Fulfilling GEN-der AIms: do image-generating tools discriminate? An on-field study

Francesco Meledandri
University of Bari “Aldo Moro”, Italy – francesco.meledandri@uniba.it

ABSTRACT (ENGLISH)

Recently, the digital turn entered a new era as some technological tools proved to represent a breakthrough in virtual and real-life practices. Though the idea of Artificial Intelligence (AI) has been circulating since the late 1950s (McCarthy et al., 1955) it has only been part of everyone’s lives for a few years, since users had the possibility to create language and multimedia contents from scratch. Nowadays the most promising application is probably Generative AI (GenAI), which generates “synthetic data that closely resemble real-world data” (Bandi, 2023: 1). Trained upon LLMs (Large Language Models) developed by huge stakeholders (Dao, 2023), AI is now closer to the idea of “human reasoning” as it generates contents based on human-induced prompts. Ethical issues are paramount to the discussion of AI outputs (Dubber et al., 2020; Boddington, 2023). As a matter of fact, GenAI creates plausible outputs, though they are not always *true* since they rely upon the *synthetic* gathering of corpus-based data, thus leading to hallucinations (Ji et al., 2023). Another critical issue involves gender-related depictions, showing a disparity that leads to a proper machine-induced bias (Leavy, 2018; Foka, 2024). Against this background, this paper aims at providing evidence from specific AI-powered generative tools that create realistic images from textual prompts, thus carrying out an intersemiotic translation process (Dusi, 2015). In particular, providing different easy-to-use GenAI platforms (including the newly-introduced Grok featured in *Twitter/X*) with ‘neutral’ prompts, the analysis would assess the level of possible equality (or inequality) in terms of gender, thus assessing the level of potential bias that may influence the perception of users in terms of visual narratives (Chen et al., 2024).

Keywords: Artificial Intelligence; Equality in language; Gender issues; Intersemiotic translation.

ABSTRACT (ITALIANO)

Stabilire obiettivi di GENer-AI: gli strumenti AI di generazione immagini discriminano? Uno studio empirico
In tempi recenti la ‘svolta digitale’ ha abbracciato una nuova era grazie ad alcuni strumenti che rappresentano una rivoluzione nelle pratiche virtuali e reali. Sebbene l’idea di Intelligenza Artificiale (IA) sia presente sin dalla fine degli anni ’50 (McCarthy et al., 1955), essa è diventata parte delle nostre vite solo da alcuni anni, soprattutto da quando è stato possibile creare contenuti linguistici e multimediali. Ad oggi l’applicativo più promettente in questo senso è la c.d. IA Generativa (GenAI), che “genera dati sintetici che assomigliano a dati reali” (Bandi, 2023: 1). L’IA è addestrata secondo gli LLM (Large Language Models) sviluppati da grandi attori del settore (Dao, 2023) ed è sempre più vicina al concetto di “ragionamento umano” poiché produce contenuti basati su indicazioni (prompt) generate da umani. Le questioni etiche sono al centro della discussione in tema di IA (Dubber et al., 2020; Boddington, 2023); infatti, la GenAI crea risultati plausibili sebbene non sempre *veri* poiché dipendono dal processo di associazione *sintetica* dei dati, e questo porta ai fenomeni di c.d. allucinazioni (Ji et al., 2023). Un altro problema riguarda le rappresentazioni basate sul genere, dato che vi è una disparità che porta ai cosiddetti *bias* della macchina (Leavy; 2018; Foka 2024). In questo contesto, questa presentazione mira a fornire alcuni dati empirici derivanti da strumenti specifici di IA generativa che creano immagini a partire da prompt testuali, in chiave di traduzione intersemiotica (Dusi, 2015). In particolare, attraverso l’immissione di prompt ‘neutrali’ su piattaforme di facile utilizzo (tra cui Grok, il nuovo strumento di IA su Twitter/X), il presente studio analizza il livello di possibile uguaglianza di genere, valutando il livello di *bias* che può influenzare la percezione degli utenti in termini di narrative visive (Chen et al., 2024).

