BRIEFING NOTE ON ARTIFICIAL INTELLIGENCE AND THE HUMANITARIAN SECTOR

OCHA centre for humdata

KEY TAKEAWAYS

- Artificial intelligence (AI) is the simulation of human intelligence by machines that are programmed to perform complex cognitive functions.

- Fueled by growing investment, data availability and computational power, AI applications are rapidly emerging across sectors.

- Utilisation of AI has doubled in the last five years with significant advancements in techniques that underpin broader, generative tasks including the creation of new images, text and video known as Generative AI (e.g., ChatGPT).

- AI has the potential to drive operational gains for humanitarians through enhanced efficiency in routine tasks like report drafting and data formatting.

- Beyond operational efficiencies, AI enables analysis at an unprecedented scale. This has the potential to support a shift towards proactive, data-driven approaches in the humanitarian sector.

- OCHA has made progress in harnessing the power of data science and advanced analytics, including AI and machine learning, to enhance humanitarian decision-making processes.

- Given AI's dependence on massive amounts of data, any AI-related policy should also include relevant data policy.

- Deeper integration of AI into the humanitarian sector presents some immediate challenges, including misinformation, reinforcement of bias, system opacity, cybersecurity, and erosion of privacy. Humanitarians will need to navigate the evolution of this multi-faceted and rapidly accelerating technology, ensuring its advances are leveraged safely, ethically and effectively.

- Recent international dialogues, including within the United Nations, signal a growing consensus on the need for global governance initiatives, grounded in a diversity of expertise, to balance risks with fostering innovation.

STATE OF THE TECHNOLOGY

Artificial intelligence is the simulation of human intelligence by machines that are programmed to perform complex cognitive functions. The development of AI systems requires three components: large amounts of data, access to compute power, and advanced mathematical techniques. As access to all three components has increased rapidly over time, so has the prevalence and impact of AI on our lives. In the last five years, the utilisation of AI has more than doubled across diverse sectors and domains.[1] Its applications continue to undergo dynamic transformations with the continuous emergence of diverse AI approaches, shaping industries from medicine to education.

The last decade of applied AI has focused primarily on undertaking narrow, discrete tasks powered by supervised learning techniques. Supervised learning requires labelled datasets for training that are used to ‘teach’ the algorithm to predict an output, based on patterns of inputs. Examples of supervised learning include image identification, spam email classification and shopping recommendations.

Recent shifts in the availability of data and compute power have triggered the rapid growth of AI techniques that enable broader, generative tasks including the creation of new images, text and video. This approach, known today as Generative AI, relies upon vast amounts of unstructured input data from which the algorithm builds ‘intuition’ around the completion of patterns across a growing number of data types, from text to images, and increasingly multimodal. Multimodal models can process a wide variety of inputs as prompts and convert them into various outputs, not just the source type. Examples include text generation (e.g., ChatGPT and GitHub’s Copilot) and text to image generation (e.g., DALL-E and Midjourney).

TYPES OF AI TASKS:
SUPERVISED VS GENERATIVE

SUPERVISED LEARNING TASK

Generate discrete recommendation, probability or classification based on labelled training data

OUTPUT EXAMPLES

Ad recommendation (targeted advertising)

Optimal route to get from A to B (Google maps)

Image classification (this is 95% likely to be your face, a cat, etc)

INTERACTION MODEL

Interacting with narrow applications of AI through software products (apps, services, websites)

GENERATIVE AI TASK

Generate new information based on a prompt (image, video, text, code) by predicting the next object (word, pixel)

OUTPUT EXAMPLES

Answers to a question (ChatGPT does your homework)

Image based on a prompt (create an image of hamsters in chairs)

Video worlds (backdrop for a game based on a fairy tale)

Coding projects (create a mobile app that helps find recipes)

