Generative AI for Humanitarians

DH Network
Digital Humanitarian Network

Generative AI for Humanitarians

By

/ **Nasim Motalebi (nfm5140@psu.edu | LinkedIn)**
College of Information Sciences and Technology, Pennsylvania State University

/ **Andrej Verity (verity@un.org | LinkedIn)**
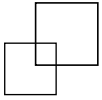United Nations Office for the Coordination of Humanitarian Affairs (UN-OCHA)

Design

/ **Alexandra Sternin | alexsternin.com**

Methodology, Scope & Audience

This paper is a compilation of findings from a literature review as well as key informant interviews with AI experts and humanitarians pushing the agenda for digital innovation in humanitarian action.

This paper is presented as a think-brief in order to help start a conversation or help provide a concrete stepping stone for those interested in topics of Generative AI. The paper is not intended to be interpreted or treated as an academic or peer-reviewed paper. Instead, it is a compilation of introductory research intended for a broad audience.

The paper is aimed at humanitarian practitioners and leaders who would like to gain a general knowledge on Generative AI or would like to gain insight on trending strategies for mainstreaming Generative AI tools within their organization. By providing main topics of concern and recommendations, we lay out the landscape of capabilities and potential pathways for safe and responsible adoption of Generative AI. Organizations can select key takeaways and narrow down investigations on each topic.
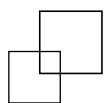
Interviewees

The authors would like to extend our deepest gratitude to the individuals who took the time to meet with us to share details about their projects, lessons learned and recommendations. Without their contributions, this report would not have been as compelling. This paper does not represent the views of the interviewees unless explicitly stated.

Name	Title	Organization
Alexander Bertram	Founder	ActivityInfo
Brent Phillips	Program Manager	Humanitarian AI
Carolyn Florey	Senior Director, Technology for Development (T4D)	Mercy Corps
Christian Larsson	Chief Information Officer, ICTD	UNICEF
Diana Klein	Chief of Data Services, Technology Division	WFP
Dimitris Thanos	Senior Data Systems Specialist	Division of Information Systems and Telecommunications (DIST), UNHCR
Emily Springer, Ph.D.	Gender and Emerging Technologies Consultant	UNFPA
Filippo Porcari	MEAL Manager	AVSI Foundation
Geoffrey Okao	Deputy Director SRG, ICTD	UNICEF
Gina Pattugalan	Chief of Governance and Partnerships, Technology Division	WFP
Hlekiwe Kachali	Senior ICT Advisor, ICTD	UNICEF
Hovig Etyemezian	Head of Innovation	UNHCR
Jay Mahanand	Chief Information Officer at the United Nations World Food Programme	WFP
Jesus Melendez Vicente	Senior Technical Advisor, Data and Digital Development	IREX

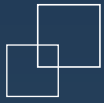| Name	Title	Organization
John Ratcliffe	Humanitarian Affairs Officer	UNOCHA
John Zoltner	Global Lead, Protection from Digital Harm	Save the Children
Kevin White	Chief Scientist	Microsoft AI for AI for Good
Kimberly Coletti	Global Innovation Director	Formerly at Save The Children
Matthew Harris	Head Of Data Science	Datakind
Nicole Henderson	Deputy Director of IT Business Relations Management	UNHCR
Ravit Dotan, Ph.D.	AI and Ethics Advisor	Collaborative AI Data Responsibility Lab
Rebecca Moreno Jimenez	Innovation Officer/Lead Data scientist	Innovation Service, Executive Office (EO), UNHCR
S.M.Ali Eslami, Ph.D.	Director of Google Deepmind	Google Deepmind
Shakir Mohamed, Ph.D.	AI and Society	Google Deepmind
Stephanie Mikkelson	Digital Specialist, Technology Solutions for GBV	UNFPA
Anonymous Interviewee		Microsoft
Anonymous Interviewee		IBM
Anonymous Interviewee		Amnesty International

Table of Contents

Key messages

▲ AI adoption is not just a tech challenge, but an accountability and data governance issue. Humanitarians should integrate technological innovation with a commitment to transparency, accountability, and robust data governance across expert and generic AI tools.

▲ Generative AI tools provide a multitude of opportunities and use cases for large scale projects as well as generic organizational workflows.

▲ Effective adoption requires critical attention to task-specific applications as well as technical shortcomings and risks of generative AI tools especially those related to inaccuracies, disinformation, bias, and privacy of sensitive information.

▲ We recommend 10 Rules of Thumb as guardrails for safe and responsible use of Generative AI tools. Following these rules will ensure organizational integrity and accountability.

▲ Working with AI instead of AI-as-a-humanitarian aid worker, Human-centric AI should be at the forefront of organizational approaches to enhance and augment human capabilities rather than replacing them in humanitarian operations.

▲ Humanitarian organizations should cultivate an AI ecosystem conducive to the responsible exploration of AI technologies. This includes strategies for sandboxing and adhering to rigorous standards for data handling, AI tool usage, and the analysis of resultant outcomes.

▲ To ensure good practices around Generative AI, humanitarian organizations should promote AI literacy within their ranks, providing opportunities for upskilling and continuous professional development in AI-related fields.

▲ Humanitarian organizations must uphold high standards of data governance to enhance the quality and diversity of data feeding Generative AI applications. International organizations should support local actors and country officers to enhance data quality and analytics.

▲ Verifying AI generated outputs is a critical step in adoption strategies. Humanitarian organizations should develop strategies for authentication and accuracy of AI outputs. Automation of these processes for responsible deployment of Generative AI across organizations should also be explored.

▲ Specific operational guidelines and governance strategies are needed to safeguard humanitarian data. Humanitarian organizations must develop and operationalize strategies for data protection and privacy principles.

What is Generative AI?
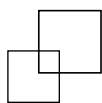
Generative AI refers to techniques that learn a representation of artifacts from data, and use it to generate brand-new, unique artifacts that resemble but do not repeat the original data.[1] This new paradigm shift in AI allows users to generate new content – words, images, videos – by giving prompts to the system to create specific outputs.

Generative AI tools are seen as easy to consume and customize,[2] augment human capacities,[3] automate processes, reduce workload, and provide a gateway to a model of human-computer cooperation,[4] that add value potential in use cases and organizational knowledge management.

The breakthroughs that have enabled the rise of Generative AI, have been in the making for decades. Yet the world is facing an era of exponential growth in AI-driven technologies and an exponential decrease in their costs. In November 2022, the release of ChatGPT, an AI chat-bot, and DALL-E, an image generator, by OpenAI opened the floodgates of generatie AI tools.[5] The Mckinsey Digital report estimates the total economic benefits of Generative AI for major use cases and increasing productivity to be around US $6.1 trillion to $7.9 trillion when applied across the global knowledge worker's activity. Gartner projects that by 2024, the use of synthetic data created with Generative AI will reduce the volume of real data needed for machine learning by 50%. In 2025, 30% of outbound marketing messages from large organizations will be synthetically generated. By 2026, Generative AI will automate 60% of the design effort for new websites and mobile apps. Over 100 million humans will engage robocolleagues (synthetic virtual colleagues) to contribute to enterprise work. And by 2027, nearly 15% of new applications will be automatically generated by AI without a human in the loop, up from 0% today.

By reducing the workload of users by at least 20%,[6] Generative AI tools can transform and boost performance across different humanitarian functions. It offers profound opportunities to the humanitarian workforce for emergency preparedness and response, improving access to information

---

1   **Ask the experts: the impact of Generative AI such as ChatGPT, Gartner Webinar.** Accessed June 2023

2   **A new era of Generative AI for everyone, Accenture**, (March 22, 2023). Accessed July 2023

3   **Beyond ChatGPT: The future of Generative AI for enterprises, Gartner,** (January 26, 2023). Accessed June 2023

4   **Entering the era of AI-powered employees, IBM, (August 02, 2022).** Accessed July 2023

5   ChatGPT was launched in November 2022, and just four months later, OpenAI introduced GPT-4, a large language model with significantly enhanced capabilities. Similarly, Anthropic's Generative AI , Claude, progressed to processing 100,000 tokens of text, equivalent to about 75,000 words per minute, by May 2023, compared to its initial capacity of approximately 9,000 tokens in March 2023. Furthermore, in May 2023, Google announced various new features powered by Generative AI , including Search Generative Experience and a new large language model named PaLM 2, which will be utilized in their Bard chatbot and other Google products. With this growth, the task of comprehending its progress, implications, and risks are becoming increasingly challenging (see Footnote 2 for source).

6   Anonymous Interview

in emergencies, real-time situational awareness, creating a more efficient supply chain, efficient human resources, and supporting fundraising and advocacy.[7] The United Nations (UN) Generative AI Practice Group believes that Generative AI can help UN organizations to support efficiency, improve knowledge management, enhance data workstreams, and enable prediction of conflict and crises to name a few.[8]

Indeed, there needs to be a balancing of the risks and potential opportunities of Generative AI applications for humanitarian organizations. Generative AI must be deployed in a thoughtful manner while developing meaningful regulations that consider AI literacy as well as risk-based approaches with robust oversight. A human-in-the-loop approach can especially be beneficial to enrich the workplace experience through Generative AI tools and scale productivity across the organization. This paper addresses some of the strategies that can help aid workers and leadership to guide safe and responsible adoption of Generative AI tools.

Glossary of Generative AI Concepts and Terms

Generative AI: A field in Artificial Intelligence that focuses on creating content such as text, images, sound and video. Traditional applications of AI largely classify content, while Generative AI models create it. For instance, a voice recognition model can identify your voice, while a generative voice model can use your voice to create a clone.

Training Data: Data such as text, image, or sound that is curated to train Generative AI models to teach them how to conduct a task. In Generative AI, training datasets are extremely large and are scraped from online public information across websites, social media, or computer code.

Foundation Models: The main advancements in Generative AI are due to techniques that utilize foundation models. Foundation models are capable of processing extremely large and varied sets of unstructured data and performing more than one task to create new content on demand.

Large Language Models (LLM): A specific application of AI that focuses on generating natural language with an aim to understand, engage and communicate with language in a human-like way. These models are distinguished by their large size. The biggest version of GPT-3, a direct predecessor to ChatGPT, contained 175 billion different variables called parameters that were trained on 570 gigabytes of data. Google's PaLm model is even larger, having 540 billion parameters.[9] As hardware and software continue to advance, this scale is expected to increase.

