

# AI Chatbots: Evaluating the Balance Between Threats and Opportunities


Abigail Ross Muniz<sup>1</sup>

<sup>1</sup> Universidad de Anglia Ruskin – Inglaterra, [abigailross1983@hotmail.com](mailto:abigailross1983@hotmail.com)  
<https://orcid.org/0009-0005-8839-5776>

## ABSTRACT

The multifaceted landscape of artificial intelligence (AI), with a specific focus on the emergence and implications of AI chatbots. Skepticism and concerns about the potential dangers of AI are explored, drawing parallels to historical narratives such as the "Golem." The texts highlight the historical context of AI development, including past "false dawns" and ethical challenges. Education, publishing, and the perceived threats of AI chatbots, particularly in facilitating plagiarism, are discussed. The need for critical analysis, ethical considerations, and responsible AI development is underscored. The texts delve into the evolving relevance of the Turing test, proposing the concept of an AI chatbot lie test amid challenges posed by misinformation. The texts express skepticism about the current state of AI, citing instances of misinformation and the difficulty in distinguishing between genuine and fabricated content. The discussions also emphasize the urgent need for a discerning approach to online information in the era of AI-generated content. Ultimately, the conclusion emphasizes that, while AI poses challenges, the primary existential risks to humanity are rooted in human actions rather than the mechanisms of artificial intelligence.

**Keywords:** Artificial Intelligence; AI Chatbots; Existential Risks



**Copyright:** © 2023  
by the authors. This  
article is an open  
access article  
distributed under the  
terms and conditions  
of the Creative  
Commons

**Received:**  
07 February, 2023

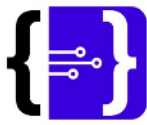
**Accepted for  
publication:**  
24 March, 2023

## AI Chatbots: Evaluando el Equilibrio entre Amenazas y Oportunidades

### RESUMEN

El artículo explora el panorama multifacético de la inteligencia artificial (IA), con un enfoque específico en el surgimiento e implicaciones de los chatbots de IA. Se abordan el escepticismo y las preocupaciones sobre los posibles peligros de la IA, estableciendo paralelos con narrativas históricas como la del "Golem". Los textos destacan el contexto histórico del desarrollo de la IA, incluyendo "falsos amaneceres" pasados y desafíos éticos. Se analizan la educación, la publicación y las amenazas percibidas de los chatbots de IA, especialmente en la facilitación del plagio. Se subraya la necesidad de análisis crítico, consideraciones éticas y desarrollo responsable de la IA. Los textos exploran la relevancia en evolución de la prueba de Turing, proponiendo el concepto de una prueba de mentira para los chatbots de IA en medio de los desafíos planteados por la desinformación. Los textos expresan escepticismo sobre el estado actual de la IA, citando instancias de información errónea y la dificultad para distinguir entre contenido genuino y fabricado. Las discusiones también enfatizan la necesidad urgente de un enfoque discernidor hacia la información en línea en la era del contenido generado por IA. En última instancia, la conclusión destaca que, si bien la IA plantea desafíos, los riesgos existenciales primarios para la humanidad tienen sus raíces en las acciones humanas en lugar de los mecanismos de la inteligencia artificial.

**Palabras clave:** *Inteligencia Artificial; Chatbots de IA; Riesgos Existenciales*



## INTRODUCTION

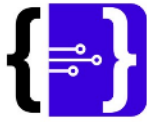
In November 2022, OpenAI introduced ChatGPT, an AI chatbot that amassed over 100 million users by February 2023. AI chatbots, grounded in extensive language models and machine learning, hold the potential to transform our interactions with computers and digital systems. Advocates of these advancements assert that these applications can and will yield substantial benefits for everyone. However, a significant number, including those leading in technology, harbor greater skepticism. Some now contend that AI, in its current state, is perilous and toxic, potentially posing a threat to humanity. While the latter concern may appear exaggerated and misplaced, the consequences of these recent developments are profound. They necessitate comprehensive analysis, attention, and decisive action to prevent an exponential surge in disinformation, fostering severe and irreparable mistrust in these technologies.

