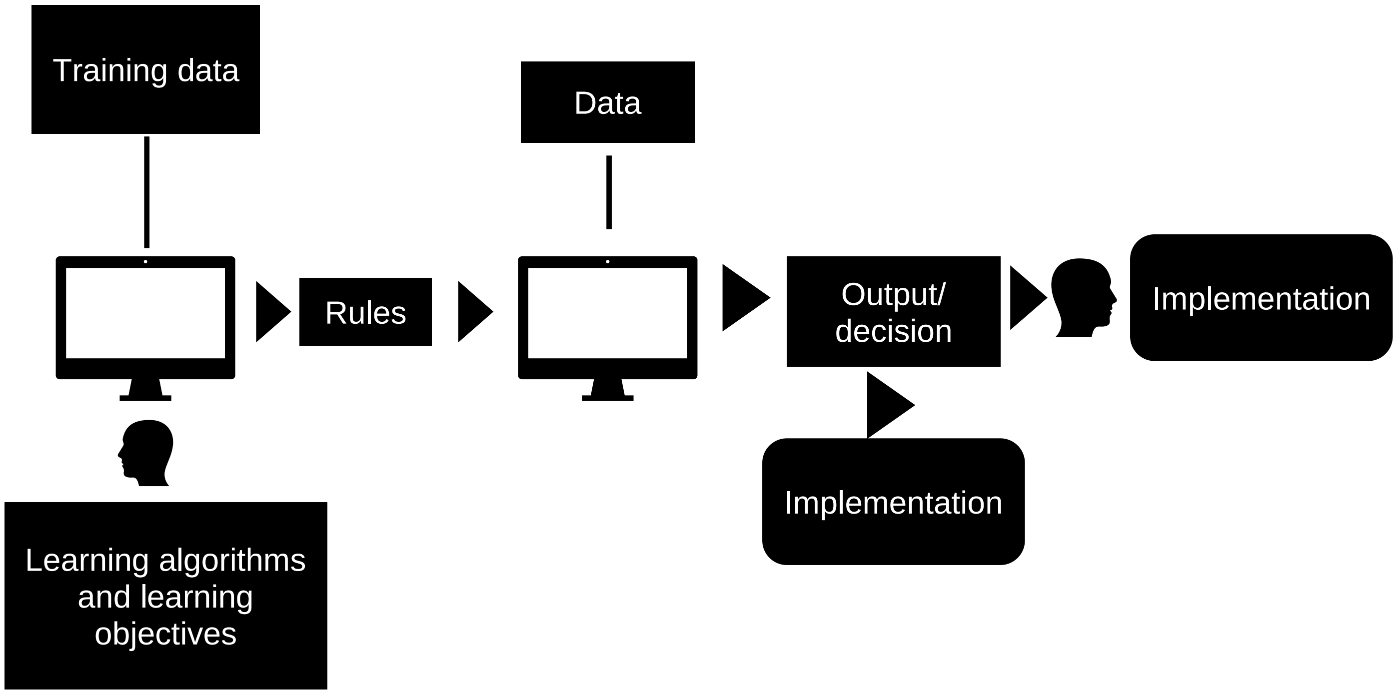

Figure 2: System Architecture for Generative AI assisted Decision Making

In the context of public administration, this level of potential autonomy raises the question as to how much autonomy the underlying algorithms should have. The choice is between an expert system, and a more autonomous GPT-type system.

Expert systems are a specific type of artificial intelligence (AI) technology that mimics the decision-making capabilities of human experts in a specific domain or field. Also known as knowledge-based systems, expert systems are designed to capture and utilize the knowledge and expertise of human specialists to solve complex problems or provide intelligent recommendations. They are one approach within the broader field of AI and whereas other AI techniques, such as machine learning and deep learning, focus on learning patterns and making predictions from data, expert systems rely on explicit knowledge representation and rule-based reasoning. They are particularly useful when the knowledge is well-defined and can be explicitly represented, while machine learning techniques excel in situations where there is a vast amount of data available for learning patterns or making predictions. Expert systems have already been successfully applied in various domains, including medicine, finance, engineering, and troubleshooting complex systems. They excel in situations where the knowledge of human experts is valuable but limited in availability, scalability, or consistency. Expert systems can capture and reproduce this expertise, enabling consistent decision-making and problem-solving capabilities, even in the absence of human experts.