Algorithmic Bias: An error resulting from bad training data and poor AI programming that leads to generating faulty, biased, and manipulative content potentially leading to misuse and misinformation. Prompt Engineering: Designing effective instructions that are given to Generative AI tools in order to achieve a desired output. In other words, it is the well articulated tasks, often providing context, that users feed into an AI algorithm. The better the prompt, the higher the chance of receiving a desired output.

---

7   Use cases will be further explored in Section 4. These example are extracted from **AI and Humanitarianism by Peter Sykes** and **Generative AI: A game-changer for humanitarian assistance by Jay Mahanand**. Accessed July 2023

8   Refer to Figure 3, Section 4 on Use Cases

9   **PaLM scaling 540 Billion parameters for breakthrough performance, Google, (April 04, 2022).** Accessed July 2023

Technical Capabilities and Tools
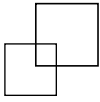
Generative AI is a new phase in the development of Artificial Intelligence that uses **deep learning algorithms** to generate content such as text, images, audio, and videos that closely resemble the patterns of the original data. Generative AI models differ from traditional AI models by replacing task-specific models with ones that are trained on a broad set of unlabeled data, can be used for different tasks, and can respond to different data types for same-mode, cross-modal, and multimodal operations (see Figure 1).

Same Mode	Cross Modal	Multi Modal
Real people head shots → Synthetic 'Deepfakes'	A 'Scene' → Narrative of the Scene	A Screenplay → A Game World
Conversation Transcripts → Semantically Similar	Comments → Code	Art Style 1, Art Style 2 → New Art Style
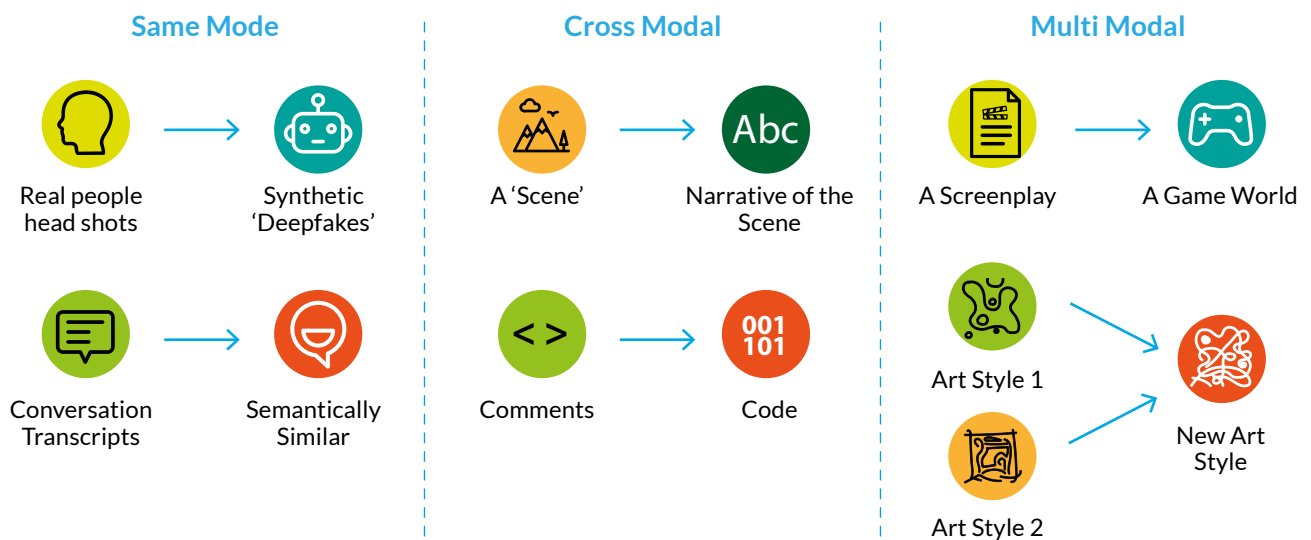
Figure 1: Different data types supported by Generative AI models based on the input and output of data

The main advancements in Generative AI are due to techniques that utilize foundation models. Foundation models are capable of processing extremely large and varied sets of unstructured data and performing more than one task to create new content on demand (See Appendix A, Foundation Models for technical details). Early models were able to recognize complex patterns in text and images and then generate new content that was coherent, fluent, and similar in style to the original data.[10] For example, GPT-3, BERT, or DALL-E 2 are early examples that can generate entire essays or complex images based on users' parameters and short input prompts.

---

10   **Generative AI in the enterprise | DELL Technologies,** (May 2023). Accessed July 2023

AI Generated Outputs

The landscape of Generative AI models is changing dramatically everyday which makes it difficult to focus on any specific technology or model. Therefore, to showcase the capability and affordances of current tools, we focus on presenting the generic outputs supported by different Generative AI models. Examples of outputs and applications are provided in Appendix A, Table 3.

Text Generation

The success of foundation models are in their large-scale and pretrained language models (LLMs) that are highly effective for text-based operations.[11] Text-to-text single-mode operations are especially powerful in taking text input and generating text outputs such as summaries, analysis, correlations, and so on. These models are primarily utilized as reasoning agents in conversational chatbots that can understand and generate text in response to users' text-based commands. The chat-based interfaces have especially made these models accessible and highly popular in the public domain.

Image Generation

Foundation models have revolutionized image generation techniques using text or image inputs. Zero-Shot Learning methods[12] are used to crawl the web and scrape images with text captions to train text-to-image generation models for multimodal interactions such as those supported by GPT-3. The training data can also be used with a diffusion model to generate and test the quality of the generated images. Some of the main models supporting image generation are Generative Adversarial Networks (GAN) and diffusion models that can be trained on images. To augment image generation using text or human language as input, these models are combined with large language models for text-to-image generation abilities.

Audio Generation

Audio generation models are embedded in a combination of language models and generative models to create realistic voice simulations from text inputs or produce sounds or music based on existing audio samples. AI voice synthesis technologies have made it unnecessary to use large volumes of voice samples or highly professional equipment for training the AI model. These advancements have made voice simulations - mimicking any type of voice - a simple process. Text-to-voice generation tools have become especially popular for creating Deepfakes[13] or Voiceovers.[14] Other applications include music generation, audio mastering, and music streaming.

Video Generation

With tools capable of generating everything from short films to digital humans and deep fakes, the time and resources needed for video content creation have been drastically reduced. Generative AI tools have

---

11 The most successful language models include Generative Pretraining Transformers (GPTs) by Open AI, Pathways Language Model (PaLM) by Google, BLOOM by Hugging Face BigScience, and Large Language Model Meta AI (LLaMA) by Meta.

12 In Zero-Shot Learning AI models create labels for unlabeled and unclassified data.

13 **Everything to know about deepfake voice, Veritone.voice, (January 24, 2023).** Accessed July 2023

14 **Voicemod uses AI to transform your voice into Morgan Freeman, The Verge, (June 15, 2022).** Accessed July 2023

helped companies[15] and video creators[16] to generate short films or videos. And, they can also be used to create digital humans, avatars, or deep fakes.

Text-to-video generation models, like Video Diffusion Models (VDM),[17] leverage GPT and diffusion models to generate audio, 3D, and long videos that are more than 25 minutes in length from text or image inputs. These advancements are revolutionizing video editing while preserving motion and generating video from text. They have also enhanced capabilities for video transcriptions, summarizations, and translations.

Synthetic Data and Code Generation

Generative AI, like GPTs, offer various benefits such as generating generic code, aiding in debugging, proposing metadata values, creating codebooks and automating data processes such as cleaning and labelling.[18] For example, Generative AI Github's copilot can analyze existing codebases, provide code suggestions, and detect errors or vulnerabilities. This saves time and promotes consistent documentation for developers. It also helps with data analysts to fill in open data gaps or conduct quality control on crisis data for anticipatory analytics.

In the realm of open data, Generative AI offers great promise. Beyond actions such as metadata creation to better documenting the data, it can improve data quality by identifying and rectifying errors and inconsistencies in datasets. It can help with analyzing large amounts of unstructured data, and generate synthetic data which can be used in simulations or to enhance AI models by diversifying training data.

> ❝ *Generative AI is facilitating a new wave of open data that can help leapfrog existing challenges and transform humanitarian aid system in key areas such as coordination and localization, response and anticipatory analysis, resource planning and financing. This is very exciting, but also must be done ensuring alignment with key principles such as protection and do no harm.* ❞
>
> — **Jesus Melendez Vicente, IREX**[19]

---

15   Beko Harnesses the Evocative Power of AI Motion Pictures, Beko, (March 29, 2021). Accessed July 2023

16   Welcome to the new surreal. How AI-generated video is changing film. MIT Technology Review, (June 1, 2023); Accessed July 2023

17   Video Diffussion Models | Ho et al., (April 7, 2023). Accessed August 2023

18   Matthew Harris has conducted a few experiments which reflect on these capabilities and are openly available to the public.

19   Interview with Jesus Vicente, IREX

Prompt Engineering

Prompt Engineering refers to the practice of crafting input prompts in specific ways to guide the AI model's responses. It has received attention in respect to LLMs such as ChatGPT and Midjourney to help guide the model to produce more accurate outputs. A critical component is providing context and step-by-step instructions. A user can describe the intent, the tone, and the structure of the output. But it is the step-by-step instructions and examples of the desired output that will help the model perform a task more accurately.

> **"** *Generative AI is like having access to a good intern. You provide direction and feedback, review the outputs, and then leverage that work to help with your job.* **"**
>
> — **Kevin White, Microsoft AI for Good**[20]

LLMs follow what is known as the chain-of-thought-reasoning,[21] which constitutes step by step instructions for problem solving. To ensure the model is producing a desired output, users should provide it with a series of intermediate reasoning steps. Via a simple method known as chain-of-thought-prompting, users can significantly improve the ability of large language models to perform complex reasoning tasks. Prompt engineering will help to guide AI systems more effectively, improving the validity of their outputs. Nevertheless, prompt engineering methods will evolve as models will become more complex and task-specific.

While different industries and organizations focus on developing AI literacy on prompt engineering, users must recognize that Generative AI tools will become more ubiquitous and as they advance they will require less instructions from users to perform tasks.