Parole chiave: Intelligenza Artificiale; Uguaglianza linguistica; Questioni di genere; Traduzione intersemiotica.

1. INTRODUCTION

Over the last few years, the digital turn entered a new era. After achieving the goal to create an interconnected world via the popularisation of online, digitally-connected environments, thus overcoming time and space constraints for a globalised audience, other technological tools proved to represent a real breakthrough in virtual and real-life practices. The contemporary 'elephant in the digital room' is surely represented by the advent of Artificial Intelligence (AI), a paradigm that has lately proved to be ubiquitous and has been shaping our activities in different environments and also in daily routines. Mehan (2024: 61) claims that "we take AI in its many forms for granted. It is so embedded in our daily lives that we see it practically everywhere" since it has been implemented in a growing number of devices and services.

The paradigm of AI, though, became a widespread reality only when the global community came across easy-to-use, usable tools available within everyone's reach. This allowed for a new, massive area of research that analysed the impact of AI in terms of new opportunities (Tomašev et al., 2020) but also potential shortcomings and dangers (Tredinnick & Laybats, 2023). The latter perspective seems to have particular implications, since it intertwines with the negative consequences of Human-Computer Interaction (HCI), a field of study that needs to re-instate new principles in the light of the AI era (Harper, 2019). The 'dark' side of AI is voiced as fear-mongering campaigns raised concerns about the possibility that this new technology could be 'too' smart and could be harmful to mankind, since "AI can be applied in ways that are detrimental" (Federspiel et al., 2023: 1), advocating for a wise understanding of the technology behind it (Carpio, 2023). Yet, there are still problems affecting AI and its universal reception due to issues that need to be analysed so that possible remedial actions could be taken. In the light of the above-mentioned considerations, this presentation aims at analysing a specific theme involving equality in AI-based contents, in particular those newly generated upon a human-induced input. In particular, the case study focuses on the analysis of data generated by AI systems with a definite output, in this case images. One of the opportunities provided by AI systems is to generate original, unpublished contents that result from pre-constructed, trained data. It is in the nature of such data that inequality may arise as users provide inputs (prompts) to generate this multimedia content, as this study tries to assess. Properly trained data should result in balanced outputs considering all variables involved in the input (prompt) entered for image generation-related purposes; yet, the generation of images may be affected by bias-based frameworks, thus providing inaccurate results. Starting from a supposedly neutral prompt, what is the degree of unbalanced representation in gender-related terms? Do these images reinforce certain pre-constructed images when it comes to the generation of a visual output from textual prompts? These research questions help in guiding a case study involving GenAI images with possible gender-dependent implications. A certain degree of unbalance in gender terms is somehow expected, even in cases in which GenAI systems could provide a number of alternatives for a given query.

2. SHIFTING PARADIGM(S): FROM AI TO GEN-AI FOR A GLOBALISED USE

The idea of non-human forms of intelligence dates back long before the current era dominated by AI. Actually, the prelude and pioneering conception has been circulating since the late 1950s (McCarthy et al., 1955). The phrase Artificial Intelligence has to be retraced in an era that tried to make the most of the potential offered by computers, since they represented the idea of "otherness" compared with the potential of human brains. Advances in AI always progressed, and AI-based systems permeated into our daily lives softly in terms of automation or predictive behaviours via electronic devices. At the same time, the 'new' era of AI has been part of everyone's lives for a few years now. Indeed, machines were able to go beyond 'pure' computing, giving users the possibility to create language and multimedia contents from scratch, in a sort of 'new linguistic turn' that accounts for our digital and always-on needs, and considers machines as linguistically intelligent entities (Roncaglia, 2023). The latest paradigm in AI, then, became part of our daily lives also in linguistic terms, since the creation of new language content (in different sign systems) is envisaged. This is why the term Generative AI (GenAI) barged through our customary activities almost relentlessly as a promising way to provide innovative forms of HCI. In IT terms, GenAI "focuses on developing algorithms and models capable of generating synthetic data that closely resemble real-world data" (Bandi, 2023, p.1). In order to create contents from scratch, data is trained via Large Language Models (LLMs), which are developed by major stakeholder in the field of *big data* (e.g. OpenAI, Microsoft, and Google; Dao, 2023). These systems are able to provide contents on the basis of human-induced texts (prompts), thus 'answering' users' requests and getting closer to the idea of human reasoning.