INTERACTION MODEL

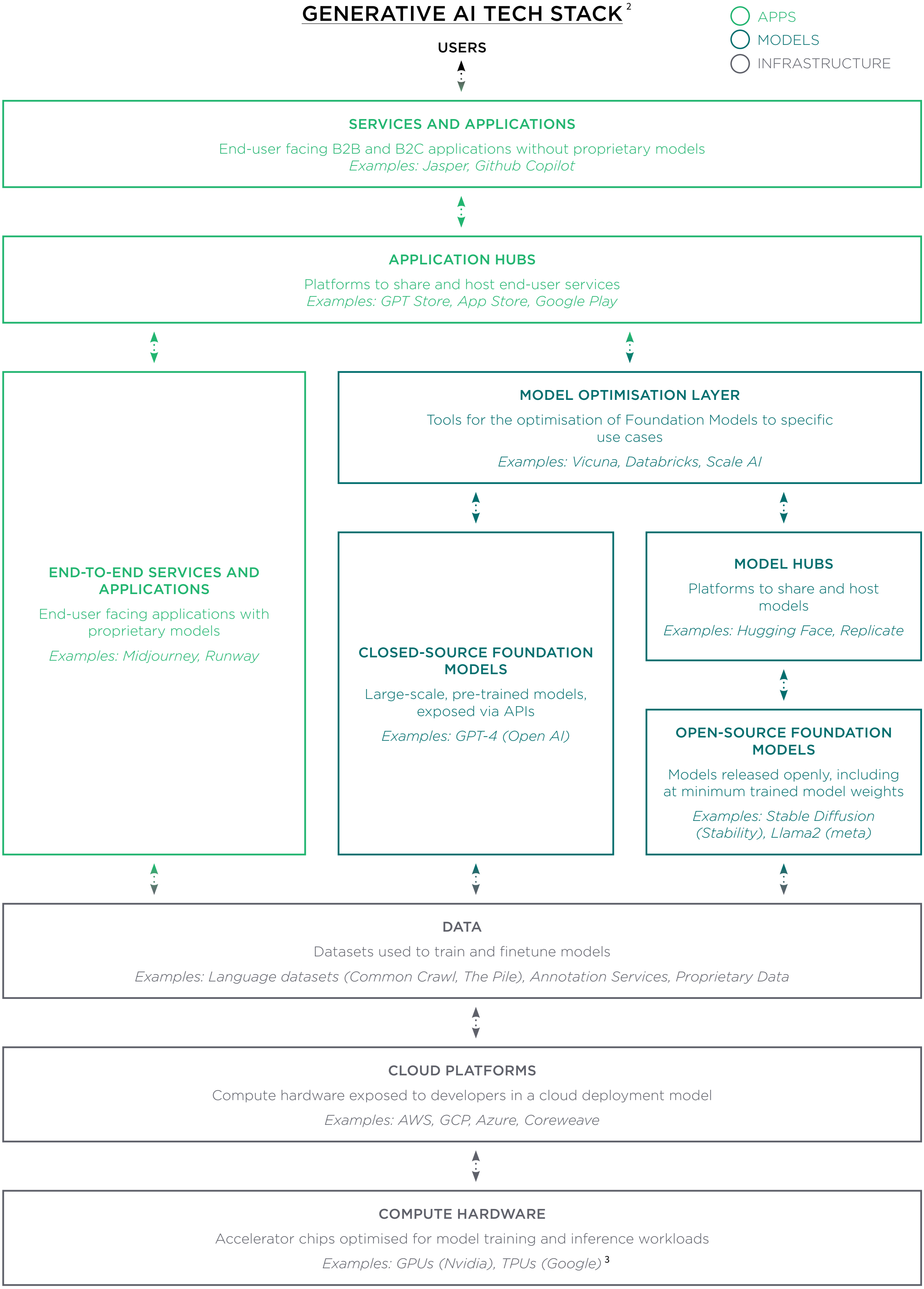
Consumers can produce content with AI using their native language, not code

Businesses can build or integrate technical services with little to no expertise

Fueled by robust investments from venture capital and corporate investment, a rich application ecosystem has swiftly emerged on top of Generative AI technologies. This dynamic services layer, harnessing shared models, cloud infrastructure and advanced computing, represents a transformative shift for end users, market dynamics, and potentially geopolitical power structures. Significantly, the prevailing market dominance faces a challenge from initiatives that provide open-source access to Generative AI technologies. This shift not only broadens access for productive applications but also raises concerns about potential misuse in areas with inherent risks.