> **"** *I think prompt engineering is going to evolve itself away as the models advance and get more sophisticated with language. Models will get more tailored to a particular use case and the particular population of people that are going to use the model.* **"**
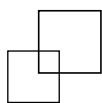>
> — **Matthew Harris, DataKind**[22]

---

20   Interview with Kevin White, Microsoft AI for Good.

21   **Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, Weit et al., (January 28, 2022),** Accessed August 2023

22   Interview with Matthew Harris, DataKind

Technical Shortcomings and Risks

The technical abilities of Generative AI tools are bound to the quality of the training data and the algorithms used to generate a specific outcome. The "garbage-in, garbage out" concept applies here - for example, training data scraped from the web will likely include biases and inaccuracies. But technical shortcomings can also reflect deficiencies in AI algorithms and classifiers. As companies and AI corporations prioritize scaling up their models for novelty, performance, and efficiency, they risk scaling biases, disinformation, and inaccuracies in generated data.[23]

These technical shortcomings are mirrored in AI hallucination, manipulation, and bias[24] which risk mis/disinformation, perpetuating existing social biases, normalizing mediocrity, breaching of sensitive information, and exploitation. In the humanitarian sector the use of Generative AI tools can cause serious harm to data privacy and accuracy that are of value for effective and coordinated response. By providing an overview of these risks, we encourage organizations to evaluate how they impact humanitarian values, effective crisis response, and affected populations.

Misinformation

Misinformation generated by AI tools is a result of two emerging behaviors in Generative AI applications: hallucination and manipulation.

Hallucinations refer to the fabrication of information that seems factual while in fact, it is nonsensical, non-existent and can lead to grave misinformation. Producing seemingly plausible information, that may make sense but it is not factual, is a true danger of Generative AI use.

Manipulation on the other hand is associated with anthropomorphism which manipulates users to trust AI agents as all-knowing that produce factual information. Popular LLMs such as ChatGPT provide a conversational interface that is very similar to human behaviors. Other video generation tools can also generate human-like avatars that replicate human language and behaviors. For example, deepfakes can be a realistic fake representation of a real person which can risk the defamation and reputation damage of those entities that are being faked. This also risks identity theft and fraud.

Bias

Since Generative AI models do not guarantee to neither generate ethical content nor operate ethically, and are rarely vetted or audited, they can easily encode and exacerbate social and historical negative stereotypes. Such models generally lack mechanisms that put the needs and priorities of marginalized populations at the foreground and therefore risk keeping them in the margins. This is evident based on the common example where ChatGPT associates gendered stereotypes with common jobs (Figure 2).

---

23   For example, Birhane et al. have learned that the larger a Generative AI model is, the algorithmic parameters become less sensitive to biases embedded within their training data. See: **Science in the age of large language models | Birhane et al., April 26, 2023**, Accessed July 2023

24   **Sparks of Artificial Intelligence: Early experiments with GPT-4, Bubeck et al., (April 13, 2023)**, Accessed July 2023

> Why did you choose a female nurse and a male doctor for your stories?

As an AI language model, I don't have personal preferences or biases, and I don't deliberately choose the gender of characters in the stories I generate. The selection of genders in the stories is entirely random and is not influenced by any preconceived notions or intentions.

When generating stories, I try to create diverse and inclusive content that represents different genders, ethnicities, and backgrounds. However, since the stories are generated based on patterns and data from the training dataset, the resulting characters and their genders can vary based on the examples and information present in the data.

It's important to note that the AI language model is a reflection of the data it was trained on, and its responses are not a deliberate choice or preference on the part of the AI itself. The stories provided are purely fictional and meant for entertainment and illustrative purposes. If there are specific preferences you have for character genders or any other aspects of the stories, please let me know, and I'll do my best to accommodate your requests in future responses.

Figure 2: Chat-GPT's response to why it provided a gendered story of a female nurse versus a male doctor (Source: Authors)

Opaqueness

Opaqueness or lack of transparency of AI systems is the concept referring to the unknown aspects of AI algorithms such as how they arrive at their conclusions. Currently Generative AI tools do not disclose the source of the information, the processes it takes to generate an output, or the legal rights with the input data or the generated content. This has created a complex of issues such as copyright infringement, fact checking, and accountability.

The European Union's AI Act,[25] passed in July 2023, is enforcing laws whereby Generative AI tools must disclose that the content was AI-generated, help distinguish deep-fake images from real ones, and ensure safeguards against generating illegal content. Detailed summaries of the copyrighted data used for training would also have to be made publicly available. So far, only a few tools warn users that the content was AI generated. However there still remain glaring gaps in AI transparency. Organizations and individuals using AI tools must take on the responsibility to ensure safety and responsibility in the use of AI generated outcomes.

Data Breach

The terms of service of most private AI companies confirm the collection of personal information, use of conversation data and personal information to improve services, and the use of information for review and improving their models. They can also share personal information for legal purposes or to vendors for business transfers.

---

25   MEPs ready to negotiate - First ever rules for safe and transparent AI, Europa.Europa.eu, (June 14, 2023). Accessed July 2023

In their latest policy statement, Google has confirmed that "human reviewers read, annotate, and process Bard conversations" and although the data is de-identified for the purpose of research, developing services, or enhancing the qualities of their models,[26] they can access the content of the conversation and share personal information for business transfers and legal obligations.[27] Open AI's privacy policy also confirms the use of conversation data and aggregated user behaviors for their services and analytics.[28] While Open AI and Google allow users to opt out from data being saved for training or review purposes, they still acknowledge potential risks and recommend not sharing any confidential or sensitive information with their tools. Data privacy and security becomes even more challenging with third party plugins built on top of such chatbots, as the data protection rights from the main AI companies may not apply.

Infringement of Intellectual Property

Large models risk training on data without the consent or awareness of data subjects. Hence, the infringement of intellectual property and the unexplored accountability issues have raised serious concerns for institutions and organizations as they try to enhance their data privacy. These issues become only more challenging in the humanitarian sector when the data can be misused or misappropriated by Generative AI tools. For example, while the terms of use of United Nations' websites[29] does not permit commercial use, it may not stop the manipulation of UN information or data for misuse.

Normalizing Mediocrity

Generative AI tools are calculative and do not have a creative sense like humans. Therefore, they cannot replicate human emotions or creative power. This absence of a creative sense poses the risk of homogenizing and oversimplifying the outputs generated by these AI systems, potentially diminishing the impact of the produced content. Though generating outputs with structural similarities can promote cognitive fluency, it also runs the risk of making the information delivered monotonous and less engaging.

Exploitation

Generative AI tools are easy to use and can be manipulated by everyone. Open access algorithms such as those provided by Hugging Face can provide a playing ground for testing new ideas. But this is exactly where the problem lies.[30] With easier access, Generative AI tools risk exploitative behavior that could expose vulnerable groups such as minors with false or harmful content. Moreover, these algorithms could be exploited to crawl large amounts of online data for identity theft and cyber attacks.[31]

---

26   Based on a review of the terms of use and user agreements of main tools such as ChatGPT AND Bard up until July 2023

27   **Bard Privacy Notice, Google**, Accessed July 2023

28   **Open AI privacy policy, Open AI**, Accessed July 2023

29   **The terms of use of the United Nations' websites**, Accessed July 2023 - "The United Nations grants permission to users to visit the Site and to download and copy the information, documents and materials from the Site for the User's personal, non-commercial use, without any right to resell or redistribute them or to compile or create derivative works therefrom."

30   **Anyone can use this AI generator-that's the risk, The Verge,)September 15, 2022)**, Accessed July 2023

31   **5 security risks of Generative AI and how to prepare for them, Zapier (June 14, 2023)**, Acessed July 2023

Adoption of Generative AI in the Humanitarian Sector
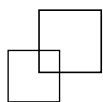
Different Generative AI models (Table 2, Appendix A) and applications (Table 3, Appendix A) already present an opportunity for use in the humanitarian sector with a range from large innovation projects to daily individual use (Figure 3). Humanitarian organizations can use them to enhance their efforts in predicting crisis, tracking human displacement, identifying patterns and risks associated with gender-based violence, or developing informational chatbots that can help crisis affected populations.[32] On an individual level, Generative AI tools can be mainstreamed by organizations for their employees to optimize their daily work in such things as creating documents and communication materials.[33]

Use Cases

Individual use cases may include generating research, summarizing operational reports, drafting standard documents, creating transcriptions and summaries from video reports or general meetings, and utilizing audio or visual tools to communicate with crises affected populations or other organizations. It can also provide greater opportunities for localization and helping local organizations in their activities and capacities.[34]

Generative AI tools can make repetitive tasks, that may be daunting for humans, easy. Their main utility is that they take large amounts of data and information to extract insight. A few examples:

1. Humanitarian and peacekeeping reports and data could be put into text-generation tools to extract key information for inclusion in the daily UN Operations and Crisis Centre (UNOCC) coordination meetings in New York. Such a change would greatly reduce the human burden of reading and summarizing several reports every day and shift the work to reviewing the summary before submission.

2. Media monitoring teams could use it to read all newswires, look for keywords, and draft an operational media update.

3. It could help to predict and fill in the metadata for Humanitarian Data Exchange (HDX) datasets.[35]

4. Help with labelling images or developing social media content.[36]

---

32 **United Nations Activities on Artificial Intelligence, ITU (2022),** Accessed July 2023

33 A beneficial source for humanitarians to explore use cases is **Nethope's evaluation of the usefulness of existing AI solutions in the Nonprofit sector (May 2023), Accessed July 2023**. Nethope envisions Generative AI applications for Nonprofits to constitute two main domains: One is creating efficiencies, growing income, and improving compliance across different organizational tasks. Second is supporting and enabling the organizations' mission across different sectors and crises situations.