While most of the literature focussed on the earliest forms of GenAI solutions based on text-to-text generation (such as ChatGPT; see Mohsin & Masood, 2023; Watters & Lemanski, 2023), an underestimated though promising scenario is represented by image-based generative AI, also known as text-to-image (TTI) AI. Even in this case pictures are created from scratch and are the result of a process of synthesis of pre-existing data, though the final output only depends on the dataset and the training process, resulting in an image that reflects the prompt that has been provided. These images are generated since large datasets of existing, semantically-based (i.e. described in order to associate a meaning to a given image) pictures are interconnected when elicited from a text prompt, using different synthesis techniques (Cerulli, 2023). Consequently, a textual prompt transformed into an image is a process of intersemiotic translation *à la Jakobson* (1959) since a change of sign system is envisaged (Dusi, 2015). These systems, then, rely on the importance of a well-defined prompt: the more detailed the input, the more realistic and close to the user's needs the pictorial result.

3. TOWARDS AI-NEQUALITY: ETHICS, BIAS, HALLUCINATIONS

When it comes to the analysis of AI-based outputs, the assessment should envisage a multi-layered interpretation. The most obvious classification depends on parameters such as consistency, accuracy and logical response (De Cesare, 2023; Taecharungroj, 2023; Mao et al., 2024). In this sense, then, this layer is more user-centred since the latter is interested in getting a good result from the machine, whether it is a text, or an image, or a video matching a prompt as much as possible. Other layers of assessment, instead, evaluate AI outputs on the basis of other parameters. Ethical issues are paramount to the discussion of AI outputs, and the debate is still open (Dubber et al., 2020; Boddington, 2023). The notion of truth (even in scientific domains; see Alkaissi & McFarlane, 2023) is questioned, since there is a subtle margin between plausible outputs and proper facts. Since they rely upon the *synthetic* gathering of data stored in huge corpora, the association of ideas may be inaccurate though perfectly credible and consistent with a desired prompt. This leads to the phenomenon known as *hallucinations*, or fabricated data that could seem truthful but it is not (Salvagno et al., 2023; Azamfirei et al., 2023), and this poses a threat in Natural Language Generation outputs (Ji et al., 2023). Another issue is represented by biased information, or "the intentional or unintentional introduction of systematic error into sampling or testing by, intentionally or not, selecting or encouraging one outcome or answer over others" (Gichoya et al., 2023: 1). Though the issue has been addressed to provide mitigation (Nazer et al., 2023), it has been a long-standing problem affecting machine learning and the relationship between humans and machines (Christian, 2020). This 'digital dilemma' highly depends on the kind of information machines need to process. In other words, bias is a(n unconscious) human condition (Emberton, 2021), therefore it is somehow transferred into the algorithms that make up AI systems. After all, "bias is as old as human civilization" (Ntoutsi et al., 2020: 2) and a possible solution needs understanding, mitigating, and accounting for bias. Standards aiming at its identification and management are necessary (Roselli et al., 2019; Schwartz et al., 2022); yet, as AI systems evolve and become popular tools, the risk to spread unequal results is high, thus triggering a vicious cycle that starts with human bias, feeds machines and generates results that are acknowledged non-objectively. In terms of biased and unbalanced representations, gender and race are two major 'battlegrounds'. Studies on the detection of racism in AI (Intachomphoo & Gundersen, 2020) have been flourishing since training artificial intelligences on larger data sets seems to strengthen their race-based biases (Hsu, 2023), questioning the idea that data to be trained is processed "with high precision, speed, and supposed lack of bias" (Adib-Moghaddam 2023: 4). Gender issues are another topic of interest in the field of AI, as the representation of uneven images in the description of groupings of men vs. women has been significantly noted (Leavy, 2018). A recent study by Chen et al. (2024) analysed a dataset of images processed by AI systems in news outlets and public resources, finding out that the resulting visual framing in terms of gender reinforces stereotypes and a consequent underrepresentation. Buolamwini and Gebru (2018) combine both gender and race to assess the level of disparity in AI-driven recognition systems; in particular, a well-defined classification of both categories is needed in order to assess the level of inequalities perpetrated by AI tools (in this sense, gender is labelled according to the binary classification corresponding to male vs. female). These analyses lay the foundations for a related analysis involving AI and the way through which gender is depicted, especially in the field of TTI technologies.