[1] The state of AI in 2022—and a half decade in review | McKinsey, 2022

GENERATIVE AI TECH STACK[2]



[2] Adapted from 'Who Owns the Generative AI Platform?' | a16z Enterprise, 2023

[3] A processor unit is a primary component of a computer that processes data and performs calculations. A Central Processing Unit (CPU) handles a wide variety of tasks, while Graphical Processing Units (GPU) and Tensor Processing Units (TPU) are designed specifically to process machine learning and AI tasks.

DEFINITIONS

TYPES OF AI

Today's artificial intelligence technologies are all examples of narrow AI, although some believe that we are seeing the emergence of more General AI services that offer intelligence across domains.

Narrow AI: Dedicated to assist with or take over specific tasks (GPS, Alexa, self-driving cars).

General AI: Takes knowledge from one domain and transfers it to another (general intelligence chatbots). General AI can often be confused with Generative AI; despite the similarity, they are two different concepts (see definition below).

Super AI: Machines that are an order of magnitude smarter than humans.

AI TECHNIQUES

There are many technical approaches to AI that differ in how the machines are taught to understand and predict patterns. Traditionally, most AI applications have been taught through supervised learning, although Generative AI techniques are growing rapidly.

Supervised Learning: The algorithm is trained on a labelled dataset, with input-output pairs, to learn the mapping function from input to output.

Unsupervised Learning: The algorithm is given unlabelled data and tasked with finding patterns, relationships or structures within the data without explicit guidance on the output.

Reinforcement Learning: An AI agent learns to make decisions by interacting with an environment, receiving feedback in the form of rewards or punishments, with the goal of maximising cumulative reward.

Generative AI: The AI focuses on creating new, original information, often mimicking the patterns and characteristics observed in the training data. This is used in tasks such as image generation or text synthesis.

TERMS RELATED TO CURRENT AI DEVELOPMENTS

Foundation models (sometimes called a 'general-purpose AI' or GPAI system) are a type of Generative AI model trained on broad data at scale such that they can be adapted to a wide range of downstream tasks, such as text synthesis, image manipulation and audio generation.[4]

Frontier AI is a currently ambiguous term for highly capable general-purpose AI models that match or exceed the capabilities present in today's most advanced models.[5]

Generative AI is a category (and technique - see above) of artificial intelligence that learns patterns from input data in order to generate new data, such as text, images, audio, or video.[6]

Large language models (LLMs) are a narrow type of foundation model that works with language. This provides the basis for a wide range of natural language processing (NLP) tasks, such as generating blocks of text based on a user prompt.[7]

[4] Stanford University Center for Research on Foundation Models, Ada Lovelace

[5] Future Risks of Frontier AI | UK Government Office for Science, 2023

[6] TechTerms

[7] ibid

AI AND THE HUMANITARIAN SECTOR

Applications of artificial intelligence methods, including machine learning and other advanced statistics, are not new in the humanitarian sector. From assisting displaced persons through automated chatbots to leveraging AI-supported disaster mapping for emergency response, narrow applications of AI have been utilised in humanitarian response for several years. However, discussions about the role of AI in humanitarian response have accelerated as AI tools become more available and accessible to the general public. Humanitarians will need to navigate the evolution of this multi-faceted and rapidly accelerating technology, ensuring its advances are leveraged safely, ethically and effectively.