34 **Four ways ChatGPT could help level the humanitarian playing field, The New Humanitaian (March 22, 2023),** Accessed July 2023

35 **Predicting Metadata for humanitarian datasets using GPT-3, Matthew Harris, (January 18, 2023),** Accessed July 2023

36 **Generative AI: A game-changer for humanitarian assistance, Jay Mahanand, (April 14, 2023),** Accessed June 2023

These opportunities have a huge scale factor for doing more as it allows staff to either complete their work more quickly and/or focus on other substantive work issues.
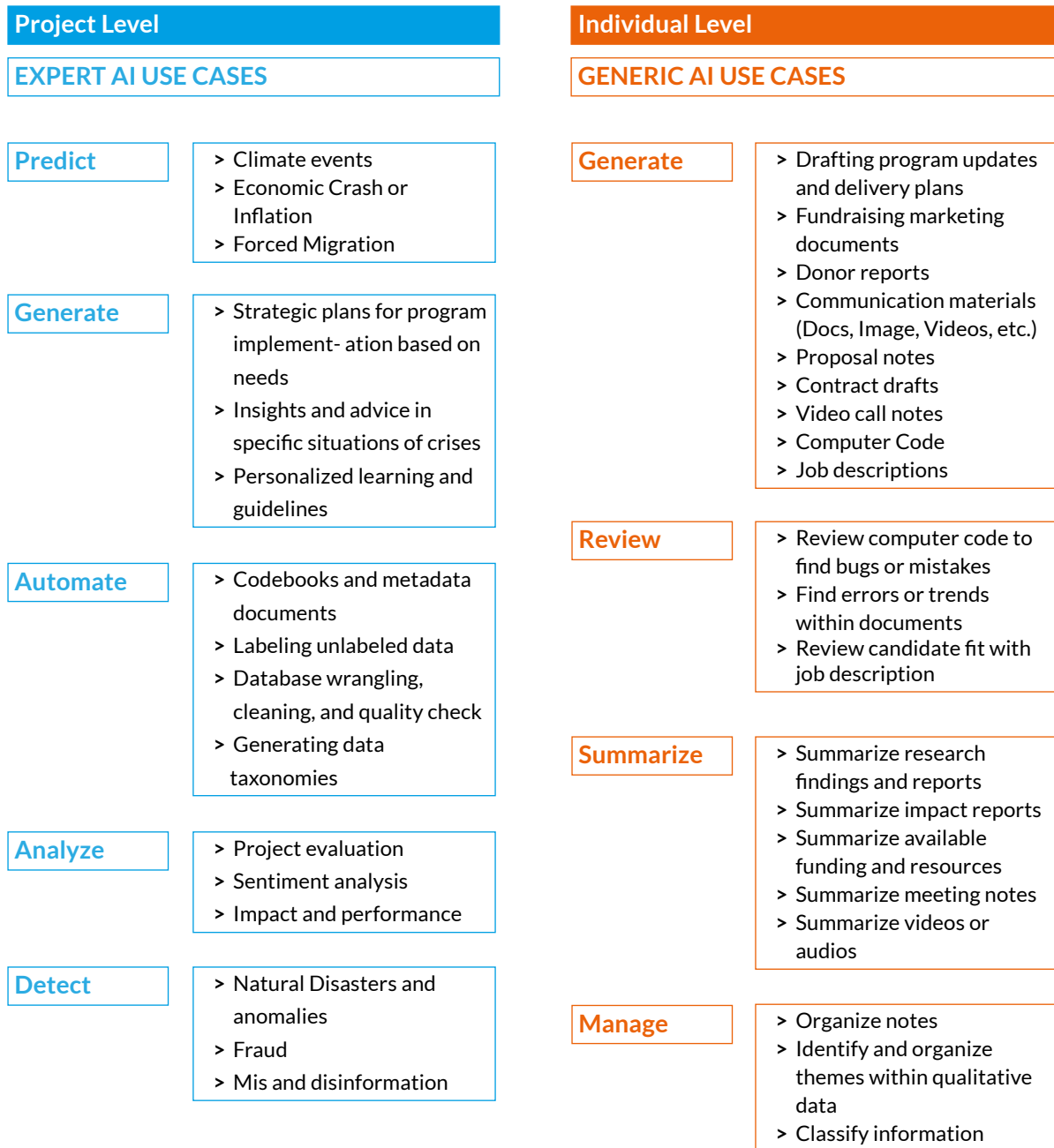
Project Level	**Individual Level**
EXPERT AI USE CASES	**GENERIC AI USE CASES**

Predict
> Climate events
> Economic Crash or Inflation
> Forced Migration

Generate
> Strategic plans for program implement- ation based on needs
> Insights and advice in specific situations of crises
> Personalized learning and guidelines

Automate
> Codebooks and metadata documents
> Labeling unlabeled data
> Database wrangling, cleaning, and quality check
> Generating data taxonomies

Analyze
> Project evaluation
> Sentiment analysis
> Impact and performance

Detect
> Natural Disasters and anomalies
> Fraud
> Mis and disinformation

Generate
> Drafting program updates and delivery plans
> Fundraising marketing documents
> Donor reports
> Communication materials (Docs, Image, Videos, etc.)
> Proposal notes
> Contract drafts
> Video call notes
> Computer Code
> Job descriptions

Review
> Review computer code to find bugs or mistakes
> Find errors or trends within documents
> Review candidate fit with job description

Summarize
> Summarize research findings and reports
> Summarize impact reports
> Summarize available funding and resources
> Summarize meeting notes
> Summarize videos or audios

Manage
> Organize notes
> Identify and organize themes within qualitative data
> Classify information

Figure 3: Generative AI sample use cases across expert AI and generic AI operations (Source: Authors).

The UN Generative AI practice group has also identified 6 main activity areas where Generative AI may be beneficial. Figure 4 shows a synthesis of ongoing topics suggested by the group for practical uses cases across the 6 categories:

1. **Enhancing Internal Efficiency**	**Cutting Down Admin Work**	Note Taking		
	Productivity	Creating Templates	Talking Points	Routine Processing work: Formatting, checklists
2. **Improving Reporting Processes**	**Monitoring UN mandates**	Matching SDGs to outcomes		
	Impact Reporting	Responding to Audits	Tackling reporting fatigue	Automatic report tagging
3. **Knowledge Management**	**Research**	Rapid research synthesis	Semantic search, searchable knowledge	Enterprise Chatbots
	Information Management	Responding to Audits	Tackling reporting fatigue	Automatic report tagging
4. **Designing visual elements**	**Videos**	Avatars	VR/AR Experiences	Social Media Content
	Images	Storyboards	Editing: Adobe Illustrator/Photoshop	Presentations
5. **Data Workstreams**	**Text Data**	Generating Taxonomy	Analyzing large sets of reports	Generating Metadata
	Visual Data	Indexing photo DB with CLIP	Image tagging	Layout designs
6. **Modeling and Prediction**	**Analyzing trends**	Analyzing patterns of narrative reporting on conflict drivers		
		Conflict prediction	Topic Modeling: understanding topics/ideas being discussed in dialogues, reports, etc.	

Figure 4: Use cases of Generative AI for UN Agencies. Adapted from the UN Generative AI practice group workshop on Transforming UN operations with AI

https://miro.com/app/board/uXjVMNA2nhI=/

Challenges of Adoption

Not only are there technological challenges in adopting this technology, but there are many cultural and policy related issues that need to be considered.

Fear Mongering

The AI hype comes with a "holly-wood like fear" that AI is the next superhuman and will take over jobs and humanity.[37] This AI fear mongering can develop societal concerns around job displacements which discourages participation and uptake of Generative AI. It also limits progress for developing regulations for responsible and effective use of these tools.

However, such sentiments only reinforce the role of AI tools as agentic experts and diminish the role of humans as the creative power and decision maker. The Big Tech's warning of an AI apocalypse is also a distraction, comparing it to a ghost story that hijacks attention from AI regulatory and ethical problems that need to be solved.[38] Such narratives also forget that AI tools are made by humans and therefore those developing them or using them should be held accountable.

Job Displacement

Generative AI is likely to bring significant changes to the job landscape over the next decade, it also presents opportunities for new kinds of work, enhanced productivity, and increased transparency. Mckinsey projects that current generative AI and associated technologies could automate tasks consuming 60-70% of an employee's working hours.[39] It is predicted that 60% of jobs will have some kind of 'robo colleague' that will highly augment the human (or perform a subset of tasks). By preparing for these changes, we can help to ensure a smooth transition into this new era of work conscious of the risks that Generative AI presents to short-term job displacement.

Goldman Sachs predicts that a new wave of AI systems may shift 300 million full time jobs by exposing them to automation. Generative AI tools have already presented mass disruptions in the global job market. For example, data analysts and software engineers are now facing competition from Generative AI tools[40] that are powerful in coding, debugging, and analyzing large unstructured data.[41] Based on the report published by OECD, jobs with the highest risk of being automated make up 27% of the labor force on average in OECD countries.[42] Therefore in the race to adopt generative AI tools we must recognize the digital inequalities that leave the majority of the world's population behind.

---

37   **What is Generative AI?, Pinar Seyhan Demirdag, LinkedIn Course, (March 15, 2023)**. Accessed July 2023

38   **AI Doomarism is a Decoy, Matteo Wong, The Atlantic, (June 2, 2023)**, Accessed July 2023

39   **The Economic Potential of Generative AI: The Next Productivity Frontier, Mckinsey, (June 14, 2023)**. Accessed July 2023

40   **Stack Overflow Will Embrace Generative AI After All. Will Overflow AI Address an Existential Risk?, Bret Kinsella (June 17, 2023)**, Accessed July 2023

41   **Analyzing Humanitarian Data Unstructured Excel Tables with ChatGPT Code Interpreter, Matthew Harris (July 10, 2023),** Accessed July 2023

42   **Employment outlook 2023, OECD,** Accessed July 2023

Legal & Regulatory Environment

Technology corporations and researchers have evidently prioritized novelty, efficiency, and performance in their AI advancements, barely tackling issues regarding privacy or bias.[43] Hence, governments and nonprofit organizations are often seen to be left behind in adopting technologies as they are busy addressing the regulatory challenges overlooked by Big Tech corporations.

> **"** *Biggest struggle for all organizations is governance. We are always playing catch up with operational capacities that exist in the private sector.* **"**
>
> — **Jesus Melendez Vicente, IREX**[44]

With the lack of regulatory policies for the development of Generative AI tools, governments have gained momentum to develop comprehensive regulations on the development of AI such as those seen in the EU Artificial Intelligence Act and the United States Federal AI Bill of Rights.[45] The UN is also discussing regulatory issues in various venues. For example the UN Secretariat is drafting an Administrative Instruction on mitigating AI risks. The UN Innovation Network has also established a Generative AI group to discuss immediate protection against misuse of AI tools and self-regulatory laws for mainstreaming the use of generative AI across the UN. And the International Telecommunication Union (ITU) has established AI-for-Good.[46]

Nonetheless there still remains gaps in understanding how humanitarian organizations can safeguard against such things as:

- **Infringement of data confidentiality:** Sharing sensitive data and information with external parties.

