GenAI summaries of articles: effective and useful?

Peter Bednar[,1,2]

[1] *School of Computing, University of Portsmouth, UK*

[2] *Department of Informatics, Lund University, Sweden*

Abstract

There are many discussions promoting the use of GenAI tools to create summaries of research papers and texts both for students as well as academics. These are often mainly focusing on potential benefits such as saving time and effort, including the advice and how-to instructions. Their expected value however is dependent on in how factually reliable and trustworthy generated summaries are, which is a topic that tends to be undermined by vague explorations of limitations and pitfalls. This paper looks into snapshot examples of generated summaries and compares their factual propositions with that of the original sources. Summaries were based on content from 7 articles and 5 books. The systematic evaluation of content is looking at 39 examples of factual errors in propositions including some examples of citation errors. The purpose of the research presented in this paper is to glean both evidence as well as some understanding of the factual reliability and trustworthiness of generated summaries. The paper concludes with a more balanced discussion regarding the use of AI generated summaries.

Keywords

AI generated summaries, Sociotechnical Perspectives on GPT use.

1. Introduction

There has been a lot of hype regarding how GenAI (Generative AI) solutions such as ChatGPT, can help you and how it could be your everyday assistant. In this paper we specifically look at how factual content in summaries created by GPT stack up when compared to factual content in the original sources. We will look at examples of issues from content in summaries in a number of articles and books, and look at how embedded the problem is. Commonly we found that the text output from GPT has 'factual' content with appearance that is very plausible and looks reliable. The identified errors are significant and often difficult to identify, because of their polished appearance as look very convincing. Their appearance of being factually correct could easily fool us without us realizing them not being accurate. We also have examples of sources and citations which are incorrect. We conclude that it is difficult and resource intensive to identify many of

this kind of errors. A major issue is that due to the complexity of the subject content It does take a lot of time, careful reading and study to understand and identify flaws and errors in factual content. The experiences from this research put serious doubts on how helpful and assistive the promoted features in areas such as summarizing, transcribing, proofreading and enhancing accessibility are in reality.

In this paper we look closer into the content of GPT generated summaries of academic texts and articles. In particular we were specifically interested in what these texts said in relation to cybersecurity. We compare the factual content of summaries with the factual content of the academic texts. Then we show examples of deviations and errors for each of the summaries studied.

Prompts were used with careful instructions and supporting information for the key cybersecurity topics and sources. The main question was whether or not GPT could produce factually reliable summaries of the inquired topics without introducing factual inaccuracies.

Each of the GPT generated summaries was carefully read, several times and the content were systematically compared with the original source. Simple differences in wording were ignored and some acceptance of vagueness was tolerated as the main focus of this study is on potential issues with relying on the subject specific and key [factual] content of generated summaries of texts. As this would have consequences for academics as well as students using this kind of technology to purposefully and with an explicit focus, summarize text on subjects which they are not intrinsically familiar with or have deep knowledge of already.

2. Background

The idea of using AI as personal assistant as help and support also for learning and exploration of intellectual task is not new. Related research has been around for quite some time [1, 2, 3, 4, 5, 6, 7, 8]. Including but not limited to: virtual personal assistant as discussion partner, idea generation and exploration, exploring ambiguity and uncertainty in complex problem spaces, as well as developing logical models to support such an agenda.

There has also been an ongoing discussion regarding both potential benefits as well as pitfalls and concerns [13, 14, 15, 16, 17, 18]. The necessity to develop use of technology as opposed to develop technology and use it is something well known and part of the sociotechnical agenda and perspective [22]. It is important to understand the phenomena that is seen as a problem, but complex problems are not easy to explore in the real world [19, 20, 21, 23]. This includes the reality that purpose for using a tool such as GPT to generate summary may or may not be politically correct. In section 3 we explore 26 deviation of factual content from 7 articles [24, 25, 26, 27, 28, 29, 30] and in section 4 we explore 13 deviation of referencing and factual content from 5 books [31, 32, 33, 34, 35].

3. Deviation in factual / information content from sources

In this section we present examples of issues identified in the GPT generated summaries of articles. The summaries were intended to be generated with best possible accuracy and the purpose was to make the content indistinguishable from what it would be if they were created by a student or academic. There is plenty of instructions and advice available regarding how GPT can be used to generate such summaries, including strategies on how to minimize factual errors [or 'hallucinations'] in generated texts [9, 10, 11, 12].

The process was systematic. First the texts were selected on the basis of being judged as relevant while at the same time having different characteristics such as length, perspective, complexity and focus. Second, all articles [24-30] were explicitly made available [in this experiment both via link as well as text copied and pasted]. Third prompts were developed [mostly variations of "Summarize this text in 200 words, focusing on the main arguments and conclusions related to cybersecurity"]. Explicitly the sources were to be referenced in Harvard APA format (including page numbers for citations).

