CAIRO INSTITUTE
FOR HUMAN RIGHTS STUDIES
Institut du Caire pour les études des droits de l'homme
مركز القاهرة لدراسات حقوق الإنسان

رواق عربي
دورية محكمة
ROWAQ ARABI

ISSN: 2788-8037
Publication details, including guidelines for submissions:
https://rowaq.cihrs.org/submissions/?lang=en

Views: ChatGPT as a Human Rights Defender

Hakim Abdelnaeem

To cite this article: Abdelnaeem, Hakim (2023) 'Views: ChatGPT as a Human Rights Defender,' *Rowaq Arabi* 28 (3), pp. 5-12, DOI: 10.53833/BUYJ4616

To link to this article: https://doi.org/10.53833/BUYJ4616

Disclaimer
This article may be used for research, teaching and study purposes, as long as it is properly referred to. The Rowaq Arabi editors make every effort to ensure the accuracy of all the information contained in the journal. However, the editors and the Cairo Institute for Human Rights Studies make no representations or warranties whatsoever as to the accuracy, completeness or suitability for any purpose of the content. Any views expressed in this publication are the views of the authors and not necessarily the views of the editors of Rowaq Arabi or the Cairo Institute for Human Rights Studies.

Copyright
This content is published under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Licence.

cc BY NC

Views: ChatGPT as a Human Rights Defender

Hakim Abdelnaeem

Keywords: Artificial Intelligence, ChatGPT, Human Rights, Arab Region, Critical Discourse Analysis

A handful of investors founded OpenAI in 2015, among them Elon Musk, the CEO of several other companies, and Sam Altman, the current CEO of OpenAI. As a non-profit organisation established to develop and promote artificial intelligence research and systems, OpenAI aims to freely collaborate with other institutions and researchers by making open-source patents and research available to the public; the organisation also seeks to address the risks of artificial intelligence. In 2018, Musk resigned from OpenAI's board, citing a potential conflict of interest given his role as the CEO of Tesla Motors, which is involved in AI development for self-driving cars, but he remained a financer of the organisation. In 2019, OpenAI accepted a $1 billion investment from Microsoft, one of the world's leading tech companies, and on 23 January 2023, Microsoft invested $10 billion in the organisation.

ChatGPT, one of OpenAI's most significant contributions, is a large language model-based chatbot that generates texts using generative pre-trained transformer models (GPT), which aims to create natural language texts in a style that mimics human writing. The tool is trained on a massive, diverse database of texts to learn grammar, context, and various linguistic styles.

In March 2023, OpenAI released GPT-4, the latest version of its chatbot, which had seen a tremendous surge in popularity since early 2023 and has revolutionised AI-based applications. While the free version of ChatGPT operates using the GPT-3.5 language model, the new version, which is now available to ChatGPT Plus subscribers, offers much greater capabilities, including the ability to analyse images and respond to related questions. It should be noted that some proportion of the information on ChatGPT (public version) is outdated because its knowledge base stops at January 2022. Most developments or changes after this date are not incorporated in the application's memory, but it still has a tremendous ability to perform all the tasks asked of it.

The year that GPT-4 was released, specifically in September 2023, the Human Rights Education Unit (HREU) of the Cairo Institute for Human Rights Studies (CIHRS),[1] where I am responsible for programming, launched a new training program for human rights defenders: Tamkeen–Critical Insights on Human Rights.[2] The programme allows participants from the Arab region to develop the knowledge and skills acquired from their human rights work through the philosophical, political, and practical immersion in human rights issues in the region. When the programme was first launched, the Education Unit received hundreds of applications from human rights defenders from Morocco, Algeria, Tunisia, Libya, Egypt, Sudan, Palestine, Syria, Lebanon, Iraq, and Yemen.

In the process of reading and reviewing the applications for the first round of vetting, the HREU staff observed the same formal and compositional patterns in answers to questions on a few of the applications, and at times found even identical responses. More suspicious was the arrival of new applications exhibiting the same formal and compositional patterns and again containing answers quite similar to those on other applications. The responses were similar in their language, theses, and the ideas discussed, which exhibited no clear authorial point of view and were notably information-heavy and formulaic despite the question being crafted to give respondents free latitude to express their ideas. When another set of applications was received that contained these similar types of responses, a member of the HREU team confirmed that not only were the applications edited using the ChatGPT chatbot, which to some extent is not problematic, they were moreover written entirely by ChatGPT, save for the personal data. This raises a fundamental question that I attempt to explore here: Can human rights defenders formulate their thoughts on complex political challenges with the support of artificial intelligence?