The system architecture consists of four components. The first is the knowledge base which stores the domain specific knowledge and expertise and contains the collection of facts, rules and heuristics that represent the knowledge of human experts in a particular field. This base is created through a process of knowledge engineering which is mainly focused on the task of arranging in a structured format the knowledge which can be elicited from domain experts so that a machine can use it. The second is the brain of the system which is the inference engine which reasons and makes decisions or recommendations based on matching facts and rules contained in the structured information of the knowledge base. The third is the user interface which can take various forms (text or graphical), the most useful being a natural language interface like that of ChatGPT. A good user interface is interactive and allows the inference engine to accept input from users, present questions or prompts for clarification and display the system outputs and recommendation. Finally, expert systems have an explanation facility that can explain the reasoning behind the system's decisions or recommendations. This helps users understand why a particular solution or recommendation was provided, increasing transparency and trust in the system.

GPT architectures are trained on massive amounts of data to generate responses based on patterns and information, so they are not systems where the rules are devised by experts (humans) as is the case for Expert Systems. GPT are self-learning whereas Expert Systems are curated and encoded by humans. Because GPT responses are based on statistical patterns, they can generate contextually relevant responses which may or may not exhibit deep reasoning or understanding. Expert Systems, however, use logical reasoning and rule-based approaches and are systems that leverage explicit domain knowledge but do not leverage the full potential of machine learning. One big disadvantage is that at Expert Systems need to be developed specifically for the applications we are discussing here and will therefore be hostage to funding and proprietary issues. GPT systems have more long-term potential, but they also raise certain key issues that we highlight below.

6. ON PLATFORMS AND PROMPTS: DESIGN CHALLENGES AND APPROACHES

As with any transformative technology, potential dangers, pitfalls, and challenges arise, some very serious. The motorcar, for example, transformed human mobility more than any other invention, yet has caused hundreds of thousands of deaths through accidents and pollution. While it is easy to say safety must be paramount, as a society we constantly make compromises among costs and benefits. With generative AI too, a process of discovering problems will no doubt appear. Specifically, regarding the use-case we present here, a few of these are enumerated below as an attempt to both contribute to better design and flag important issues of concern.

6.1. Prompt Engineering

One way in which working with LLM-based AI platforms is fundamentally different from working with other digital technologies is that these platforms are specifically designed to try and understand what is being asked of them and to train themselves to do so more unerringly through their interactions with humans. Traditional digital platforms required the humans to be trained to instruct the machines in an unambiguous manner to do exactly what was desired. In the latter case, the machines had no ability (or programming) to understand human intentions. LLM platforms, which are trained and self-learning rather than programmed can deal with normal human communication. In this sense Jen-Hsun Huang, the CEO of NVIDIA (one of the leading AI microchip firms) recently stated that natural language processing could lead to the end of the digital divide amongst those who possess computer programming and literacy skills and those who do not.

Nevertheless, computers do not have a strict understanding of intention. Language models produce text based on the probability for a word to occur based on previous words in the sequence. ChatGPT was trained on 45 terabytes of text from the internet to calculate that some sequences of words are more likely to occur than others. Therefore, the output from these platforms still depend to a lesser or greater degree upon the prompts that they receive and communicating with these platforms will still be more rewarding for those who are able to express their requirements more analytically and intelligently.

In response to the deployment of LLMs, a whole sub-field of prompt engineering has emerged, which involves specifying the desired outputs as clearly and as unambiguously as possible. This involves the development of a process of categorization of prompts and focuses on how to structure these prompts, which kinds of instructions to provide in the prompts and what kinds of information to include to guide the computer.