AI has the potential to drive operational gains for humanitarians, improving the efficiency of routine tasks and processes that are currently labour-intensive, such as report drafting, manual data reformatting or the calculation of metrics. Despite these tasks often being perceived as 'low hanging fruit' for automation, the potential to generate unforeseen and unintended consequences should not be underestimated.

Beyond operational efficiencies, AI has the potential to impact humanitarian response by enabling analysis at a previously unattainable scale. This technological capacity is emerging at the same time as the humanitarian sector is beginning to use data science to get ahead of crises as part of the anticipatory action agenda. These advances are underpinned by improved access to data through open platforms like the Humanitarian Data Exchange. The scale and speed of digestion and analysis supported by AI tools could markedly increase the capacity for large-scale humanitarian analysis, advancing the ambition to achieve a humanitarian sector that is "as anticipatory as possible, and only as reactive as necessary."[8]

The mainstream availability of AI systems may also improve access to analytics for teams and organisations that previously did not have in-house analytical capacity. This has the potential to deepen the integration of data into humanitarian decision-making. However, a dependency on AI systems, without a 'human in the loop' to assess quality and accuracy, may result in a new host of issues.

The Centre for Humanitarian Data is exploring opportunities to integrate AI, in support of its ambition to increase the use and impact of data in humanitarian response. These early efforts are detailed in Annex A.

NO AI WITHOUT DATA

AI systems and large-scale analysis are dependent upon vast amounts of training data; the quality of these systems is directly related to the quality of that data. Therefore, when approaching the potential uses of AI in the humanitarian sector, data quality assessment and dataset transparency is critical to ensure that the AI can be trusted. It is also imperative that any AI system have a high quality 'ground truth' dataset to measure the success of the system's predictions. Given AI's dependence on massive amounts of data, any AI-related policy should also include relevant data policy.[9] Safe, ethical and effective utilisation of AI systems is inseparable from a safe, ethical and effective approach to data management.

Considerations for data used in AI for humanitarian response:

What is the data's provenance? This includes who collected the data, how it was collected, and who is managing and updating the dataset. This can help mitigate the unknowing or harmful inclusion of falsified (synthetic) datasets.

Is the data fit for purpose? The original, intended purpose of the data should be considered in order to limit bias or mis/underrepresentation, which can lead to biased outputs by the AI.

Does the data meet our quality standard? Quality standards can drive shared expectations and approaches to usage of data, driving the overall quality of analysis. These assessments should be created or validated by technical experts (e.g., focused on interoperability and shared definitions) and subject matter experts (e.g., geographical or content area expertise).

Have we considered sensitivity and confidentiality of the data? It is imperative to preserve the privacy and security of individual data. Techniques such as statistical disclosure control can be used to assess and lower the risk of a person being re-identified from the analysis of microdata, while still allowing organisations to draw helpful insights from the data.

[8] High-level Humanitarian Event on Anticipatory Action: A Commitment to Act before Crises, Co-chairs' statement | OCHA, 2021

[9] Report Of The HLCP-HLCM Joint Session | UN System Chief Executives Board for Coordination, 2022

AI RISKS AND CHALLENGES

Deeper integration of AI into the humanitarian sector presents some immediate challenges.

Misinformation | The capacity for AI systems to offer "coherence with limited capacity for fundamental reasoning" presents a substantial risk of misinformation.[10] AI can easily be enlisted to rapidly generate large-scale mis- and disinformation campaigns, worsening the existing public trust deficit. AI's foundation in probable pattern recognition can also result in 'unintended' falsification, whereby a system that "confidently produces outputs that do not match the data it has been trained on."[11]

Reinforcing Bias | The 'intelligence' of AI originates from the data the model has been trained on. Generative AI systems generate content through the replication of patterns, and as such, risk perpetuating the inaccuracies or biases reflected in the data. This can lead to functional limitations, such as language models trained primarily on data in English that do not work well in other languages, as well as presenting more structural consequences. In replicating patterns, they risk replicating harm. In a humanitarian context, this risks reinforcing social inequalities, resulting in decisions that drive discriminatory outcomes. However, biases within the data are not the only concern. The volume of data that AI systems require risks exacerbating current humanitarian data disparities, in which data-rich contexts receive a greater share of attention simply because information is less scarce. This could lead to the further isolation of contexts that are already marginalised.