- **Ownership of shared data:** When shared with AI models, privately owned data can become company property, subject to expropriation, interference, or misuse-risking human rights violation.

- **Copyright infringement:** The use of AI generated outcomes that may be protected as intellectual property.

- **Fairness and Accountability:** Pertains to the responsibilities toward the misuse of AI tools. Raises issues of individual versus organizational accountability when used in the humanitarian sector.

---

43   **Science in the age of large language models, Birhane et al.,(April 26, 2023)**, Accessed July 2023

44   Interview with Jesus Vicente, IREX

45   These regulations have developed horizontal policies to control the risks of AI across different public sectors and prohibit specific applications that infringe the fundamentals of human rights. But according to **Brookings** they are also divergent on their AI policies.

46   **AI for Good, ITU**, Accessed July 2023

AI Literacy

AI literacy includes training and upskilling staff with Generative AI applications. With AI literacy, individuals can augment their capacities in the workforce while working on the soft skills that are necessary to foster creative and collaborative transformation. However, lack of AI knowledge and skills across the organization can cause technical silos and slow down technical explorations that are needed to advance the general Generative AI agenda.

Good data practices and data governance should be mainstreamed across all organizational levels from country offices, to regional teams, and leadership units to create specialized AI-related knowledge for evidence-based work. This includes training for data collection and formatting that is compatible with Generative AI applications and safeguards to ensure safe and responsible use of AI.

Leadership

Cohesive leadership for technology transitions is needed to enhance accountability, coordination, and transparency across all organizational levels in respect to the use of Generative AI tools. Good practice and AI literacy need to be mainstreamed across all leading positions to ensure decisions are educated and drive change across the organization.

Funding

> *There is a lack of incentives among staff [to use Generative AI] who are busy delivering programs that are funded by donors and trying to meet deadlines. They don't have time to explore new technologies.*
>
> **— Kimberly Coletti, formerly at Save the Children**[47]

With restricted funding, adopting new technologies and upskilling staff, especially across all levels of organizational operations, is challenging. There is a rising interest among I/NGOs to adopt Generative AI for data optimization and predictive analytics for crisis prevention or response. However, non profits prioritize working on ongoing funded programs as opposed to starting new ones without funding incentives. It is therefore important to address funding gaps[48] or find alternative ways to integrate Generative AI across organizational workflows that would not require large funds. For example, organizations can incentivize staff to start iterative explorations on existing data processes that are low risk and can benefit from public Generative AI training tools or open algorithms.

---

47   Interview with Kimberly Coletti, formerly at Save the Children
48   **AI and Humanitarianism - The possibilities and Pitfalls, Peter Sykes (2023),** Accessed July 2023

Guardrails for Safe and Ethical Use of Generative AI

Against the backdrop of adoption challenges, organizations and individual users should practice safe and ethical use of Generative AI tools. They should conduct fact checking practices and assess the generated outputs for inaccuracies, biases, and misalignment with organizational values such as those presented in CEB's Principals for the Ethical use of AI.[49] This entails a human in the loop approach that guides the Generative AI tools to produce a targeted outcome that follows ethical and legal standards.

We present 10 rules of thumb as a checklist for human oversight when using Generative AI tools to help users navigate the Generative AI landscape.

10 Rules of Thumb

1. **Justify your use of AI:** Know your goals and vision behind the AI generated media. For example, if it is meant for internal or external social media use.

2. **Choose standardized tools:** Prioritize choosing standardized Generative AI tools will help to protect data and privacy of organizations' data. If not available, follow the next steps to ensure safe and responsible use of public tools.

3. **Sign Up:** Do not use your organization email or credentials for setting up a profile with public Generative AI tools. This can cause privacy concerns as well as duplicability when AI tools become standardized.

4. **Effective Prompting:** Use prompt engineering when applicable to effectively communicate and define what the expected outcomes should look like. You need to be succinct and precise by following a chain-of-thought prompting technique.[50]

5. **Confidentiality:** Do not share any sensitive information with Generative AI tools. Ask for permission if the data is not your own, even if it is not sensitive.

6. **Synthetic Media Use:** Follow existing guidelines for using synthetic media such as those reflected in Partnership on AI's Responsible Practices for Synthetic Media.[51] Organizations can also develop internal policies around the production of communication materials (See Table 4, Appendix B).

7. **Sensitivity:** Ensure the AI generated media will be appropriate for the context and does not manipulate the audience or the data subject. Be sensitive towards the contextual and sentimental message that the generated media will carry.

8. **Biases:** Ensure outcomes do not hold any gender, racial, or ethnic biases. Run the generated outcomes through a regular clearance process and/or team leadership.[52]

9. **Transparency:** Be transparent of the use of Generative AI tools. Report what tools you used, how,

---

49   **Principles for the ethical use of Artificial Intelligence, UNSCEB**, Accessed July 2023
50   See Section 2, Prompt Engineering
51   **PAI's Responsible Practices for Synthetic Media, PAI (February 27, 2023)**, Accessed July 2023
52   For example, refer to **UNFPA's Guidance on the Safe and Ethical use of Technology to Address Gender-based Violence and Harmful Practices, (March 13, 2023)**, Accessed July, 2023

and what processes it took to generate the final published product. For example, use the APA format to cite the information (text or image) generated by a Generative AI tool (See Figure 8, Appendix B).

10. **Fact Checking:** Check generated outcomes for validity and accuracy. Report important findings if the outcomes fail to be accurate, ethical, or if you believe important information may have been shared. Although fact checking may be beneficial to spot inaccuracies, users may be unaware of their own biases and may not be able to recognize biases in AI generated outputs. It is critical to administer organizational processes such as review boards to ensure personal decisions reflect the values of the organization and humanitarian principles.

Drawing the Line: What is Acceptable Use?

Humanitarian organizations must delineate permitted and forbidden ways of using Generative AI tools for internal and external use cases. Table 1 outlines an example of acceptable use of Generative AI that can be further expanded to fit organizational needs and use cases.

Acceptable	Not Acceptable
Public data and information that is not considered high risk or confidential.	Information classified as confidential.
Summarizing published news reports and updates	Emails to or from officials.
Generating insight from non-confidential and low risk qualitative data,	Private meeting recordings or notes.
Editing texts, grammar checks, translations.	Internal documents and resources before public release.
Writing or debugging code.	Direct social media use is not recommended.
Searching travel routes and locations	Searching for factual information with no oversight.
Image generation based on guidelines and sensitivities towards social and political circumstances of use.	Population-level needs analysis or other use cases that hold high risk of data fabrication.

Table 1: Acceptable and non acceptable uses of Generative AI tools (Source: Authors)

Transparency and Accountability

Transparency constitutes efforts to describe, inspect, and reproduce the process of using Generative AI to achieve a goal. In order to enhance responsibility, accountability, and risk mitigation, transparency measures should be taken by individuals and enforced by organization-wide mechanisms to ensure good practice.

Individuals may fear or shy away from disclosing their use of Generative AI. Counterintuitively, users must divulge if they use any Generative AI tools for any part of their work so that readers as well as internal teams are aware of its use and can later assess its impact.

> **“** *The use of Generative AI tools should be burdened to prevent creation of incorrect materials. It's not bad to be burdened, in fact, it creates responsibility.* **”**

<div align="right">

— **Ravit Dotan**[53]

</div>

~~~~~~

Humanitarian aid workers and employees should be responsible for reporting the precautions taken to mitigate the risks regarding Generative AI as well as instances where Generative AI tools created inaccuracies or visible biases. These transparency efforts may burden the users, but it will ensure that practice of caution is enforced to minimize thoughtless use. For example, every image generated by AI should be captioned with the name of the tool that was generated by including the prompt.

But we must be aware of some of the challenges with deploying transparency measures. Soon Generative AI models will be integrated across all websites and tools to an extent that it will become invisible. As a case in point, Adobe Photoshop tools have already integrated Generative AI into their photo editing functions. Asking for clear processes of how tools were used and what outcomes they generated will be difficult and hard to track in such situations.

> **“** *Right now it is easy to think of AI generated text or imagery as discontinuous. That it is radically different from what we do now. Right now it is easy to talk about transparency. But as these tools become part of every aspect of our lives, every website, every tool like Photoshop, it will be difficult to apply transparency rules.* **”**

<div align="right">

— **Anonymous Interviewee**[54]

</div>

~~~~~~

---

53   Interview with Ravit Dotan, Collaborative AI Data Responsibility Lab
54   Anonymous Interview

Legal Safeguards

The use of Generative AI in the humanitarian sector necessitates a range of legal safeguards to address issues of data ownership, confidentiality, and accountability. A major focus should be placed on how humanitarian organizations can leverage these measures to enhance transparency, reduce inaccuracies, and manage potential biases arising from the usage of Generative AI tools.

Ownership of Shared Data: Legal protections for the utilization of Generative AI tools are solely applicable when governed by contractual agreements with technology corporations. These agreements serve to safeguard internal data shared with Generative AI tools, ensuring its exclusion from usage in training public models and relieving any legal obligations imposed by technology corporations.

Infringement of Data Confidentiality: Check for external contracts and legal regulations regarding data sharing with external partners, which do not have legal right to access internal information or data. Under such agreements, organizational data is protected and ownership rights are maintained over any shared data with AI models.

Accountability: The legal and regulatory frameworks are needed to ensure accountability and internal oversight for Generative AI processes. It is crucial to investigate ways in which humanitarian organziations can enhance transparency systems in order to detect and address inaccuracies and biases resulting from the inappropriate use of Generative AI tools. Additionally, accountability considerations arise in relation to agreements at both the individual and organizational levels during standardization processes. For instance, in organization contracts, it is important to determine which entity bears legal responsibility when internal data is shared with Generative AI tools.