Clarity is of course key to obtaining better information and output. This can be done with less ambiguous language and by supplementing the prompt with context and examples to reduce uncertainty and ambiguity. Remember, machines have no conceptual understanding of what they do. Rather, LLMs predict the next word and sequences of words given a prompt, so using an incorrect or inappropriate term could easily set the machine barking up the wrong tree. Setting limitations, constraints and priorities can also improve the quality of outputs. These limitations and constraints might even be obvious to the human interlocutors but obviously the machine has no such innate ability.

This will not come naturally to the users from low-capacity areas, and as the preceding discussion about the capacity conundrum has made clear, reaching out and training these people will not be possible. So, in some way, the model has to be trained to understand better what the human intention is so that the chances of value misalignment are reduced.

6.2. Platform Specificity

If the people cannot be trained, it must be the platform which must be trained to understand the limitations of its interlocutor and to interact in such a way as to ask the interlocutor to be more specific, asking several questions regarding objectives, resources, abilities and capacities before providing the output. A general ChatGPT-type platform would by itself be unfit for purpose. While such platforms could serve as the underlying basis for generating outputs a more specialized and dedicated interface would be needed, one which is specifically trained to recognize the limitations of its users. This means that the available platforms cannot just be used for this task. What will be required are more specialized platforms.

Another aspect which would need to be addressed is regarding the data and libraries which the platform would use to generate its outputs. The internet is full of incorrect and biased information. Current AI platforms do not distinguish well between valid sources of information and contaminated and spurious sources. The accessible libraries for such platforms will have to be restricted in some way to ensure only high-quality information is used. This creates an engineering conundrum since the volumes of such information might not be large enough for the type of machine-based analysis and learning that these GPT platforms need.

Engagement is another key challenge related to platform usefulness. Designing a system that promotes engagement requires multidisciplinary teams that undertake extensive discussions with potential users. In corporate applications, there is usually a product owner who is the interface with the technical team from the users' side and this person generally exercises control over the technical team to ensure the product reflects the users' requirements rather than the designer's' preferences. This engagement with stakeholders needs to be at three stages: scoping (where the boundaries of the problem are defined), developing and deploying.

A final observation regarding the platform is that it would be essential that several competing platforms are developed. Unlike consumer products where competition and the availability of alternative leads to increased offering and better value, network and platform economies are built upon quasi-monopolization. In the case of platforms which may be used for climate change management tasks by personnel from capacity and resource constrained environments, a proliferation of competing donor (or otherwise) funded platforms would probably just result in niche demo projects and not a widely dispersed and widely used platform.

6.3. Legitimacy, Accountability and Transparency

In public policy and administration, the issue of the legitimacy of decisions is an enduring concern. Institutional protocols governing representative democracy and professional bureaucracies have been developed and refined over centuries to reflect societal preferences and choices regarding participation, transparency, authority, prioritization, and decentralization. Even with the same training and access to the same informational resources, human decision processes are idiosyncratic and lead different individuals and teams to different decisional outcomes. So, legitimacy relates not so much to the output as to the process and in public administration in representative bureaucracies, the process of decision making is itself the legitimation mechanism. Legitimacy also links with issues of accountability, responsibility and transparency.

These issues are not easily reconciled with the black-box nature of AI programs, more so when they are based on proprietary technology. If AI is privatized, centralized, and black-box in its essential decision-making nature, then will legitimacy be reduced to only technical proficiency?

Another important concern is how responsibility will be assigned if generative AI is used widely. This is not a question of being right or wrong. Humans are make the wrong decisions all the time, often with severe consequences, but what matters is intent. Since machines do not have intentionality, how will society deal with machine induced mistakes?

Those who research AI currently underestimate the complexity and relevance of politics in governance and administrative process. Public action is infused with political concerns that are left unstated in official rules that bureaucracies must follow but are implicit in every decision and action. Some of these are positive and desirable biases; others negative and undesirable. AI programs, to the extent that they learn from real world data which is saturated with societal biases and discrimination, often reproduce biases and discriminatory practices. For example, fraud detection programs used by Dutch authorities routinely falsely accuse those of minority backgrounds. In this case algorithms reproduce but also provide a veneer of neutrality and legitimacy to a deep and widespread bias in Dutch society against minorities.