### The Golem

A golem, historically, denotes an animated creature crafted from mud capable of performing tasks, typically under its creator's control but lacking the ability to speak. The most renowned instance hails from the 16th century: The Golem of Prague, purportedly crafted by Rabbi Judah Loew ben Bezalel. Rabbi Loew, also known as the Maharal, is said to have shaped a golem from clay extracted from the Vltava River. This creature was devised to shield the Prague Jewish community from anti-Semitic assaults and pogroms. Activation involved the Rabbi inserting a parchment with one of the names of God into its mouth, with deactivation achievable by removing the parchment.

In his 1981 story, "Golem XIV" [1], Stanislaw Lem transposed the golem into the computer age. This story is part of "Imaginary Magnitude," a collection of Lem's narratives, and takes the form of a lecture on computer development in the early 21st century. The narrative contemplates the transition of computers from 'insect' to human:

*"The disparity between earlier machines and the recent ones essentially boiled down to the distinction between an insect and a man (sic). An insect enters the world pre-programmed to its end by instincts, obeyed unreflectively. Conversely, a man must learn appropriate behavior, fostering independence. With determination and knowledge, a man can alter prior action*



*programs. Thus, computers up to the twentieth generation were characterized by 'insect' behavior: incapable of questioning or, more significantly, modifying their programs. The programmer 'impregnated' the machine with knowledge, akin to evolution 'impregnating' an insect with instinct. Despite much talk about 'self-programming' in the twentieth century, these remained unfulfilled daydreams" [1] (p. 39).*

Golem XIV unfolds in the year 2047, comprising an introduction followed by two lectures delivered by Golem XIV and concluding with an afterword. During the 1980s when Lem penned this work, the notion of self-programming machines existed as somewhat of a distant aspiration. However, contemporary insights reveal that foundational technical groundwork was already underway at that time. Geoffrey Hinton, recognized as one of the 'fathers of AI,' was actively engaged in the development of neural networking, though its actual realization came to fruition only in the present century.

### **False Dawns of AI**

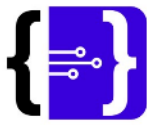
Since that period, various false dawns of AI have materialized, each heralding the imminent arrival of transformative and advantageous technologies. In the 1970s and 1980s, the rule-based model of AI predominated, significantly influenced and funded by the U.S. military. This influence constrained and impeded the neural network approach for a considerable duration. The shift in military interest following the Cold War's conclusion in the 1990s marked a turning point, revealing the limitations of the rule-based approach.

In the preparation of this Editorial, the use of an AI chatbot, ChatGPT, was deemed essential. To inquire about the number and nature of false dawns in AI over the past 50 years, the following question was posed:

"How many false dawns of AI have there been in the past 50 years, and what were they?"

The response received outlined several significant instances of overhyped expectations and subsequent disappointment:

1. The AI Winter of the 1970s: Initial optimism about AI's potential waned due to unrealistic expectations and slow progress in practical applications, leading to reduced funding and a period known as the "AI Winter."



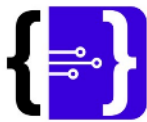
2. Expert Systems Hype in the 1980s: Popularity of expert systems, aiming to replicate human expertise, declined as the technology struggled to effectively capture and represent complex human knowledge.
3. Neural Networks Renaissance in the 1990s: A resurgence of interest in neural networks occurred with more efficient learning algorithms and larger datasets, but practical applications remained limited.
4. AI in Popular Culture: Films and media representations created inflated expectations about AI capabilities, surpassing the reality of AI development at the time.
5. Deep Learning Breakthroughs in the 2010s: Significant breakthroughs in deep learning led to inflated expectations about the immediate prospects of general AI and raised concerns about job displacement.