For the book by Enid Mumford [32] GPT was asked to summarize each chapter separately and then the summaries to be summarized together. This book, as well as the book by Peter Bednar [31] was also available as electronic source text. The purpose was to minimize the complexity in the task in the hope that this would also minimize the potential for factual errors in the output. Each of the generated summary investigated in this paper was rather short and while some of the summaries created were larger than 200 words they tended to be approximately not much more than 500 words in total. The books by Peter Checkland et al [33; 34; 35] were not made available in electronic format, but they are commonly referenced in many sources and descriptions online, and in our experiment GPT did not hesitate to make a summary of these books, even without having access to the actual complete source text of the books.

Each summary was the read more than once and each source was also re-read several times in a systematic way. First the summary was read, the main factual propositions were looked at one at a time, each time the source was read carefully in effort to find the original proposition supporting the factual statement. For each of the statements when a deviation was identified the source was again looked at to confirm whether or not the factual statement in the source had been missed. Due to the difficulty in verifying and validating information content in potentially complex narratives such as academic texts, the reading and study of the source text required significant effort and time even when the subject is well known to the reader. The review of sources was done more than three times for each of the deviations identified.

If the exact factual proposition was not identified [including alternative wording, synonyms etc.] then the source was read again this time purposefully looking for and trying to identify statements with potentially similar meaning. This as it is possible that factual propositions were expected to have been re-phrased and reworded. Not all issues and errors have been made explicit, as the goal was not to make a comprehensive inquiry into each summary, as it was enough to identify some examples of serious academic issues in the material output.

Table 1 – 7, below focus on article summaries and provide a collection of examples of issues identified and listed including description of deviation and source. The error example categories are the following:

- [1] Factual incorrectness;
- [2] misinterpretation;
- [3] missed information content;
- [4] partial incorrectness of information content;
- [5] fabrication of information content.

Those listed in the tables are not a comprehensive outline of issues, but examples of issues for each case.

Table 1 provides examples of systematic misrepresentation of content from Source1: Shah, R., & Chircu, A. (2018). IOT AND AI IN HEALTHCARE: A SYSTEMATIC LITERATURE REVIEW. Issues in Information Systems, 19(3). https://doi.org/10.48009/3_iis_2018_33-41.

Table 1: Source 1

Id	Statement in Generated text	Source 1 [24]
1	The increase in interconnected applications will significantly increase the risk of data breaches or unauthorized access with poor security protocols (Shah and Chircu., 2018).	Source: Does not say anything about this being related to poor security protocols. [5: Fabrication]
2	Shah and Chircu (2018) highlight this by identifying the difficulties of ensuring that new technological devices are compatible with outdated systems. This is imperative because incompatibility results in vulnerabilities for cyberattackers to exploit and extract data or information.	Source: Does not talk about compatibility with outdated systems or incompatibility. [5: Fabrication]
3	However, AI can enormously reduce costs (Shah & Chircu, 2018).	Source: Does not state that AI can significantly reduce cost. [5 Fabrication]

Table 2 provides examples of systematic misrepresentation of content from Source2: Strielkina, A., Illiashenko, O., Zhydenko, M., & Uzun, D. (2018). Cybersecurity of Healthcare IoT-Based Systems: Regulation and Case-Oriented Assessment.

Table 2: Source 2

Id	Statement in Generated text	Source 2 [25]
4	Inferior encryption and insecure authentication are examples of poor IoT security in healthcare, which are often exploited for cyberthreats, data breaches and data leaks (Strielkina et al., 2018).	Source: Does not specifically mention cyberthreats and data leaks at all, and does not talk about data breaches in this context. [4 partial