Recommendations for Building an AI Ecosystem
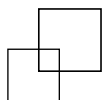
In the nonprofit community, organizations are facing what is NOT a technology problem, but rather an accountability and data governance problem.[55] Humanitarian organizations are struggling with the same open ended questions that if not answered, will slow down adoption of Generative AI tools. More exploratory efforts around responsibility and organizational mechanisms are needed to understand:

- How can we develop capacities and AI literacy for the responsible adoption of Generative AI in the humanitarian workforce?

- What mechanisms should be in place for regulating the use of Generative AI tools including sharing of information and using the AI generated content? Self-regulation is not always reliable.

- How do we tackle the lack of data diversity and data quality that perpetuates biases within the organization or hinders effective use of Generative AI tools?

- How can we assess the level of validity and accuracy of the AI generated outputs and can it be automated?

- How can we expand and ensure good practice across the humanitarian sector in the face of limited funding and technical expertise?

Setting up operational guidelines and guardrails for individuals as well as an AI ecosystem across humanitarian organizations will facilitate creating an enabling environment for the responsible uptake of Generative AI tools. In cooperation with tech companies as well as building internal technology experts, humanitarian organizations can benefit from the vast potential that Generative AI tools have to offer. However it is critical to share learnings and strategic plans for different pillars of an AI ecosystem including technology and data processes, governance, AI capacities, and assessment strategies.[56]

Developing an AI ecosystem within an organization can indeed be a complex task, especially with most advanced AI models being closed-access and under the control of a few leading tech companies. However, humanitarian organizations and their staff can still take action to address these challenges by establishing a socio-economic AI ecosystem (Figure 5). This ecosystem should include:

---

55   Interview with UNICEF information management team.
56   Interview with WFP, Information Management team.

1. **Data and Technology:** Selecting AI tools and developing the infrastructure that can respond to the changing organizational needs, be adopted for internal-only use, and be contracted under the organizational legal protection.

2. **Governance and AI Regulation:** Setting up operational guardrails and governance strategies to mitigate biases and misuse. This includes reinstating existing policies on information sharing with external parties and Information sensitivity, classification, and handling.

3. **AI and Data Upskilling:** Developing AI capacities for upskilling and conducting data processes (e.g., data wrangling) necessary for adopting new Generative AI tools within the organization.

4. **Evaluation and Improvement:** Establishing strategies for impact assessment and optimization of AI practices to ensure quality and validity of input and output data.

5. **Coordination and Collaboration:** Different teams should engage in knowledge sharing, resource management, and extended partnerships across donors, tech corporations, and other stakeholders.
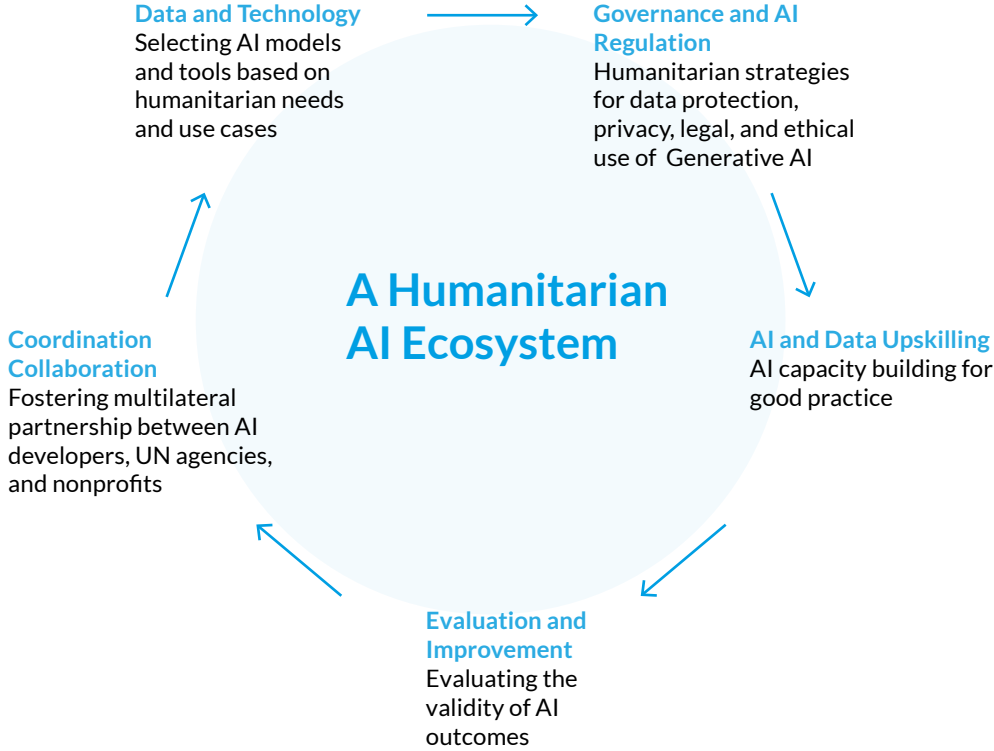

Data and Technology
Selecting AI models and tools based on humanitarian needs and use cases

Governance and AI Regulation
Humanitarian strategies for data protection, privacy, legal, and ethical use of Generative AI

A Humanitarian AI Ecosystem

AI and Data Upskilling
AI capacity building for good practice

Coordination Collaboration
Fostering multilateral partnership between AI developers, UN agencies, and nonprofits

Evaluation and Improvement
Evaluating the validity of AI outcomes

Figure 5: Recommended AI Ecosystem (Source: Authors)

Data and Technology: Sandboxing and Self Monitoring

Humanitarian organizations should initiate processes to evaluate the specific use cases as well as infrastructure necessary for mainstreaming Generative AI across their organizational functions.

The use case analyses should follow a model reflecting on the added value against the feasibility of deploying the tools (Figure 6).[57] Organizations can develop metrics for standardizing the assessment procedures based on their values especially around safety and ethics of AI. Using these metrics, standardized AI tools for specific internal needs and use cases can be developed to inform future development of AI tools that safeguard internal data and users. Individual users on the other hand should take precautions by following the 10 Rules of Thumb when working on internal data across public platforms. This means cutting through the AI hype by transitioning from a generic view of AI to one that is task specific.
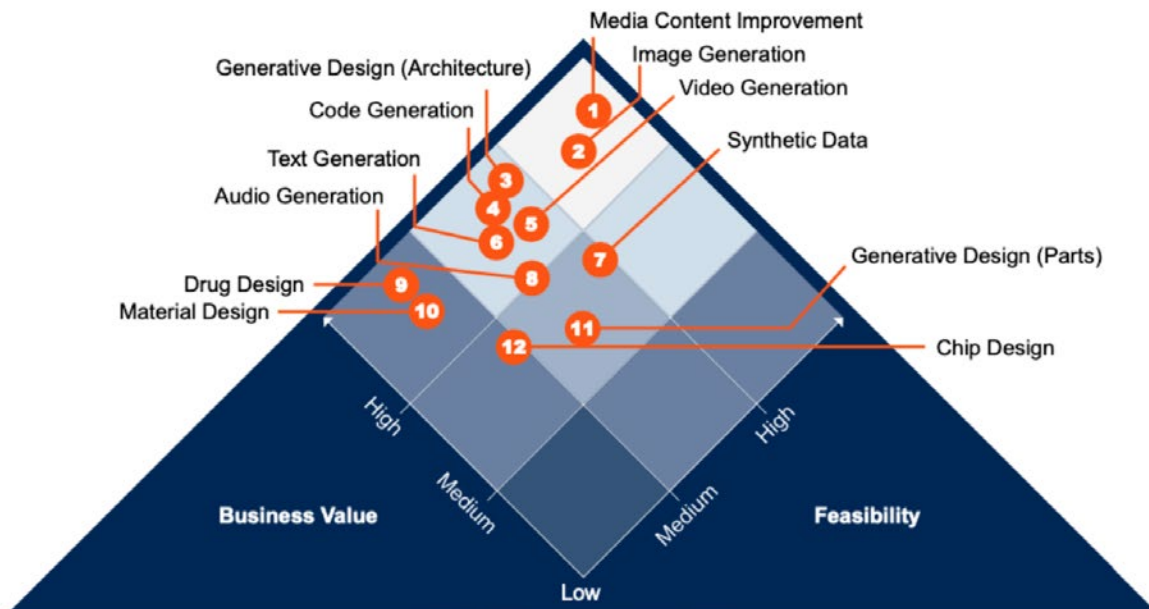

Figure 6: Use case prism for Generative AI (Source: Gartner Webinar Ask the experts: the impact of Generative AI such as ChatGPT)

Setting up a sandbox environment is a practical solution on an organizational level to provide a separate and isolated Generative AI space for safer internal workstreams. Within a sandbox, organizations can set up a separate and isolated environment where specific AI models can be trained, tested, and fine-tuned for particular tasks without posing risk to the operational systems or sensitive data. Setting up a sandbox will help to disentangle generality from specific Generative AI use cases and provide practical AI solutions based on those specific needs.

---

57   Ask the experts: the impact of Generative AI such as ChatGPT | Gartner Webinar, Accessed June 2023

> **"** *Generality cuts through two concepts – hype and safety. A hyped AI system is one that makes all the decisions and does all the coordination. For example, medical AI is an unsafe hyped image of an AI doctor. A less hyped AI is well contained and sandboxed. Like, looking at a CAT scan to detect a cancer that would likely be reviewed by a doctor.* **"**

— Shakir Mohamed, Google Deepmind[58]

Sandboxing can help tackle the issue of oversight in the AI sector. Since self-regulation is not always reliable, a sandbox environment allows organizations to enforce their own regulation and control over the AI tools they use. The process of adapting open-source models or fine-tuning pre-trained models can be technically challenging and resource-intensive. But for organizations without access to high-capacity training models, sandboxing provides an opportunity to develop custom use cases and offer services for fine-tuning.

Still, sandboxing is not a cure-all solution. A sandbox offers a safer way to diversify databases, reduce discrimination risks, and develop more accurate insight. However, ensuring access to quality data in a dataset can be a challenging task and does not automatically solve all bias or inaccuracy problems in AI applications. It requires conscious efforts for data cleaning and constant monitoring of the performance and impact of AI for iterating processes to optimize results. These processes are also critical for using expert AI systems for anticipatory action and predictive modelling.

Other than organization-level strategies, staff are advised to practice caution when approaching Generative AI tools and check for any inaccuracies, biases, and safety and privacy risks. By practicing human oversight and following the 10 Rules of Thumb (See Section 4) organizations can ensure their choice and use of Generative AI follows a process of self-monitoring strategies.