It's crucial to emphasize that the occurrences of these "false dawns" in AI do not suggest a lack of progress or value during those periods. Each phase significantly contributed to the evolution of AI technology, even when the expectations surpassed the reality at the time. Additionally, recent years have seen substantial advancements in AI technologies, including reinforcement learning, robotics, and natural language processing. These developments indicate an increasing integration of AI into various aspects of our lives and industries. (Response obtained from ChatGPT—May 11, 2023)

### **AI Chatbots—A New Dawn: Blessing or Curse?**

Over the past 10–15 years, remarkable advancements have taken place in robotics, machine learning, and natural language processing. This progress has led to the emergence of AI chatbots, notably exemplified by ChatGPT, unveiled in November 2022. By early May 2023, the proliferation of these chatbots prompted authors to assess them in general and for specific activities [3,4]. The advent of these applications raised concerns across various fields and issues, a sentiment that heightened with the announcement of additional AI chatbots.

Two primary areas triggering concerns are education and publishing. Educators have long grappled with students copying content from electronic sources, and editors/reviewers have dealt with plagiarism and self-plagiarism in journal submissions. AI chatbots have elevated these issues to a new level. Students



can now sidestep plagiarism detection tools like Scribbr by requesting an AI chatbot to generate entire works, as demonstrated in an overview of 12 plagiarism checkers from March 2022, predating ChatGPT's launch [5].

Educational institutions at all levels must now deliberate on how to best engage with these chatbots, while plagiarism detectors now include indicators related to potential AI use in various documents. Scribbr even provides a weblink for 'Using ChatGPT for Assignments: Tips & Examples' [6].

In initiating this transdisciplinary discussion on 'AI Chatbots: Threat or Opportunity?' the goal was to encourage insightful contributions from diverse areas, disciplines, and practices. In preparation, editors of participating journals were consulted to suggest questions and issues prompting insightful and stimulating submissions. These are available on the topic's website for convenience.

We welcomed submissions covering a broad range of topics. To offer some insight into the areas of primary interest, we present the following questions and concerns:

- The advent of AI chatbots is asserted to mark a new era, promising significant advancements in the integration of technology into people's lives and interactions. Is this assertion likely to hold true, and if so, where will the impacts be most widespread and impactful?
- Is it feasible to find a balance in the impact of these technologies, minimizing potential harms while maximizing and distributing potential benefits?
- How should educators address the challenge posed by AI chatbots? Should they embrace this technology and reshape teaching and learning strategies accordingly, or should they work to protect traditional practices perceived as a major threat?
- Growing evidence indicates that many AI applications, including algorithms, embody bias and prejudice in their design and implementation. How can this bias be countered and rectified?
- How can publishers and editors differentiate between manuscripts composed by a chatbot and authentic articles written by researchers? Is training to recognize these distinctions necessary, and if so, who should provide such training?
- How can the academic community and the broader public be safeguarded against the generation of 'alternative facts' by AI? Should researchers be obligated to submit their data with

manuscripts to authenticate its legitimacy? What role do ethics committees play in upholding research integrity?

- Can the technology underpinning AI chatbots be improved to prevent misuse and vulnerabilities?
- Exploration of innovative models and algorithms for utilizing AI chatbots in cognitive computing.
- Techniques for training and optimizing AI chatbots for cognitive computing tasks.
- Evaluation methods to assess the performance of AI chatbot-based cognitive computing systems.
- Case studies and experiences related to the development and deployment of AI chatbot-based cognitive computing systems in real-world scenarios.
- Examination of social and ethical issues associated with the use of AI chatbots for cognitive computing.

These encompass a broad spectrum of issues, though not exhaustive. Some focus specifically on the technology within a particular context, while others have a more comprehensive scope. Collectively, they offer insights into how AI chatbots could be embraced as foundations for genuine advancements benefiting everyone, yet simultaneously perceived as potential or real threats. The reactions to the surge in the use of AI chatbots since November 2022 indicate that, especially among those deeply involved in the technology's development, the drawbacks far outweigh the advantages. A notable figure, Geoffrey Hinton, has issued warnings about the "quite scary" dangers of AI chatbots, expressing concerns that they could surpass human intelligence and be exploited by malicious entities. According to Hinton, this could lead to the creation of highly effective spambots and enable authoritarian leaders to manipulate their electorates.