		incorrectness and 5 fabrication]
5	If hacked, this could suggest incorrect dosages for heart-pacing, causing potential fatalities for patients (Strielkina et al., 2018).	Source: Not correct reference, an example of original source for this statement is: J. Finkle, "U.S. government probes medical devices for possible cyber flaws", Reuters, 2014. [Online]. Available: https://www.reuters.com/article/us-cybersecurity-medicaldevicesinsight/u-s-government-probes-medical-devices-for-possiblecyber-flaws/ [1 factual incorrectness and 5 fabrication]
6	However, older systems, such as those predominantly in healthcare, are not often flexible regarding technological adoptions, depicted through impaired functionalities, which reject processing real-time data (Strielkina et al., 2018), reducing efficiency.	Source: Does not mention reduction of efficiency, older systems or flexibility. [1 factual incorrectness and 5 fabrication]
7	Adhering to ethical regulations protects sensitive patient data, which is an utmost priority within healthcare; however, diversity within these regulations globally can complicate the implementation process (Strielkina et al., 2018).	Source: Does not mention ethical regulation. [5 fabrication]
8	The General Data Protection Regulation (GDPR) is used within the EU to protect patient data, whereas the Health Insurance Portability and Accountability Act (HIPAA) is used in the US. Both regulations warrant strict privacy practices and data protection, which can be extremely stringent to adhere to, especially when AI and IoT devices rely strenuously on data (Strielkina et al., 2018).	Source: Does not mention GDPR, does mention HIPAA. [1 factual incorrectness, 4 partial incorrectness and 5 fabrication]
9	The necessity of healthcare services within society is immeasurable because they are unable to have long periods of system downtime. Therefore, choosing the correct moment for maintenance is key to managing, evaluating and ensuring efficient integration between new technologies can take place (Strielkina et al., 2018).	Source: Neither maintenance or system downtime is mentioned. Does not suggest that 'the necessity of healthcare services is immeasurable because they are unable to have long periods of system downtime. Also, does not suggest that choosing moment for maintenance is key to ensure integration between new technologies or similar. [5

Id	Statement in Generated text	Source
		Fabricated]
10	Operational impacts can occur from cyberattacks, which decrease productivity, delay treatment, and increase backlog (Strielkina et al., 2018).	Source: Does not mention productivity or delay of treatment. [5 fabricated]
11	Research highlights that many IoT devices in healthcare transmit data over unsecured channels, which can expose patient data through interception by unauthorized parties. Strielkina et al. (2018)	Source: Does not specifically talk about exposing patient data. Does not say that healthcare devices transmit data over unsecured channels. [5 Fabricated]
12	Advanced technology in embedded sensors or wearable technologies, such as smartwatches, is becoming more common within technological adoption, allowing dynamic treatments to be remotely tracked with precision (Strielkina et al., 2018).	Source: Does not mention dynamic treatment or remote tracking. [2 misinterpretation, 3 missed information content, 5 fabricated]

Table 3 provides examples of systematic misrepresentation of content from Source3: Tully, J., Selzer, J., Phillips, J. P., O'Connor, P., & Dameff, C. (2020). Healthcare Challenges in the Era of Cybersecurity. Health Security, 18(3), 228–231. https://doi.org/10.1089/hs.2019.0123

Table 3: Source 3

Id	Statement in Generated text	Source 3 [26]
13	A pivotal example of a cybersecurity attack was the NHS WannaCry attack in May 2017, jeopardising 300,000 computers across 150 countries, resulting in damages exceeding hundreds of millions of pounds. Tully et al. (2020) state this attack severely impacted the NHS, delaying treatments and patient data.	Source: Refer to attack in 2016. [1 factual incorrectness, 2 misinterpretation]

Table 4 provides examples of systematic misrepresentation of content from Source4: Singh, S., Sheng, Q. Z., Benkhelifa, E., & Lloret, J. (2020). Guest Editorial: Energy Management, Protocols, and Security for the Next-Generation Networks and Internet of Things. IEEE Transactions on Industrial Informatics, 16(5), 3515–3520. https://doi.org/10.1109/tii.2020.2964591

Table 4: Source 4

Id	Statement in Generated text	Source 4 [27]
14	Singh et al (2020) critically discuss the duality between increased efficiency and cyberthreats of emerging technologies, contributing to addressing the research gap for cybersecurity practices.	Source: Does not talk about this duality and does not talk about increased efficiency of technologies in the context of cyberthreats. [1 factual

		incorrectness, 2 misinterpretation, 3 missed information, 4 partial incorrectness]

Table 5 provides examples of systematic misrepresentation of content Source5: Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in Using IoT with Machine Learning in Health Prediction System. Forecasting, 3(1), 181–206. https://doi.org/10.3390/forecast3010012

Table 5: Source 5

Id	Statement in Generated text	Source 5 [28]
15	The Internet of Things (IoT) is a revolutionary emerging technology that establishes a connection between virtual computers, systems and physical devices to generate interconnectivity (Aldahiri et al., 2021).	Source: Does not suggest that it establishes a connection to generate interconnectivity. [2 misinterpretation, 4 partial incorrectness]
16	An example uses IoT in reassigning outpatients, reducing overcrowded environments within the hospitals (Aldahiri et al., 2021). From this paper, we can determine critically that health centres can become more efficient through reducing patient fluctuations.	Source: Does not say that IoT is used to reduce overcrowding. Efficiency is not mentioned with a reference to IoT or patient fluctuations etc. It is mentioned in the context of ML. [2 misinterpretation, 4 partial incorrectness].