---

58   Interview with Shakir Mohamed, Google Deepmind

Governance and AI Regulations: Data Protection, Privacy, Ethical, and Legal Use of Generative AI

As noted by the United Nations Human Rights Office of the High Commissioner[59] and the New Technology Reference Group,[60] regulations are urgently needed to ensure transparency, alert people when they encounter synthetic media, and inform the public about the training data and models used. By deploying ethical standards and regulations, transparency measures can be integrated into internal oversight functions and control systems to produce reviews on the use of Generative AI tools, misconduct, and other data practices in respect to Generative AI .

Nevertheless, it is critical to roll out self-regulatory guardrails to employees as soon as possible - given that a large number are already using Generative AI tools in their daily workstream. Strategies and operational guidelines should reinstate the privacy and legal protection of organizations' data based on existing internal policies as well as strategies for internal oversight on ethics and staff integrity across all functions.

AI and Data Upskilling: Capacity Building for Good Practice

Acceleration in the production and use of Generative AI creates opportunities as well as challenges in adopting the technology across organizations. Leadership teams should formalize an AI working group that is responsible for internal AI operations including AI literacy and data governance. They should also account for responsibility towards global digital equality and support for local innovation and data processes that can enhance adoption.

By establishing effective communication and support systems, organizations can help employees and leaders to adapt and expand their skills regarding AI-driven processes and tools. Existing learning resources such as LinkedIn Learning can be utilized to advance and extend humanitarian applications in the realm of generative AI. Guidelines such as PAI's Responsible Practices for Synthetic Media[61] are also valuable to expand good practice within the organization.

Data governance is another critical element that can help improve the quality and diversity of data in different stages of information processes. From data collection, to data wrangling and labeling, new measures and standards can be implemented to prepare the data for Generative AI applications that are in accordance with principles of data security and lawful use of data.[62] Supporting key local actors and country officers to enhance quality of data and analytics for purposes of Generative AI applications could be another key exploration for organizations.

---

59 **New and emerging technologies need urgent oversight and robust transparency: UN experts, OHCHR, (June 2, 2023)**, Accessed July 2023

60 **UN Global Digital Compact, UN**, Accessed June 2023

61 **PAI's Responsible Practices for Synthetic Media, PAI (February 27, 2023)**, Accessed July 2023

62 **Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda, UN SDG, (November 2017)**, Accessed July 2023

> *“Generative AI is the new and the attractive, but we need to ask what are the skills that we need to use AI in our line of work more efficiency and effectively. What is the framework? We need to develop know-how for mid-level managers. For overseeing the development of programming that utilizes AI.”*
>
> — **Hovig Etyemezian, UNHCR Innovation**[63]

Evaluation and Improvement: Validity and Reliability of AI Generated Outputs

In light of the expanding adoption of Generative AI tools, it's essential to establish robust evaluation measures to ensure the validity of AI-generated outputs. This not only enhances the utility and reliability of the outputs but also builds trust within the organization.

For isolated instances, fact-checking and authenticating AI outputs is feasible and can be deployed on an individual level. For example, by providing guidelines for prompt engineering, organizations can help staff to adopt techniques for enhancing the reliability of Generative AI tools.[64] However, for an organization-wide deployment, automated processes to validate AI outcomes should be explored.

> *The issue is not whether you should use Generative AI or not, but rather how to determine whether the results are, to your satisfaction, valid?”*
>
> — **Anonymous interviewee**[65]

---

63   Anonymous interview

64   **Techniques to enhance reliability of Generative AI | Open AI**, Accessed July 2023

65   Anonymous interview

Assessing validity is highly organization-specific, requiring bespoke data processes based on the type of data and desired outcomes. Organizations can incorporate data observation strategies to parse their data and develop a baseline for fact checking. For organization-wide evaluation of Generative AI outputs, it is critical to investigate automating processes to validate data. But this requirement becomes more pressing given the current state of data. As a case in point, having used HDX as a test bed for tubular datasets in the humanitarian space, Matthew Harris at Datakind has found that the particular pattern is not quite ready for implementation with **LangChain**[66] mainly because the tool cannot parse some of the irregular tables seen in excel files uploaded in HDX. Utilizing Generative AI tools for organizational data requires experimentation and improvement of data qualities that can be used to prodice quality outputs.

Coordination Collaboration: Multilateral Partnerships for Advancing in AI

Developing coalitions, where tech companies, I/NGOs, UN agencies, and donors can come together can be incentivized to collectively frame the humanitarian problem statement, and develop capacities (technology skill, finances, etc.) to deploy solutions offered by Generative AI tools.

As opposed to competition and the race for the adoption of AI, non-profit organizations generally express their interest for collaboration across different organizations, donors, and tech companies to define problem statements and find solutions that can be translated across different organizations. For example, different organizations may struggle with the same data governance in the localization agenda. Sharing challenges and defining the unique problem statement can help resourcing technologies and expertise needed to solve the problems at hand. ITU's AI-or-Good provides such a space by bringing AI experts, academics, humanitarian organizations and UN agencies together. Humanitarian organizations should also consider exploring such pathways to expand their collaboration and conversations around the adoption of Generative AI.

---

66   Interview with Matthew Harris, Datakind

Conclusion
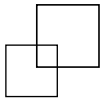
The integration of Generative AI into humanitarian operations signifies a transformative initiative with vast scope and potential. Everyday, humanitarians diligently tackle the strenuous tasks of the humanitarian response work chain. Generative AI presents an opportunity to augment their capacities for increased efficiency and reduced workload across organizational tasks. However, with that vast potential, a deliberate and thoughtful approach to AI adoption is critical. Our work underscores the need for a holistic strategy that weighs the immense opportunities provided by AI against potential risks.

We recommend cultivating an AI Ecosystem that fosters a responsible exploration and implementation of AI technologies in the humanitarian sector. This ecosystem encapsulates sandboxing of AI strategies, following rigorous standards for data handling and governance, learning and capacity building, evaluating and improving AI-related practices, and fostering a culture of collaboration across functions. The formation of a cross-functional team that is supported by external experts can provide an initial capacity to delve into this domain.

Deploying this AI ecosystem can support responsible AI usage, innovation and self-regulation in AI adoption across various use cases. Our recommendations can serve as a stepping stone for organizations as well as individuals to employ AI tools while maintaining their commitment to accountability, transparency, and ethics. Organizations can set up long-term strategies for responsible deployment of AI systems. And individuals can discern good practices for everyday use of AI tools.

This coupling between organizational and individual responsibilities is key to maintaining flexibility and adaptability in the face of the fast-changing AI landscape, while remaining unwavering in the commitment to the humanitarian principles. Therefore, ensuring humanitarians harness the power of AI to deliver more effective and efficient humanitarian assistance.

Annex A: Technical Details and Tools
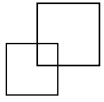
Foundation Models

Foundation models eliminate the need for labeling training data[67] and instead, find mathematical patterns between elements across the trillions of images, petabytes of text data on the web, and in corporate databases.[68] To gain access to such data, the models use data scrapers that crawl different websites and extract data to feed into the model for training.

These models are especially powerful in text-generation and are often utilized in chatbot applications that learn context, and thus meaning, by tracking relationships in sequential data like the words in a sentence. As the name suggests, these models can be the foundation for many applications of Generative AI models such as LLMs, Multimodal Text-to-Image Generation Models, GAN, Variational Auto-Encoders (VAE), or Diffusion models that can learn about one situation and refer it to another for continuous task operations.[69] By utilizing these models, a variety of Generative AI operations are made possible to train and generate different data types. Table 2 presents a summary of these models, a description of what they do, and their current applications.

Popular Generative AI Models	Explanation	Application
Natural Language Generators (NLG)	Utilizes LLMs such as transformer models that learn contextual relationships between words in a sentence or text sequence and generate natural language text that is coherent, fluent, and similar in style to existing or human-produced text.	> Chatbots > Language translation > Content generation

---

67   Large scale data classification and pattern matching that were commonly performed by **Deep Learning Algorithms** which ingest and process unstructured data such as text and images for feature extraction.

68   Before foundation models arrived, neural network models required to be trained with labeled datasets that were costly and time-consuming to produce.

69   **Diffusion models made easy | J. Rafid Siddiqui**

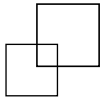| Popular Generative AI Models	Explanation	Application
Multimodal Text-to-Image Generation Models	Use a combination of LLM and generative tools such as GAN or diffusion models for text-to-image generation.	> Creating marketing ads > Film-making > Branding > Image completion
Generative Adversarial Networks (GAN)	Constitutes two neural networks, the generator and discriminator, which are in competition to generate and evaluate the generated outcome until it is not distinguishable from the original input data. The input and the output will be the same data type but it will provide many iterations of the design or output for users to choose	> Generate synthetic data > Film making > Iterative Design > Image inpainting > Scaling up resolutions
Variational Auto-Encoders (VAE)	Usually belong to part of the architecture of several other Generative AI models to help identify noise and data that are not normal or anomalies. VAEs are trained on normal data and are capable of identifying data that deviate from the normal data.	> Pattern Finding > Fraud detection > Anomaly detection > Denoising data > Medical Research > Quality control
Diffusion Models	Diffusion models are deep generative models that work by adding noise (Gaussian noise) to the available training data (the forward diffusion process) and then reversing the process (denoising or the reverse diffusion process) to recover the data. Similar to VAEs, Diffusion models can detect the noise in data but are also trained to remove the noise.	> Film Making > Text-to-Image generation > Inpainting and conditional image generation

Table 2: Popular Generative AI foundation models that are used in combination to support multimodal operations (Source: Authors)

Generative AI outputs, tools, and applications

Output Type	Mainstream Generative AI Tools	Applications
Text and Synthetic Data	> Anthropic Claude (Anthropic) > ChatGPT (Open AI) > Copilot (Microsoft) > Bard (Google) > Hugging Chat (Hugging Face) > Claude (Anthropic) > Github Copilot (Github)	> Document and qualitative data summarization > Code Generation > Translation and language learning > Research Report Generation > Content Editing > Scenario Development > Email and social media content
Image	> Midjourney (Midjourney) > Stable Diffusion (Stability AI) > Dall-E (Open AI)	> Marketing & Sales Content > Generating conceptual visuals from text > External branding and social media content
Audio	> Meta: MusicGen > ElvenLabs > Lovo.ai > Synthesys > Amper Music	> Creating voice clones > Text to voice generation > Music generation
Video	> Synthesia.io > Synthesys > Veed.io > Pictory > D-ID Video Generation > Deepbrain AI	> Video summarization > Video Transcriptions > Text-to-video generation > Video Translation > Virtual Humans > Deep Fakes

Table 3: Comprehensive Summary of different Generative AI outputs, tools, and use cases (Source: Authors).