In seeking to answer this key question, we must try to find an appropriate framework to understand the complexities of the political challenges confronting human rights defenders, starting from the previous question. Perhaps the most appropriate theoretical framework is critical discourse analysis, which is geared towards understanding language and discourse through the social, cultural, political, and historical contexts in which it occurs, showing the influence of these contexts on the use of language and power relations.[3] Applying this theory to our question will give us insights into the multifaceted nature of that process. Critical discourse analysis will also help us analyse the discourse strategies used by human rights defenders in specific contexts, highlighting how language shapes their understanding of political challenges and their ability to respond effectively. Critical discourse analysis has been the research focus of several prominent scholars. In his renowned *Discourse and Social Change*, considered an authoritative reference for studies on discourse, power, and social context, Norman Fairclough takes an in-depth look at how language and discourse influence the formation of power and authority in different societies and cultures, and he discusses how language and discourse are used to convey meaning and construct cultural and social identities. The book focuses on the relationship between language and power, and how to use discourse to have an impact in diverse social contexts.[4]

Analysing critical discourse with a focus on context will allow us to assess how AI-based technologies affect human rights defenders' discourse and to understand the interaction between AI-generated content and human rights discourse in specific contexts, taking into account the ethical concerns and power dynamics associated with AI support.

ChatGPT and the Role of Context in Creating Meaning

ChatGPT is based on generative pre-trained transformer technology, which is a model for generating natural language used in the analysis and production of texts. It works through five basic stages. I asked ChatGPT a question about how it works in general and the details of each stage, and the chatbot gave the following meticulous response:

1) Pre-training: ChatGPT is trained on a large, diverse set of texts taken from the Internet. In this stage, it learns linguistic relationships, grammar, vocabulary, and various linguistic patterns.
2) Fine tuning: After initial training, ChatGPT is modified using smaller, domain-specific datasets, such as responding to medical questions or generating ads. This phase helps the model customise its responses and understand users' needs.
3) Interaction and production: When ChatGPT receives a text or question from a user, the model analyses and understands the text using the knowledge acquired from its training. It then produces a logical response based on the context and information present in the given text.
4) Text generation: ChatGPT uses the methods learned during training to generate new texts. If asked to complete a sentence or write a paragraph, it uses its understanding of language and meaning to generate an appropriate, natural text.
5) 5. Following directions: Users can guide ChatGPT by providing more specific instructions or directions. This guidance can include requesting specific information or confirming certain facts.

The above points are typical of ChatGPT's responses to a user who asks a specific question. It is important to note that ChatGPT can sometimes provide inaccurate or illogical responses due to its method of training, which relies on already existing data. Continuous follow-up and improvement of the models used may be needed to ensure optimal performance. Understanding the nature of potential errors makes it easier to ensure that the system is periodically upgraded, thus ensuring a better user experience.

On a personal level, my experience with AI applications has led to multiple discoveries that spurred me towards greater engagement with the field, especially applications that generate text, images, sounds, and all manner of symbolic production. Although AI technologies like ChatGPT could potentially produce highly accurate text and audio-visual content, they still have limitations that make it difficult, if not impossible, for them to replace humans. ChatGPT and similar applications can largely understand texts, but they cannot understand context and meaning in the same profound way that humans do, which can sometimes lead them to generate inaccurate content. In addition, any creative product requires the kind of deep reflection and creativity that is still beyond the capabilities of this type of technology. I also suspect that effective communication with humans requires the ability to interact and respond dynamically according to context, and that ability is still not fully operational in any existing AI system.

It is therefore unlikely that AI will replace the work of artists, creators, thinkers, and writers. It is designed primarily to facilitate work in all these fields, acting like a capable assistant who cannot perform to the fullest unless provided with adequate information and details; otherwise, the results will be unsatisfactory. This field—the issue of prompts—is in fact the key area of focus for work on ChatGPT thus far.

Prompt Chaos and Accuracy

It should be understood that assertions that AI applications in general, and ChatGPT in particular, are not wholly accurate and at times offer false and completely arbitrary responses are largely correct. As already mentioned, the key point in using these applications is prompts, the detailed inputs and information that the user enters in order to obtain a written text or piece of information. The application is not designed to answer abstract questions or do any and every thing. Accordingly, when using the application, specificity and precision are indispensable. What we might call ‘prompt chaos’ is the chief cause of disappointing results for many users. Users can only obtain the desired results through a precise, diligent engagement with the chatbot; in the end, it is operated by the user, not vice-versa. For example, if we ask ChatGPT, ‘How is the human rights situation in the Arab region affected by developments on the regional and international landscape?’ using formulations of varied structure and complexity, how would it answer? Will the answers be satisfactory, or must they be carefully edited afterwards? In the following example, I entered the data for ChatGPT and received responses that I present here without additional editing.