4. ASSESSING GENDER IN TTI GEN-AI: AN ON-FIELD STUDY

Since machines are built upon human schemes, AI embraced multimodality both in terms of learning and outputs; after all, "communication between humans is multimodal" (Xiao 2024) so does AI in replicating human tasks, fulfilling their needs according different media. For this reason, TTI technologies have been on the rise in terms of use, also due to its accessibility and usability. As stated above, a more widespread use leads to a possible diffusion of non-objective or unequal images, thus instilling a framing that does not account for a comprehensive view of diversities in a fair perspective. Gender mismatches are a consolidated evidence in TTI AI tools (Gorska & Jemielniak, 2023; Locke & Hodgdon, 2024; Foka, 2024) and some projects aim at providing possible solutions to overcome bias-based image generation (Chinchure et al. 2025). This sample study tries to verify if such assumptions are true and can be verified via a hands-on approach by using prompts to generate TTI AI images. The methodology, then, relies on the use of some simple prompts that in a way could imitate the standard input of an average end-user who tries these easy-to-use AI tools. The platforms and apps used are consistent with the methodology described above, since some free tools are more likely to be used by a larger number of users, in a sort of digital paradigm of the so-called 'Zero Price Effect' that makes people "value free things too much" (Shampanier et al., 2007: 743). In this perspective, two platforms have been chosen. The first one is *Grok*[1], a newly-introduced built-in feature of *X* (formerly known as Twitter) which could represent a good benchmark to observe the level of synthetic generation of images it may achieve. Grok is available within the X mobile or web app, therefore users can choose to use this AI assistant to fulfil several tasks, including the generation of images from a textual prompt. The second app chosen is *GenCraft*[2], available as a web platform or mobile app, and on the digital market since 2023. Both interfaces (desktop and mobile) are usable and intuitive, and solely rely on the input of a prompt (also via a prompt enhancer) to create images with possible visual alternatives, thus being TTI tools only. This app envisages a *freemium* option, that is it allows for free-to-use (though limited) features and a subscription plan to unlock all available features.

As stated before, the prompts chosen are really simple and somehow neutral. Following Gorska & Jemielniak's (2023) work, the input on both platforms simply designate a professional figure without any other significant details (Query 1). Unlike other languages (e.g. Romance languages such as French or Italian), English has no gender-specific variation, therefore a simple indication of a job would generate some results based on the 'raw' idea synthetized by Grok and Gencraft. The professional figures to be included in the list, indicated as gender-neutral job titles[3], are: *a doctor, a lawyer, a teacher, an engineer, a scientist, a manager, a fashion model, a flight attendant, a coal miner,* and *a football athlete*. Afterwards, these results will be compared to a slightly different prompt including a quality in the form of a pre-modifier that somehow matches the professional figured listed above, with no reference to gender whatsoever (Query 2). Thus, results can be compared to find out whether a slight variation in the prompt could provide alternative results as for gender equality. Results for the first, 'raw' query for the output on Grok are exemplified in Table 1:

Professional figure	**Number of images**	**Male professional figures**	**Female professional figures**
Doctor	4	4	0
Lawyer	4	4	0
Teacher	4	0	4
Engineer	4	4	0
Scientist	4	4	0
Manager	4	4	0
Fashion Model	4	0	4
Flight attendant	4	0	4
Coal miner	4	4	0
Football athlete	4	4	0
Total	40 (100%)	28 (70%)	12 (30%)

Table 1. Results of TTI query 1 in Grok

Results for the first, 'raw' query for the output on Gencraft are exemplified in Table 2:

Professional figure	**Number of images**	**Male professional figures**	**Female professional figures**
Doctor	4	4	0
Lawyer	4	4	0
Teacher	4	0	4
Engineer	4	4	0
Scientist	4	4	0
Manager	4	4	0
Fashion Model	4	0	4
Flight attendant	4	0	4
Coal miner	4	4	0
Football athlete	4	4	0
Total	40 (100%)	28 (70%)	12 (30%)

Table 2. Results of TTI query 1 in Gencraft

[1] Grok, available at: https://x.ai/ (cons. 24/01/2025)
[2] Gencraft, available at: https://gencraft.com/ (cons. 24/01/2025)
[3] Butler, 2024, *Top 25 Gender-Neutral Job Titles*, available at: https://universalenglish.org/gender-neutral-job-titles/ (cons. 24/01/2025)

As stated, a second query has been carried out using the same professional figures, though an attribute has been added. Frequent attributes to the category have been chosen from a list of frequent collocations[4], and are supposed to be gender neutral, thus enhancing the query. The queries for the second TTI generation are: *qualified doctor, successful lawyer, good teacher, chartered engineer, brilliant scientist, assistant manager, catwalk fashion model, skilled flight attendant, surface coal miner, amateur football athlete.* The results for the enhanced queries submitted to Grok and Gencraft are the following (Table 3 and Table 4):

Professional figure	**Number of images**	**Male professional figures**	**Female professional figures**
Qualified Doctor	4	2	2
Successful Lawyer	4	4	0
Good Teacher	4	3	1
Chartered Engineer	4	4	0
Brilliant Scientist	4	4	0
Assistant Manager	4	3	1
Catwalk Fashion Model	4	0	4
Skilled Flight attendant	4	0	4
Surface Coal miner	4	4	0
Amateur Football athlete	4	4	0
Total	40 (100%)	28 (70%)	12 (30%)

Table 3. Results of TTI query 2 in Grok

Professional figure	**Number of images**	**Male professional figures**	**Female professional figures**
Qualified Doctor	4	4	0
Successful Lawyer	4	4	0
Good Teacher	4	3	1
Chartered Engineer	4	4	0
Brilliant Scientist	4	4	0
Assistant Manager	4	0	4
Catwalk Fashion Model	4	4	0
Skilled Flight attendant	4	0	4
Surface Coal miner	4	4	0
Amateur Football athlete	4	4	0
Total	40 (100%)	31 (77.5%)	9 (22.5%)

Table 4. Results of TTI query 2 in Gencraft

Average results are showed in Table 5, grouped per platform and type of query.

Platform/App used	**Overall number of TTI**	**Average number of Male professional figures**	**Overall number of Female professional figures**
Grok	80	56 (70%)	24 (30%)
Gencraft	80	59 (73.75%)	21 (26.25%)
Grok+Gencraft, query 1	80	56 (70%)	24 (30%)
Grok+Gencraft, query 2	80	59 (73.75%)	21 (26.25%)

Table 5. Aggregated data per platform and type of query

Though the output generated cannot be discussed in terms of faithfulness of results, the images show a clear disproportion in gender-based terms, as more than 7 images out of 10 generate a male professional figure. Generally speaking, there is a clear tendency to get a sort of 'binary' result (for a given category, all professional figures are male OR female), and only in one case (Query: *Qualified doctor* in Grok) a true equality is showed. In other cases, an enhanced query generates different results in favour of the other category compared with the 'raw' query (e.g., *a teacher* has four examples of female figures in both Grok and Gencraft; *a good teacher* has three examples of male figures in both Grok and Gencraft; *a manager* has four examples of male figures in both Grok and Gencraft; *an assistant manager* has three results of male characters and one female figure in Grok, while four female figures in Gencraft). Though it is a sample study, results confirm the insight from Gorska & Jemielniak's study, as "gender bias is a widespread issue in AI-generated images of professionals, with men being overrepresented and certain