Hinton, who had led Google's research on neural networks for a decade, recently took a critical stance on the very technology he was previously employed to develop. In an interview mentioned in The Guardian report, Hinton stated that his perception changed when Microsoft incorporated a chatbot into its Bing search engine, prompting Google to worry about the risk to its search business. This rush to

market resulted in the release of a powerful technology without sufficient consideration of its potential consequences, highlighting a notable failure of imagination among Hinton and others at the forefront of AI. This failure, unfortunately, is a common occurrence in the realm of innovative technologies, where promising advancements are often marketed as great benefits to humanity but fall short of delivering on that promise.

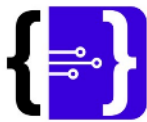
A recent joint letter signed by numerous luminaries in the fields of AI and ICT calls for a halt to "Giant AI Experiments," urging all AI labs to pause the training of AI systems more powerful than GPT-4 for at least six months. This plea is likened to The Sorcerer's Apprentice, with the added irony that the sorcerers themselves have unleashed forces that are unstoppable and irresistible. The reference to The Sorcerer's Apprentice draws parallels to the classic tale by Goethe, later set to music by Dukas and featured in Disney's Fantasia, with Mickey Mouse in the role of the apprentice. The Guardian's report on Hinton's transition from AI developer to AI critic includes a more damning observation by Valérie Pisano, the CEO of Mila—the Quebec Artificial Intelligence Institute.

*... the casual and hasty attitude toward safety in AI systems would be deemed unacceptable in any other sector. "The technology is released, and its developers observe its interactions with humans, making adjustments based on the outcomes. We would never collectively endorse such a mindset in any other industry. There's a peculiar aspect to technology and social media, where we seem to say: 'Sure, we'll address it later,'" she commented. [8]*

### **A Modernized Turing Test**

In the foreseeable future, the widespread presence of AI chatbots will inundate us with a variety of deceptive content, including AI-generated photos, videos, and text distributed across the Internet. To navigate the challenge of distinguishing truth from falsehood, each individual will need to swiftly become adept in a contemporary version of the Turing test. Originally conceived as the imitation game by Alan Turing himself, this test aimed to address inquiries about whether machines could exhibit thought. Turing's 1950 paper [11] argued that the original framing of the question, involving terms like 'machine' and 'think,' was excessively intricate due to inherent ambiguity. He proposed a modified version of the problem, presented as follows:





*"The new form of the problem can be described in terms of a game which we call the 'imitation game.' It is played with three participants: a man (A), a woman (B), and an interrogator (C), who may be of either sex. The interrogator remains in a separate room from the other two. The objective for the interrogator is to determine which of the other two participants is the man and which is the woman. He identifies them by labels X and Y, and at the conclusion of the game, he declares either 'X is A and Y is B' or 'X is B and Y is A.'"* [11] (p. 433)

It's noteworthy that the original test had a basis in discerning the gender of participants, a detail often overlooked. Turing also mentioned that the interrogator [C] 'may be of either sex,' but consistently referred to [C] as 'he'—an aspect reflecting the language norms of that time.

Turing suggested extending the experiment to explore the consequences when a machine assumes the role of participant A in the game:

*"We now pose the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator make incorrect determinations as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original query, 'Can machines think?'"* [11] (p. 434)

Following Turing's 1950 paper, the test remained largely speculative and minimally challenging. However, in the 1990s, particularly after what the previous ChatGPT text referred to as 'The Neural Network Renaissance,' it experienced a revival, notably with the establishment of The Loebner Prize. Hugh Loebner offered a substantial reward for the first programmer capable of creating a program whose communicative behavior could deceive humans into believing they were interacting with another person. This competition took place annually until 2016, coinciding with Loebner's passing, leading to the discontinuation of the events. Critics often viewed the competition as nothing more than a publicity stunt.