Question 1: How is the human rights situation in the Arab region affected by developments on the regional and international landscape?

Below is the first response given by ChatGPT. It should be remembered that ChatGPT has not been updated since 2021, and the Arabic version of the application also suffers from linguistic and contextual problems at times. Another issue complicating the work of researchers is that sometimes an AI system may use concepts that humans do not understand and have no labels for. Because of this, ‘There is little way of knowing what biases might be involved in the system, or if it is providing false information to people using it, since there is no way to of knowing how it came to the conclusions it did’.[5]

Response to Question 1

The human rights situation in the Arab region is significantly affected by developments on the regional and international landscape. Many factors and events play a role in this impact, including:

1) Regional conflicts: Conflicts such as the Israeli-Palestinian conflict and civil wars in Syria, Yemen, and Libya have caused significant human rights violations. Civilians in these areas are being killed, displaced, and arbitrarily arrested.
2) International politics: Major states' decisions and interventions in the region affect the human rights situation. For example, the impact of economic and military sanctions on the civilian population can be significant.
3) Wars and conflicts: The impact of wars and conflicts on the civilian population in the region can be destructive. This includes aerial attacks, the use of chemical and biological weapons, and sexual violations and sexual violence.
4) Political violations: Some countries suppress political freedoms and expression of opinion, which negatively affects human rights. Violations such as the arrest of activists and journalists and the banning of political parties diminish fundamental freedoms.
5) Refugees and displaced persons: Conflicts and wars are forcing many people to flee their homes. These refugees and displaced persons face significant challenges in relation to their economic, social, and political rights.
6) International intervention: Interference by other states in state affairs can exacerbate conflicts and undermine human rights. For example, external support to warring parties can increase the frequency of conflicts and increase instability.
7) Economic changes: Economic conditions in states can have a significant impact on human rights. Economic decline and high unemployment rates can exacerbate poverty and worsen the standard of living.

In addition, international institutions such as the United Nations and international human rights organisations play a role in monitoring and reporting on the human rights situation in the region and help try to improve the situation. Nevertheless, the great challenges facing human rights in the Arab region persist, and joint international and regional efforts are needed to overcome them.

The response here is highly organised and formulaic, similar to the style a schoolteacher in the Arab region may recommend to their students. It starts with an introduction containing the second half of the question but in the form of a nominative sentence, and then presents the answer in bullet-point form with subheadings for each point, followed by a brief paragraph explaining them. I have noticed, with my repeated use of ChatGPT, that this is a recurrent pattern of response in Arabic. In addition, the response contains clear political terms that have been grafted onto the information and details stored in the application, such as 'the Israeli-Palestinian conflict', 'refugees and displaced persons', 'interventions in the region affect the human rights situation'. Nevertheless, the response is superficial and non-analytical, and does not rise to the level of an answer that should be provided on an application to participate in a training programme to improve the analytical and political capacities of human rights defenders in various fields. But what if we try to hone the question itself by adding more specific contextual details and additional information?

Question 2: How is the human rights situation in the Arab region affected by developments on the regional and international landscape given issues raised by democratisation in the region and the failure of the Arab revolutions?

The following answer was generated in the same window as the first. ChatGPT engages in self-learning, continuously developing its responses based on the information provided and arranging them contextually to offer responses that evolve as the context of the questions do; they are not wholly decontextualised responses. Two responses were generated here. The first is similar to the previous one, albeit with different subheadings and phrasing, and clearly benefited from the additions I made to my question. The second response is more precise.

New Response to Question 1

The human rights situation in the Arab region is significantly affected by developments on the regional and international landscape, particularly given the issues raised by democratisation and the failure of the Arab revolutions. Many factors contribute to this impact:

Foreign intervention: Foreign intervention in the domestic affairs of Arab states may undermine democratisation processes and exacerbate conflicts. For example, some states' support for repressive governmental regimes can contribute to human rights violations.

Regional conflicts: Regional conflicts and tensions have a heavy impact on the human rights situation. For example, the impact of the Israeli-Palestinian conflict on Palestinians and tensions in the region leads to persistent human rights violations.

Economic failure: The deterioration of the economic situation in several Arab states can fuel poverty and social inequality, having a negative impact on economic and social rights.

Suppression of political freedoms: In many Arab states, freedom of expression, assembly, and political participation is suppressed, which curtails human rights and hinders progress towards democracy.

Instability: Political and security instability can make it difficult to achieve progress on human rights. Persistent turmoil makes it difficult to implement effective democratic reforms.