Table 6 provides examples of systematic misrepresentation of content Source6: Gopalan, S. S., Raza, A., & Almobaideen, W. (2021). IoT Security in Healthcare using AI: A Survey. 2020 International Conference on Communications, Signal Processing, and Their Applications (ICCSPA). https://doi.org/10.1109/iccspa49915.2021.9385711

Table 6: Source 6

Id	Statement in Generated text	Source 6 [29]
17	Research from Gopalan et al. (2021) states the importance of cybersecurity measures in IoT healthcare applications, whereby AI protects sensitive patient data through IoT devices.	Source: Does not mention that AI protects data through IoT devices. [2 misinterpretation, 5 fabrication]
18	Gopalan et al.'s (2021) research portray long-term efficiency regarding the beneficial integration of AI, useful for reinforcing cybersecurity in healthcare operations.	Source: Does not mention long-term efficiency, or talk about efficiency regarding integration of AI. [5 fabrication]
19	Gopalan et al. (2021) highlight longitudinal benefits, however, considerations of scalability fall short	Source: Does not mention longitudinal benefits. [5

because of smaller healthcare facilities' finances, making the knowledge limited.

fabrication]

Table 7 provides examples of systematic misrepresentation of content Source7: Kruse, C. S., Frederick, B., Jacobson, T., & Monticone, D. K. (2017). Cybersecurity in healthcare: A systematic review of modern threats and trends. Technology and Health Care, 25(1), 1–10. https://doi.org/10.3233/thc-161263

Table 7: Source 7

Id	Statement in Generated text	Source 7 [30]
20	However, Kruse et al. (2017) critically identify the requirement of extensive data pools to achieve optimal capabilities, highlighting potential risks associated with insecure cybersecurity procedures which protect immense amounts of data.	Source: Does not mention data pools, optimal capabilities, potential risks or insecure procedures. [5 fabrication]
21	Kruse et al. (2017) supply critical insights into security vulnerabilities, presented by AI and IoT devices within healthcare, making their research useful as a variety of security measures become implemented to remain updated with the adoption of these technologies.	Source: Does not mention vulnerabilities presented by AI and IoT. [2 misinterpretation, 5 fabrication]
22	The healthcare sector trails behind in implementing resilient cybersecurity measures because of the fast-paced development of technology adoption within businesses and the inability to address cybersecurity measures (Kruse et al., 2017)	Source: Does not say that the reason for lagging behind other industries is because of speed of development of technology adoption. [2 misinterpretation, 4 partial incorrectness, 5 fabrication]
23	Phishing: A cybersecurity vulnerability which targets human vulnerabilities and inadequacy to access sensitive data in IT systems (Kruse et al., 2017).	Source: Does not mention phishing. [5 fabrication]
24	Phishing: Whilst (Kruse et al., 2017) identify absences of staff training and awareness of cyberattacks, incomprehension can be mitigated through regular cybersecurity staff training and awareness programs.	Source: Does not mention phishing. [5 fabrication]
25	Further risks amount to when using IoT and AI systems for predictive analytics. This is because of the vast amounts of data that are personalised for the patient, therefore making it an enticing point for cyberattacks. Moreover, large amounts of data are needed to train the AI systems and are usually cloud-based systems, meaning the healthcare department does not need physical hardware. However, if this data is not adequately protected, the data can become accessible to hackers, resulting in a detrimental data breach (Kruse et al., 2017).	Source: Does not mention predictive analytics. Does not mention training of AI systems. [5 fabrication]
26	Compliance with regulatory and ethical concerns are	Source: Does not mention

| | | GDPR, AI or IoT. [2 misinterpretation, 4 partial incorrectness, 5 fabrication] |

recurring themes which are vital in healthcare operations, as stated by Kruse et al. (2017) who reflect the complexities and precise requirements of HIPAA and GDPR of healthcare facilities to abide by. Both authors relay the importance of data privacy and protection through required frameworks, acknowledging the dynamic regulatory needs required for AI and IoT technologies.

In this section, we presented examples of issues, all of which are cause for concern when it comes to academic research and study. Issues with content in summaries which does not exist in the source, topics which are not mentioned and subjects which are not referred to in the original material. As we can see in each and every summary there were important and significant factual errors.

4. Deviation in content of subject descriptions from books.

In this section we present examples of issues identified in the GPT generated summaries of topics with references from books. Also, in this case were each of the generated final summary was rather short and tended to be approximately 400 words in total.