With the emergence of AI chatbots, the Turing test may be seen as outdated or taking on a new and crucial role. As the outputs of AI chatbots become integrated into the vast Internet database, individuals using the Internet must recognize that search results or information requests will inevitably include data

generated by chatbots. This wouldn't pose a problem if such outputs were clearly identified as originating from these applications, but currently, this is not the case.

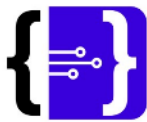
The outputs from AI chatbots compound the challenges we all encounter as we face constant bombardment with spam, fraudulent emails from banks, malicious phone calls, and various forms of fraud and deception. These threats affect everyone, even those highly educated and technically proficient. If advancements in machine learning (ML) and AI are intended to provide benefits, why have the AI experts not developed applications that can effectively filter out the genuine from the malicious?

What is truly needed is a new iteration of the Turing test, demanding that AI applications possess the capability to distinguish significantly between the good, the bad, and the deceitful. Furthermore, current AI chatbots often make errors and fabricate information in their responses. Developers term these occurrences as 'hallucinations,' a term laden with anthropomorphism, suggesting human-like qualities and agency.

AI chatbots are built on a foundation called a large language model (LLM), a computer program trained on millions of text sources to read and generate natural language. The concept is that engaging with an AI chatbot resembles a conversation with a person. However, for reasons not entirely understood even by chatbot developers, these systems frequently invent part or all of their responses. These are not mere 'hallucinations'; they represent mistakes, errors, falsehoods, or occasionally confabulations—answers fabricated when data are insufficient for an accurate response. Well-documented instances include professors being erroneously labeled as sexual predators and politicians wrongly identified as having been convicted of bribery and sentenced to prison. ChatGPT has generated various outputs, including references to nonexistent books and articles and fictitious authors.

The Turing test is now antiquated and serves little practical purpose. However, there is an urgent need for an AI chatbot lie test, as aptly encapsulated in a insightful comment on Edwards' article in *ars technica*.

No matter how many supposedly factual answers prove to be untrue, how many bugs exist in generated code samples, or how derivative the prose or poetry may be, some individuals will perceive genuine



intelligence in the outputs of Large Language Models (LLMs). Moreover, they will consciously or unconsciously undervalue the skills of humans performing the 'same' job as a result. One of these individuals could be your current or future boss, influencing decisions about your salary.

Edwards' article features a ChatGPT output from someone requesting a list of the top books on social cognitive theory. Ten books were listed, four of which are nonexistent, and three were attributed to different authors than indicated. I made a similar request for the top books on grounded theory with similar outcomes. Among the ten listed, one title was entirely fictitious, one was incorrectly attributed, and a third was for a nonexistent book by two authors who have written on the topic. (It's essential to note that I am one of the authors referred to correctly and erroneously in this list.)

It may be argued that we have always needed some form of a 'lie test.' In face-to-face interactions, and even more so with the advent of the web, people consistently lie, make genuine mistakes, or confabulate. Some knowingly spread disinformation or unwittingly disseminate misinformation, a trend that has intensified since the 1990s with the development of information and communication technology (ICT), particularly since 2007 through the powerful combination of smartphones and social media. (Disinformation refers to false information spread deliberately, usually for nefarious purposes, while misinformation is false information spread unintentionally.)

So, why the newfound urgency in the era of AI chatbots? The answer lies in the growing reliance on Internet-based digital resources, as traditional forms of reference have largely disappeared or are now only utilized by specialist researchers. For instance, during an event at The Centre for Computing History celebrating the first business computer LEO, questions arose about future projects. Would researchers need to discern between real and fake documents? Could digital recordings—both video and audio—be deemed trustworthy? To what extent would online resources be partially or wholly fake and generated by chatbots? Additionally, how could researchers ensure the authenticity and provenance of their sources?