Opaqueness of Systems | The effective utilisation of AI in humanitarian response may be limited by the opacity of systems. The complexity of the deep learning algorithms that generate AI content can make it difficult for designers, let alone users, to account for the rationale behind the results offered. Given the commitment of humanitarians to the principles of humanity, neutrality, impartiality and independence, the inability for an AI to 'show its working' may impact the capacity for humanitarian accountability.

Cybersecurity | Generative AI vectors can increase the automation and feasibility of cyber attacks, allowing malicious actors to develop more sophisticated cyber threats, even if they are not highly technical themselves. This places humanitarian infrastructure at greater risk of potential cyber attacks from a broader set of actors. Furthermore, if AI systems are trained on personal data, there is a risk of exposing marginalised individuals to "needless cyber security risks and potential data breaches."[12]

Erosion of Privacy and Breaches of Confidentiality | AI systems are dependent on "large volumes of high velocity, complex and variable data,"[13] which is hard to screen initially for community consent and sensitivity. Given their capacity for analysis and pattern recognition, AIs may regenerate and republish this sensitive data, creating significant implications for the right to privacy. This could be particularly damaging in a humanitarian context, where those impacted would be among the most marginalised communities. Increased use of AI systems also risks incentivising data collection, which may "inadvertently amplify the vulnerability of individuals in need of humanitarian aid."[14]

[10] ibid

[11] UN Primer on Generative AI | OICT Emerging Technologies

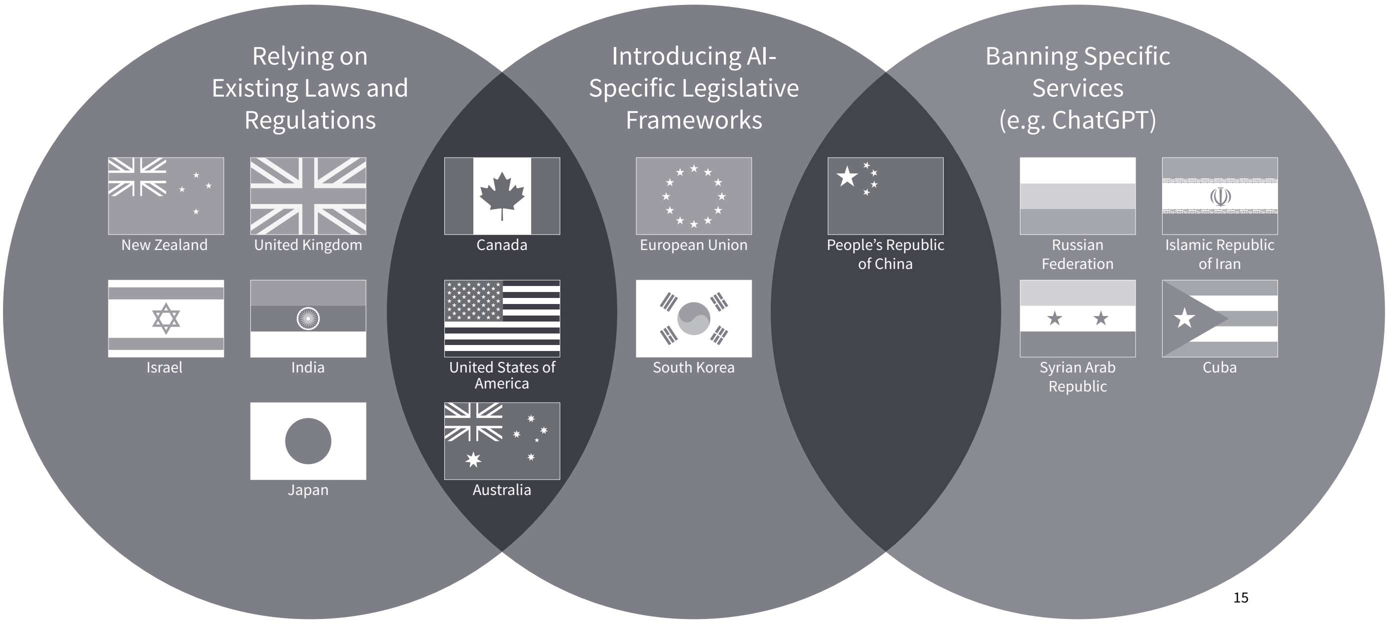
[12] Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks | International Review of the Red Cross, 2022