Weak institutions: In many Arab states, democratic institutions suffer from weakness and corruption, making it difficult to achieve progress on human rights.

Culture and values: Difference in culture and values in the Arab region can impact human rights. Some values and traditions may pose a challenge to advances in women's rights and the rights of minorities.

As a whole, the human rights situation in the Arab region is affected by a complex interaction of several factors and requires coordinated international and regional efforts to promote democracy and human rights in this region.

Response to Question 2

The human rights situation in the Arab region is significantly affected by developments on the regional and international landscape, particularly given the issues raised by democratisation and the failure of the Arab revolutions. Here are some of the ways these factors can have an impact:

The exacerbation of humanitarian crises: The failure of the Arab revolutions and the eruption of conflicts and wars have exacerbated humanitarian crises in the region. The number of refugees and internally displaced persons has increased, and wars have destroyed infrastructure, schools, and hospitals, negatively impacting the human rights of many.

Rampant grave violations: In light of instability and conflicts, grave human rights violations are rampant, including ethnic killings, arbitrary arrests, and torture. Governments and armed groups exploit chaotic situations to engage in repression and retaliation.

Suppression of political freedoms: Widespread conflicts and instability increase the likelihood of violations of freedom of opinion and expression. Governments may use security crises as a pretext to curtail political freedoms and harass dissidents.

Foreign intervention: The influence of other states and international institutions in the Arab region can significantly affect the human rights situation. Historically, some states have supported regimes regardless of their human rights records, which can encourage the persistence of abuses.

Weak democratic institutions: Many Arab states face challenges in building strong, stable democratic institutions. The poor separation of powers and minimal control over security and army forces can lead to the arbitrary abuse of power by governments.

Declining economic and social rights: Human rights have also been affected by deteriorating economic and social conditions in some states, where poverty rates and the cost of living have increased.

International monitoring and reporting: International and human rights organisations monitor, document, and report on human rights violations in the region. These reports can pressure governments to improve the human rights situation.

In short, developments on the regional and international landscape in the Arab region play an important role in impacting the human rights situation. The challenges of democratisation and the failure of revolutions make them more impactful and make it difficult to improve the human rights situation in some places.

In the previous example, by adding one sentence clarifying the topical context, geographic location, and a main feature of the current moment in the Arab region, we get a more precise, controlled response that attempts to create some space for the analysis of the information presented rather than simply convey it. With the repetition of the process, ChatGPT begins to process the information given to it, honing its style to reflect the quality and accuracy of the informational prompts given to it. Its expression also becomes more contextually sophisticated as the process is repeated. It is more like a talented, encyclopaedic assistant who is quick to learn than a lead researcher or author.

Finally, after completing this experiment with ChatGPT, it is clear to me that critical discourse analysis can enhance understanding of the conversations and information presented. Here, ChatGPT exhibits a notable capacity to grasp the critical context accurately and flexibly, making it a good tool for understanding details and sophisticated discursive content. An understanding of critical discourse analysis may enable users to draw out main ideas and key points, helping to maximise the benefits of information sharing. In addition, ChatGPT shows an ability to adapt to different contexts and understand linguistic and cultural symbols, which increases its value as a powerful and useful analytical tool in interpersonal communication. While much research is still needed, at the very least, we should be aware that technological developments and AI applications present new opportunities to support human rights defenders in analysing and understanding political contexts. Even so, we cannot as yet say that artificial intelligence will replace humans in this field. The capacity to engage with context and understand social and political challenges in a sophisticated way remains a skill in which artificial intelligence is outmatched by the human mind. Nevertheless, by combining the efforts of human rights defenders and technology, including AI, I see a clear prospect on the horizon for maximising benefits and better responding to the complex challenges of the contemporary political world.

About the author

Hakim Abdelnaeem is the programme officer for the Human Rights Education Unit of CIHRS, a multidisciplinary artist, and a writer interested in issues related to the intersection of arts and culture with the social sciences and humanities.

---

[1] CIHRS is an independent, regional non-governmental organisation founded in 1993. See: https://cihrs.org.

[2] For more about the program, see: https://cihrs.org/tamkeen-critical-insights-on-human-rights/.

[3] Paltridge, Brian (2006) *Discourse Analysis: An Introduction* (London: Continuum).

[4] Fairclough, Norman (1992) *Discourse and Social Change* (Cambridge: Polity Press).

[5] Griffin, Andrew (2023) 'ChatGPT Creators Try to Use Artificial Intelligence to Explain Itself—and Come Across Major Problems', *Independent*, 12 May, accessed 5 November 2023, https://www.independent.co.uk/tech/chatgpt-website-openai-artificial-intelligence-b2337503.html.