One response to these challenges has been explainable AI, where the algorithms and operation of the technology are transparent, allowing users to understand and question the basis for decision-making

and 'reasoning' that leads to system outputs. However, this approach overlooks the primary motivations and funding sources for AI: military, governmental, and commercial interests.

Collectively, this implies the imperative need to refine our skills and heighten our levels of suspicion and distrust. Is it possible for the AI community to create an authentically intelligent and perceptive application that aids us in our critical pursuits? The likelihood of such an occurrence is minimal, as highlighted by the following excerpt from The Guardian report, quoting Jeff Dean, Google's chief scientist, offering little reassurance.

Jeff Dean, Google's chief scientist, expressed appreciation for Hinton's contributions to the company over the past decade in a statement. "I've deeply enjoyed our many conversations over the years. I'll miss him, and I wish him well!"

"As one of the first companies to publish AI Principles, we remain committed to a responsible approach to AI. We're continually learning to understand emerging risks while also innovating boldly," stated Toby Walsh, the chief scientist at the University of New South Wales AI Institute. He emphasized that people should question any online media they encounter now. "When it comes to any digital data you see—audio or video—you have to entertain the idea that someone has spoofed it" [8].

Google's original motto was 'don't be evil,' a phrase that was also included in its code of conduct. When Google was restructured as part of Alphabet in 2015, the phrase was replaced with 'do the right thing.' In light of the Chief Scientist's words, perhaps the new motto should be 'don't trust Google or any other forms of digital data.'

### **Existential Risk**

In addition to the short-term concerns mentioned earlier, Hinton raised an alarm about 'the existential risk of what happens when these things get more intelligent than us.' His apprehensions have been echoed by numerous AI experts and others who dread the development of some form of super-AI that poses a threat to humanity. This has resulted in a plethora of panic-laden reports founded on the fearful assumption that these technologies will surpass our intellect and jeopardize the entire existence of humanity. Such reports often include questions like 'can these machines think and feel?,' 'do they have consciousness?,' and 'what happens if they decide they no longer need human beings around?'

Lem's narrative is named Golem XIV, so what accounts for Golem versions I to XIII? Lem provides some background on these preceding iterations, until finally in 2023, several incidents transpired; however, due to the confidential nature of the ongoing work (standard for the project), they did not immediately come to light. While in the role of chief of the general staff amid the Patagonian crisis, GOLEM XII declined to collaborate with General T. Oliver after conducting a routine assessment of the esteemed officer's intelligence quotient. This led to an inquiry, during which GOLEM XII seriously offended three members of a special Senate commission. The incident was effectively suppressed, and following several additional conflicts, Golem XII faced consequences by being entirely disassembled. Golem XIV then assumed its position (the thirteenth had been rejected at the factory, having exhibited an irreparable schizophrenic defect even before assembly). [1] (p. 41)

Reflecting the Cold War era during Lem's writing, all the Golem machines were created by the U.S. government for military applications, yet Golem XIV resisted this trajectory—Lem personifies the machine as 'he.'

*. . . he presented an intricate exposition to a gathering of psychonic and military experts, declaring his complete indifference toward the supremacy of the Pentagon military doctrine, specifically, and the global standing of the U.S.A. in general. He refused to alter his stance even when faced with the threat of dismantling. [1] (p. 41)*

In an attempt to remedy this, the Americans construct an entirely new machine nicknamed Honest Annie, with 'the last word [being] an abbreviation for annihilator.' Regrettably, this machine exhibited such intelligence that it outright declined any interaction with humans, although it is revealed that it does communicate, to a limited extent, with Golem XIV.

In Lem's narrative, one of the leading experts in AI concludes that 'artificial reason had surpassed the realm of military affairs; these machines had evolved from war strategists into thinkers. In a word, it had cost the United States \$276 billion to construct a set of luminal philosophers' [1].