Related to this issue is the question of the documentational procedures which need to be put in place to reach the standards of transparency that public action requires. Shall all chats with the platform be logged with unique user identification? How does team working evolve under such protocols and processes? Sharing, protecting and the transparency of model and of the data need to be resolved compatible with political and administrative norms while recognizing these systems are being developed refined and maintained in corporate environments.

6.4. Reliability and Confidence

Is the potential output reliable? Given that the underlying algorithms are statistical rather than rule-based, how vulnerable are they to producing answers or outputs that are statistically valid but logically impaired? What could be done to reduce and minimize such errors? Given that the users are also in the lower-capacity groups, would they recognize that the response they have been provided is based on a misinterpretation of their intent? In that case, how should administrators, especially from capacity-challenged areas, develop the confidence to question machine outputs that they suspect might be flawed?

In addition, given that the training of these programs is based on the English language primarily and the potential users will have imperfect language skills, is there a danger that the prompts will be misinterpreted by the machine?

6.5. Unintended Effects

Those who work with AI have always been worried about two related phenomena: emergent behavior and value alignment problems. Emergent behaviors are those which arise spontaneously from the self-learning capabilities of the AI platform and are not explicitly programmed and indeed surprise the programmer. Value alignment problems emerge when the AI agent “misunderstands” what its objectives are and no longer aligns with the goals of the users. Both of these issues have been mined for science fiction plots. How relevant are these concerns for the use-case presented here?

The effect of AI on motivation is another issue that will emerge. Will the ready availability of rapid and plausible answers from AI systems act as an incentive or a disincentive to workers seeking to upgrade their capabilities? As we academics have already started grappling with the consequences of ChatGPT in the submission of student assignments, what will be the long-term effects on capacity?

Will the values that are embedded in the socio-cultural contexts of the society conflict with values embedded or learned by the algorithm? Will this generate more controversy and conflict, of which kind and what are the logical but unforeseen consequences?

7. FINAL OBSERVATIONS

AI has reached a critical stage in its development and although the element of hype is present, the concerns acknowledged and highlighted by experienced researchers and thought leaders in the field can be taken as a sign that dramatic changes in how we work, play, and make decisions are underway. The funding being allocated to further research and development in generative AI by the large tech companies is another proxy variable which indicates that insiders expect great changes to be obtained soon, including significant breakthroughs in business application which then usually translate into governance applications.

Technological breakthroughs can rapidly change the relative positions of peoples and societies. In many cases, they reinforce inequalities. In others they attenuate them. The capacity gap is a promising use-case which can help attenuate disparities related to governance and administration in general and to climate change management in particular. To reduce the probability that AI increases disparity, we need to quickly begin to think about how it can be deployed in low-income countries.

Doing so will require attention to be paid to both the financing of the development of the processes and platforms that will be needed to facilitate the use of AI in these contexts as well as design choices. Critical issues include level of human supervision of the platforms and how it should be done. Moreover, bureaucratic rules in terms of decision processes and record-keeping responsibility will need to be adapted for incorporating generative AI.

It will also be necessary to ensure that design for the poor is not poor design. These platforms should not be developed under standard aid modalities where donor concerns and employment for the boys is of a greater priority than creating systems which work for the user population. This will require meaningful user engagement rather than box-checking participatory approaches which currently characterize aid.

Generative AI is deeply problematic and thorough safeguards are paramount. Nevertheless, the technology is also truly transformative and ultimately the basis for taking decisions on these and other relevant issues (many outside the scope of this short paper) will have to be founded on realism rather than ethical idealism

REFERENCES:

Baldwin, R., 2019. *The globotics upheaval: Globalization, robotics, and the future of work*. Oxford University Press.

Balogun, A.L., Marks, D., Sharma, R., Shekhar, H., Balmes, C., Maheng, D., Arshad, A. and Salehi, P., 2020. Assessing the potentials of digitalization as a tool for climate change adaptation and sustainable development in urban centres. *Sustainable Cities and Society*, *53*, p.101888.