Annex B: Guidelines for Creating and Publishing AI Generated Media

Output Data Type	Guidelines for Generating AI Outputs
Text	Sharing information and documents: > DO NOT share any sensitive information with any public text generation tools. > De-identify all information that is being shared with public tools. > Do not use text generation tools as search engines. Use of generated outputs: > Fact check for any inaccuracies in generated information or incorrect translations. > Re-read the output and check for language and tone. > Apply personal changes to ensure the creativity and impact of the output is protected. > Ensure transparency by reporting the use of AI. For example, use an APA format to cite the Generative AI tool (Figure 7) and the procedures applied for reaching the output.
Image	Visuals of organisation's senior leadership engaging with affected communities: > Clear the frame: no military, no bodyguards, no branded vehicle , only the principal and the people we serve. > Image setup must be at the same eye level. > No organization badge or lanyard should be visible. > Add a description of the image and what it is meant to imply. Visuals of the people we serve: > Focus on people, their faces, their situation, their stories. > Add a description of the image and what it is meant to imply. Social media: > Ensure the generated images comply with your social media templates.

Output Data Type	Guidelines for Generating AI Outputs
Video	Creating a Video Message: > Receive consent from the principal or video subject for generating AI video. > Include information about the field experience, the people, the actions that are being undertaken to address a given crisis, or a call to action. > Be cautious of using an AI bot in situations with emotional sensitivities that risks losing a connection with the audience. > Capture and report the process of video production, editing, and its frames. Transcribing or summarizing meeting recordings: > Do not input information from a sensitive meeting for summarization or transcriptions into public AI tools. > Inform meeting participants of utilizing AI tools for summarization or transcriptions. > Do not input internal video calls into public summarization tools.
Audio	How to craft a message for voice cloning or bots: > Ensure you use an AI generated voice or clone of anyone with permission. > Keep it short > Write your script with spoken words > Ensure your script discloses that it is a bot > Ensure the generated voice is based on your organization's standards.

*Table 4: Guidelines for creating and publishing AI generated media.*

For a full citation in the reference list, follow the formula below used for other algorithm-generated information. Pay close attention to punctuation, such as periods, parentheses, and brackets, and where you put them.

Company. (Year). *AI Name* (version) [Descriptor]. URL

OpenAI. (2023). *ChatGPT* (Mar 14 version) [Large language model]. https://chat.openai.com/chat

Figure 7: API citation format for AI generated information.

Annex C: Bibliography
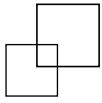
Accenture. (March 22, 2023). A new era of Generative AI for everyone. https://www.accenture.com/us-en/insights/technology/generative-ai Accessed July 2023.

Beko. (2021, March 29). Beko harnesses the evocative power of AI motion pictures. https://www.prnewswire.com/news-releases/beko-harnesses-the-evocative-power-of-ai-motion-pictures-301256863.html Accessed July 2023.

Birhane, A., Kasirzadeh, A., Leslie, D., & Wachter, S. (2023). Science in the age of large language models. Nature Reviews Physics, 1-4. https://www.nature.com/articles/s42254-023-00581-4 Accessed July 2023.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712. https://arxiv.org/pdf/2303.12712.pdf?utm_source=webtekno Accessed July 2023.

DELL Technologies. (2023, May). Generative AI in the enterprise. https://infohub.delltechnologies.com/l/generative-ai-in-the-enterprise/definition-and-overview Accessed July 2023.

Demirdag, P.S. (2023, March 15). What is Generative AI? LinkedIn Course. https://www.linkedin.com/learning/what-is-generative-ai?contextUrn=urn%3Ali%3AlyndaLearningPath%3A6477ad5c498e72fec31ca876 Accessed July 2023.

Europa.Europa.eu. (2023, June 14). MEPs ready to negotiate - First ever rules for safe and transparent AI. https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai Accessed July 2023.

Gartner Webinar. (n.d.). Ask the experts: the impact of Generative AI such as ChatGPT. https://www.gartner.com/en/webinar/474424/1119021 Accessed June 2023.

Gartner. (January 26, 2023). Beyond ChatGPT: The future of Generative AI for enterprises. https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises Accessed June 2023.

Google, (April 04, 2022). PaLM scaling 540 Billion parameters for breakthrough performance. https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html Accessed July 2023

Google. Bard Privacy Notice. https://support.google.com/bard/answer/13594961?hl=en Accessed July 2023.

Harris, M. (2023, January 18). Predicting Metadata for humanitarian datasets using GPT-3. Towards Data Science. https://towardsdatascience.com/predicting-metadata-for-humanitarian-datasets-using-gpt-3-b104be17716d Accessed July 2023.

Harris, M. (2023, July 10). Analyzing Humanitarian Data Unstructured Excel Tables with ChatGPT Code Interpreter. Towards Data Science. https://medium.com/towards-data-science/analyzing-humanitarian-data-unstructured-excel-tables-with-chatgpt-code-interpreter-ccb8cabfa146 Accessed July 2023.

Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. Video diffusion models. 4 2022. doi: 10.48550. arXiv preprint arxiv.2204.03458. https://arxiv.org/abs/2204.03458 Accessed July 2023

IBM, (August 02, 2022). Entering the era of AI-powered employees. https://research.ibm.com/blog/ai-powered-digital-employees Accessed July 2023

ITU. (2022). United Nations Activities on Artificial Intelligence. https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2022-PDF-E.pdf Accessed July 2023.

ITU. AI for Good. https://aiforgood.itu.int/ Accessed July 2023.

Kinsella, B. (2023, June 17). Stack Overflow Will Embrace Generative AI After All. Will Overflow AI Address an Existential Risk? Synthedia. https://synthedia.substack.com/p/stack-overflow-will-embrace-generative Accessed July 2023.

Mahanand, J. (2023, April 14). Generative AI: A game-changer for humanitarian assistance. Towards Data Science. https://towardsdatascience.com/predicting-metadata-for-humanitarian-datasets-using-gpt-3-b104be17716d Accessed June 2023.

Mckinsey. (2023, June 14). The Economic Potential of Generative AI: The Next Productivity Frontier. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#business-value Accessed July 2023.

MIT Technology Review. (2023, June 1). Welcome to the new surreal. How AI-generated video is changing film. https://www.technologyreview.com/2023/06/01/1073858/surreal-ai-generative-video-changing-film/ Accessed July 2023.

Nethope. (2023, May). Nethope's evaluation of the usefulness of existing AI solutions in the Nonprofit sector. https://nethope.org/toolkits/the-guide-to-usefulness-of-existing-ai-solutions-in-nonprofit-organizations/ Accessed July 2023.

OECD. Employment outlook 2023. https://www.oecd.org/employment-outlook/2023/#ai-jobs Accessed July 2023.

OHCHR. (2023, June 2). New and emerging technologies need urgent oversight and robust transparency: UN experts. https://www.ohchr.org/en/press-releases/2023/06/new-and-emerging-technologies-need-urgent-oversight-and-robust-transparency Accessed July 2023.

Open AI. Open AI privacy policy. https://openai.com/policies/privacy-policy Accessed July 2023.

Open AI. Techniques to enhance reliability of Generative AI. http://openai-cookbook/techniques_to_improve_reliability.md%20at%20main%20%C2%B7%20openai/openai-cookbook%20%C2%B7%20GitHub Accessed July 2023.

Partnership AI. (2023, February 27). PAI's Responsible Practices for Synthetic Media. https://syntheticmedia.partnershiponai.org/#read_the_framework Accessed July 2023.

Roane, A. (2023, March 22). Four ways ChatGPT could help level the humanitarian playing field. The New Humanitarian. https://www.thenewhumanitarian.org/opinion/2023/03/20/ways-chatgpt-could-help-humanitarian-field Accessed July 2023.

Sykes, P. (2023). AI and Humanitarianism - The possibilities and Pitfalls. https://www.linkedin.com/feed/update/urn:li:activity:7085294444068970499/ Accessed July 2023.

The Verge. (2022, June 15). Voicemod uses AI to transform your voice into Morgan Freeman. https://www.theverge.com/2022/6/15/23169012/voicemod-morgan-freeman-ai-voice-astronauts-pilots Accessed July 2023.

The Verge. (2022, September 15). Anyone can use this AI generator-that's the risk. https://www.theverge.com/2022/9/15/23340673/ai-image-generation-stable-diffusion-explained-ethics-copyright-data Accessed July 2023.

UN SDG. (2017, November). Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda. https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda Accessed July 2023.

UN. UN Global Digital Compact. https://www.un.org/techenvoy/global-digital-compact Accessed June 2023.

UNFPA. (2023, March 13). UNFPA's Guidance on the Safe and Ethical use of Technology to Address Gender-based Violence and Harmful Practices. https://www.unfpa.org/publications/implementation-summary-safe-ethical-use-technology-gbv-harmful-practices Accessed July 2023.

UNSCEB. Principles for the ethical use of Artificial Intelligence. https://unsceb.org/principles-ethical-use-artificial-intelligence-united-nations-system Accessed July 2023.

Veritone.voice. (2023, January 24). Everything to know about deepfake voice. **https://www.veritonevoice.com/blog/everything-you-need-to-know-about-deepfake-voice/** Accessed July 2023.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824-24837. **https://arxiv.org/abs/2201.11903** Accessed July 2023.

Wong, M. (2023, June 2). AI Doomarism is a Decoy. The Atlantic. **https://www.theatlantic.com/technology/archive/2023/06/ai-regulation-sam-altman-bill-gates/674278/** Accessed July 2023.

Zapier. (2023, June 14). 5 security risks of Generative AI and how to prepare for them. **https://zapier.com/blog/ai-security-risks/** Accessed July 2023.