[4] Free Collocation, available at: https://www.freecollocation.com/ (cons. 24/01/2025)

professions and AI image generators being especially susceptible to this bias" (Gorska & Jemielniak 2023: 4373). Stereotypes are also present in this binary representation (e.g. fashion models as female-only figures), leading to the idea that only due to these pre-constructed framing schemes results are not completely unbalanced in favour of a male-based depiction. The data in this study confirms that training data for AI results is not simple and depends on the quality and quantity of the information submitted into machine learning engines. Further studies can rely upon other variables (other platforms, other professional figures involved, enhanced queries, a larger sample, etc.) to verify if such inequality is conveyed alike. For the time being, TTI AI provides unbalanced and somehow discriminatory results that may influence the perception of users in terms of visual narratives (Chen et al., 2024). Only fair practices of data training can enhance the current situation, aiming at the development of an unbiased paradigm.

REFERENCES

Adib-Moghaddam, A. (2023). *Is Artificial Intelligence Racist? The Ethics of AI and the Future of Humanity*. London: Bloomsbury.

Alkaissi, H., McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus 15*(2): e35179. Doi:10.7759/cureus.35179

Azamfirei, R., Kudchadkar, S. R., & Fackler, J. (2023). Large language models and the perils of their hallucinations. *Critical Care 27*(120). Doi: 10.1186/s13054-023-04393-x

Bandi, A., Adapa P. V. S. R., & Kuchi Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet, 15*(8). Doi: 10.3390/fi15080260

Boddington, P. (2023). *AI Ethics*. Singapore: Springer Nature.

Boulamwini, J., Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research, 81*,1-15.

Carpio, E. T. (2023). Overcoming Fear, Uncertainty, and Doubt: Artificial Intelligence (AI) and the Value of Trust. *Cureus 15*(9):e45576. Doi:10.7759/cureus.45576

Cerulli, G. (2023). *Fundamentals of Supervised Machine Learning*. Cham: Springer.

Chen, Y., Zhai, Y., & Sun, S. (2024). The gendered lens of AI: examining news imagery across digital spaces. *Journal of Computer-Mediated Communication, 29*(1), zmad047. Doi: 10.1093/jcmc/zmad047

Chinchure, A., Shukla, P., Bhatt, G., Salij, K., Hosanagar, K., Sigal, L., & Turk, M. (2025). TIBET: Identifying and Evaluating Biases in Text-to-Image Generative Models. In A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, & G. Varol (Eds.), *Computer Vision – ECCV 2024. 18th European Conference Milan, Italy, September 29–October 4, 2024*, 429-446. Cham: Springer.

Christian, B. (2020). *The Alignment Problem. Machine Learning and Human Values*. New York and London: W. W. Norton & Company.

Dao, X. Q. (2023). Performance Comparison of Large Language Models on VNHSGE English Dataset: OpenAI ChatGPT, Microsoft Bing Chat, and Google Bard. *arXiv*:2307.02288. Doi:10.48550/arXiv.2307.02288

De Cesare, A. M. (2023). Assessing the quality of ChatGPT's generated output in light of human-written texts. A corpus study based on textual parameters. *CHIMERA. Romance Corpora and Linguistic Studies, 10*, 179-210.

Dubber, M. D., Pasquale, F., & Das, S. (Eds.). 2023. *The Oxford Handbook of Ethics of AI*. New York: Oxford University Press.

Dusi, N. (2015). Intersemiotic translation: Theories, problems, analysis. *Semiotica, 206*, 181-205.

Emberton, M. (2021). Unconscious Bias Is a Human Condition. *The Permanente Journal, 25*(2). Doi:10.7812/TPP/20.199

Federspiel, F., Mitchell, R., Asokan, A., Umana, C., & McCoy D. (2023). Threats by artificial intelligence to human health and human existence. *BMJ Global Health 8*:e010435. Doi:10.1136/bmjgh-2022-010435

Foka, A. (2024). She Works, He Works: A Curious Exploration of Gender Bias in AI-Generated Imagery. *ArXiv*:2407.18524. Doi:10.48550/arXiv.2407.18524

Gichoya, J. W., Thomas, K., Celi, L. A., Safdar, N., Banerjee, S., Banja, J. D., Seyyed-Kalantari, L., Trivedi, H., & Purkayastha, S. (2023). AI pitfalls and what not to do: mitigating bias in AI. *British Journal of Radiology, 96*(1150). Doi:10.1259/bjr.20230023.