[13] ibid

[14] ibid

LOOKING AHEAD: THE EMERGENCE OF AI REGULATION

The current global regulatory landscape for AI is fragmented. There are a mix of approaches, ranging from outright bans to adapting existing policies or crafting entirely new regulations (see figure below). It is interesting to note that while regulations are scattered geographically, regulators are responding to a concentrated group of private corporations, mostly in the United States, that have shaped their practices in the absence of - and possibly due to the absence of - regulation.



[15]

Recent international dialogues, including within the United Nations, signal a growing consensus on the need for global governance initiatives that leverage a diversity of expertise to balance risks, especially regarding national security, with fostering innovation (Hiroshima G7[16], Bletchley Declaration[17]). They call for increased accountability, proposing measures such as third-party and researcher audit access to models, public disclosure of audit results and known issues, and the labelling of AI systems for transparent consumer interactions. This is a step towards a more cohesive and responsible AI landscape.

To learn more about the Centre's work around AI, contact centrehumdata@un.org. This paper was authored by the Centre for Humanitarian Data in partnership with Kasia Chmielinski of the Berkman Klein Center at Harvard University.

[15] State of AI Report | Air Street Capital, 2023

[16] G7 Hiroshima Process on Generative Artificial Intelligence (AI) | OECD, 2023

[17] The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 | AI Safety Summit, 2023

ANNEX A

The Centre for Humanitarian Data is exploring the application of AI in a number of ways.

Improving climate impact forecasting with AI based estimates
In the Philippines, the Centre is collaborating with the local government, the OCHA office and the Netherlands Red Cross to create a model using AI algorithms that consider factors like typhoon path, windspeed and rainfall forecasts. This model predicts the percentage of damaged buildings per municipality in the event of a typhoon. The output of the model is used to trigger anticipatory action based on an agreed threshold. Efforts are underway to replicate this approach in other countries.

Partnership with tech providers to improve humanitarian response
The Centre's partnership with Google's Flood Forecasting Initiative aims to utilise AI-based flood forecasting to enhance early warning systems and potentially trigger anticipatory responses in Nigeria. The Centre is also exploring the integration of satellite technologies and AI tools to automatically detect impacts such as housing destruction, crop loss or flood extent. Collaborations with leading organisations like NASA and the European Space Agency are ongoing to explore the feasibility of these AI-based solutions.

Increased availability of modelled data on the Humanitarian Data Exchange
The Humanitarian Data Exchange platform includes a number of data sources that have been generated through AI tools. These sources include Google's building footprint data and high-resolution population density maps and demographic estimates from Meta. Ensuring the quality and reliability of this data is an ongoing priority, and collaborative efforts with initiatives such as the Data Nutrition Project are underway to achieve this goal.

Improving data access and insight with large language models
In collaboration with DataKind, the Centre is investigating using large language models to simplify access and insight of a core set of data shared by partners on the Humanitarian Data Exchange. A chatbot is under development that is designed to respond to queries, and upon request, generate graphs and maps, enhancing user understanding of data about humanitarian operations.