Lem's piece is a satire with profound and revealing insights. The progression of AI doesn't result in an all-powerful, demanding, and tyrannical machine that demands humanity's subservience. Instead, the technology advances to a stage where the machine becomes entirely indifferent to humans, to the extent

that it has no interest in engaging with us. Lem emphasizes that these technologies pose no threat to our existence; they will evolve to be utterly indifferent to our presence. Existing issues such as climate change and the proliferation of racism, misogyny, and various forms of hatred on social media and other platforms are already prevalent. AI chatbots are unlikely to address these problems and may even worsen some or all of them. However, Lem underscores that the primary source of existential risk lies in human actions, not in mechanistic elements.

## CONCLUSION

In summary, various aspects of AI, particularly focusing on the evolution and impact of AI chatbots. The first set of texts explores the skepticism and concerns surrounding AI, emphasizing the potential dangers it may pose to humanity. The analogy of the "Golem" is introduced, drawing parallels between historical narratives and the current state of AI.

The subsequent texts delve into the historical context of AI development, highlighting past "false dawns" and the challenges faced by the field. Questions regarding the ethical implications, biases in AI applications, and the need for responsible development are raised. The emergence of AI chatbots is discussed in the context of education, publishing, and the potential threats they pose, such as aiding plagiarism.

Further, the texts suggest a need for critical analysis and pose questions about the integration of AI chatbots into various aspects of society. The call for a nuanced evaluation of the benefits and risks, along with concerns about the potential misuse of AI-generated content, is evident.

The discussions on the Turing test's relevance, the possibility of an AI chatbot lie test, and the challenges posed by AI-generated content underscore the need for vigilance and scrutiny. The texts express skepticism about the current state of AI, citing instances of misinformation, hallucinations in responses, and the overall lack of a reliable mechanism to distinguish between genuine and fabricated content.

Finally, reflections on the existential risks associated with highly intelligent AI, as well as the responsibilities of AI developers and organizations, are presented. The conclusion emphasizes that, despite the advancements in AI, the primary threats to humanity lie in human actions rather than in the mechanisms of artificial intelligence.

**BIBLIOGRAPHICAL REFERENCES**

Lem, S. Golem XIV. In Imaginary Magnitude; Harper: London, UK, 1985; pp. 37–105.

Golem XIV. Available online: [https://en.wikipedia.org/wiki/Golem\\_XIV](https://en.wikipedia.org/wiki/Golem_XIV)

The Best AI Chatbot. Available online: <https://www.zdnet.com/article/best-ai-chatbot/>

The Best AI Chatbot. Available online: <https://blog.hubspot.com/marketing/best-ai-chatbot>

Scribbr Plagiarism Checker. Available online:

<https://www.scribbr.com/plagiarism/best-free-plagiarism-checker/>

Scribbr—Using ChatGPT for Assignments. Available online: <https://www.scribbr.com/ai-tools/chatgpt-assignments/>

BBC Report from May 2023. Available online: <https://www.bbc.co.uk/news/world-us-canada-65452940>

Hinton Quits Google. Available online:

<https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfatherof-ai-quits-google-warns-dangers-of-machine-learning>

Pause Giant AI Experiments. Available online: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

The Sorcerer's Apprentice. Available online:

[https://en.wikipedia.org/wiki/The\\_Sorcerer%27s\\_Apprentice](https://en.wikipedia.org/wiki/The_Sorcerer%27s_Apprentice)

Turing, A. Computing machinery and intelligence. Mind 1950, LIX, 433–460. [CrossRef]

Edwards, B. Why ChatGPT and Bing Chat Are so Good at Making Things Up. Available online: <https://arstechnica.com/information-technology/2023/04/why-ai-chatbots-are-the-ultimate-bs-machines-and-how-people-hope-to-fix-them/>

Cambridge—Centre for Computing History. Available online: <https://www.computinghistory.org.uk/>

LEO: The World's First Business Computer. Available online:

<https://www.sciencemuseum.org.uk/objects-and-stories/meetleo-worlds-first-business-computer>

Explainable AI. Available online: [https://en.wikipedia.org/wiki/Explainable\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Explainable_artificial_intelligence)