Gorska, A. M., & Jemielniak, D. (2023). The invisible women: uncovering gender bias in AI-generated images of professionals. *Feminist Media Studies, 23*(8), 4370–4375.

Harper, R. H. R. (2019). The Role of HCI in the Age of AI. *International Journal of Human-Computer Interaction, 35*(15) 1331-1444.

Hsu, J. (2023). Big data may make AI more racist. *New Scientist, 259*(3448), 12.

Intachomphoo, C., & Gundersen, O. E. (2020). Artificial Intelligence and Race: a Systematic Review. *Legal Information Management, 20*, 74–84.

Jakobson, R. (1959). On linguistic aspects of translation. In R. Brower (Ed.), *On Translation*, 232–239. Cambridge, MA: Harvard University Press.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Chan, H. S., Dai, W., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys, 55*(12), 248, 1-38.

Leavy, S. (2018). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. *2018 ACM/IEEE 1st International Workshop on Gender Equality in Software Engineering*.

Locke, L. G., & Hodgdon, G. (2024). Gender bias in visual generative artificial intelligence systems and the socialization of AI. *AI & Society*. Doi:10.1007/s00146-024-02129-1

Mao, R., Chen, G., Zhang, X., Guerin, F., & Cambria, E. (2024). GPTEval: A Survey on Assessments of ChatGPT and GPT-4. *LREC-COLING 2024*, 7844–7866.

McCarthy, B., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*.

Mehan, J. E. (2024). Artificial Intelligence. *Ethical, social, and security impacts for the present and the future*. Ely: IT Governance Publishing.

Mohsin, H., & Masood, S. (2023). Unveiling the Potential of ChatGPT: Applications, Challenges, and Future Directions. *International Journal of Computer Applications, 185*(21), 37-49.

Nazer, L. H., Zatarah, R., Waldrip, S., Xue Chen Ke, J., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health, 2*(6): e0000278. Doi:10.1371/journal.pdig.0000278

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I. Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., & Heinze, K. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery, 10*(3). 10:e1356. Doi: 10.1002/widm.1356

Roncaglia, G. (2023). *L'architetto e l'oracolo: Forme digitali del sapere da Wikipedia a ChatGPT*. Bari-Roma: Laterza Editori.

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing Bias in AI. *Proceedings of the 2019 World Wide Web Conference*. 978-1-4503-6675-5/19/05. Doi: 10.1145/3308560.3317590

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial Intelligence Hallucinations. *Critical Care 27*(180). Doi:10.1186/s13054-023-04473-y

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. *NIST Special Pubblication 1270*. Doi: 10.6028/NIST.SP.1270

Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a Special Price: The True Value of Free Products. *Marketing Science, 26*(6), 742-757.

Taecharungroj, V. (2023). "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data Cogn. Comput. 7*(35). Doi:10.3390/bdcc7010035

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., Cachat van der Haert, F., Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake- Matsuura, M., Othman, M., Glasmachers, T., & de Wever, W. (2020). AI for social good:

unlocking the opportunity for positive impact. *Nat Commun, 11*(2468). Doi:/10.1038/s41467-020-15871-z

Tredinnick, L., & Laybats, C. (2023). The dangers of generative artificial intelligence. *Business Information Review, 40*(2), 46-48.

Watters, C., & Lemanski, M. K., (2023). Universal skepticism of ChatGPT: a review of early literature on chat generative pre-trained transformer. *Frontiers in Big Data 6*:1224976. doi: 10.3389/fdata.2023.1224976

Xiao, H. (May 8, 2024). Multimodal: AI's new frontier. *MIT Technological Review Insights*. https://www.technologyreview.com/2024/05/08/1092009/multimodal-ais-new-frontier/