Bardach, E. and Patashnik, E.M., 2019. *A practical guide for policy analysis: The eightfold path to more effective problem solving*. CQ press.

Bhatia, S., 2017. Artificial intelligence for better climate governance [J]. *Journal of Artificial Intelligence Research & Advances*, *4*(3), pp.37-43.

Cohen, M.D., March, J.G. and Olsen, J.P., 1972. A garbage can model of organizational choice. *Administrative science quarterly*, pp.1-25.

Davenport, T.H. and Harris, J.G., 2007. The architecture of business intelligence. *Competing on analytics: The new science of winning*.

Dewey, J., 2022. *How we think*. DigiCat.

Duan, Y., Edwards, J.S. and Dwivedi, Y.K., 2019. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International journal of information management*, *48*, pp.63-71.

Faghmous, J.H. and Kumar, V., 2014. A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, *2*(3), pp.155-163.

Goap, A., Sharma, D., Shukla, A.K. and Krishna, C.R., 2018. An IoT based smart irrigation management system using Machine learning and open source technologies. *Computers and electronics in agriculture*, *155*, pp.41-49.

Huntingford, C., Jeffers, E.S., Bonsall, M.B., Christensen, H.M., Lees, T. and Yang, H., 2019. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, *14*(12), p.124007.

Kahneman, D. 2011. *Thinking, fast and slow*. Penguin Books.

Krauthammer, C. 1997. Psyched out by Deep Blue. Washington Post. May 16, 1997.

Lindblom, C.E., 1959. The science of" muddling through". *Public administration review*, pp.79-88.

McAfee, A. and Brynjolfsson, E., 2017. *Machine, platform, crowd: Harnessing our digital future*. WW Norton & Company.

McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E., 2006. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), pp.12-12

McEvoy, D., Fünfgeld, H. and Bosomworth, K., 2013. Resilience and climate change adaptation: the importance of framing. *Planning Practice & Research*, *28*(3), pp.280-293.

Morecroft, M.D., Crick, H.Q., Duffield, S.J. and Macgregor, N.A., 2012. Resilience to climate change: translating principles into practice. *Journal of Applied Ecology*, *49*(3), pp.547-551.

Pentland, A., 2015. *Social physics: How social networks can make us smarter*. Penguin.

Ranabhat, S., L.D. Bhatta, R. Ghate, N.K. Agrawal and S. Tankha. 2018. Policy Coherence and Interplay between Climate Change Adaptation Policies and the Forestry Sector in Nepal. *Environmental Management,* Vol. 61, No. 6.

Simon, H.A. and Newell, A. 1958. Heuristic problem solving: the next advance in operations research. *Operations Research* 6(1).

Stein, A.L., 2020. Artificial intelligence and climate change. *Yale J. on Reg.*, *37*, p.890.

Tankha, S. D. Fernandez and N.S. Narayanan. 2020. Overcoming Barriers to Climate Smart Agriculture in India. *International Journal of Climate Change Strategies and Management*, Vol. 12, No. 1.

Tankha, S., S. Ranabhat, L.D. Bhatta, R.Ghate, and N.K.Agrawal. 2019. Disinterested Agents or Mismatched Plans: The Capacity of Public Administrations to Respond to Climate Change in the Least Developed Countries. *International Journal of Climate Change Strategies and Management,* Vol 11, No. 3.

Tankha, S. & T. Rauken. 2014. Climate Politics, Emission Scenarios and Negotiation Stances in India. In Bang, G., A. Underdal and S. Andresen (eds.), *The Domestic Politics of Global Climate Change: Key Actors in International Climate Cooperation.* Edward Elgar: Cheltenham, UK.

Turing, A. M. Computing machinery and intelligence.1950. *The Essential Turing: the Ideas That Gave Birth to the Computer Age*, 2012, 433-464.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, *30.*

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M. and Fuso Nerini, F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications*, *11*(1), p.233.