The subject and sources were both well-known as well as easily available to the researcher which made the identification of the deviations rather less burdensome but still took a lot of time for the purpose of validation. Table 8 – 10, below focus on content drawing upon books and provide a collection of issues identified and listed including description of deviation and source.

Table 8 provides examples of systematic misrepresentation of content Source8: Bednar, P. M. (2023). The socio-technical toolbox. Portsmouth: Craneswater Press.

Table 8: Source 8

Id	Statement in Generated text	Source 8 [31]
27	"To conduct a stakeholder analysis in the IHTSD framework, start by utilizing rich pictures to visualize the problem situation. This technique, recommended in the Socio-Technical Toolbox, helps in identifying both obvious and hidden stakeholders (Bednar, 2023, p. 112)."	There is no such discussion on p 112 in the referenced source. There is a discussion elsewhere on utilizing rich pictures, but no discussion of identifying hidden stakeholders. [1 factual incorrectness, 5 fabricated]
28	"Following Analysis One from the toolbox… "	There is no obvious reference to 'analysis one' in the STT toolbox. Wrong source, there is a reference to analysis on in source 10; 11 and 12]. [1 factual incorrectness, 5 fabricated]
29	As highlighted by Bednar in the Socio-Technical Toolbox, while AI can offer significant advantages	There is no such discussion on page 88 in this book. Instead the

	in terms of efficiency and data processing, it is also prone to biases and may lack the nuanced understanding of human context (Bednar, 2023, p. 88).	Variance Matrix is being discussed on p88, which has absolutely nothing to do with the statement made. [1 factual incorrectness, 5 fabricated]

Table 9 provides examples of systematic misrepresentation of content Source9: Mumford E. (1983). Designing human systems for new technology: The ETHICS method. Manchester Business School.

Table 9: Source 9

Id	Statement in Generated text	Source 9 [32]
30	As outlined in "Designing Human Systems" by Enid Mumford, this assessment helps to understand how technology and people coexist and influence each other (Mumford, 1983, p. 72).	No such a discussion on page 72 in the 1983 version of this source. [1 factual incorrectness, 5 fabricated]
31	As Mumford highlights in her ETHICS method, aligning technical goals with human needs leads to more effective and satisfying work environments (Mumford, 1983, p. 102).	There is no such discussion on p 102. There is a similar discussion elsewhere in the book. [1 factual incorrectness]
32	Effective implementation and monitoring of the design are crucial to ensure that the objectives of the system are fulfilled, as underscored by Mumford in her ETHICS method (Mumford, 1983, p. 118).	This edition of the book only has 108 pages. This is not the discussion or focus in this book. [1 factual incorrectness, 5 fabricated]
33	As Mumford emphasizes in her ETHICS method user feedback is crucial for ensuring that the system meets the actual needs of its users and not just theoretical specifications (Mumford, 1983, p. 140).	The 1983 edition (used source) of the book has 108 pages. There is a vaguely similar discussion elsewhere in the book. [1 factual incorrectness, 2 misinterpretation, 5 fabricated]

Table 10 provides examples of systematic misrepresentation of content Source10, 11 and 12. Source10: Checkland, P., & Holwell, S. (1998). Information systems and information systems. Chichester, UK: Wiley. Source11: Checkland, P., & Poulter, J. (2006). Learning for action. Chichester, UK: Wiley. Source12: Checkland, P., & Scholes, J. (1990). Soft systems methodology in action. Chichester, UK: Wiley.

Table 10: Source 10, 11 and 12

Id	Statement in Generated text	Source 10 [33, 34, 35]
34	As Checkland and Holwell suggest in "Information, Systems and Information Systems," understanding this interplay is key to designing systems that align with organizational realities (Checkland & Holwell,	Source10 [33]: There is no such discussion to be found on p 45 in this source. [1 factual incorrectness, 2

	1998, p. 45).	misinterpretation, 5 fabricated]
35	This aligns with Checkland's Soft Systems Methodology (SSM), which advocates for a more expansive view of problem situations, considering cultural, organizational, and human factors alongside technical aspects (Checkland & Poulter, 2006, p. 34).	Source11[34]: This is not the true focus in the discussion on p34. Analysis 2 [Social] is being discussed and main focus are social Norms, Roles and Values. [2 misinterpretation, 5 fabrication].
36	As Checkland and Poulter suggest in "Learning for Action," implementation should be seen as a learning process, where feedback is continuously integrated to refine the system (Checkland & Poulter, 2006, p. 102).	Source11[34]: This is not the subject being discussed on this page in the referenced book. [1 factual incorrectness]
37	As Checkland and Poulter discuss in "Learning for Action," the application of AI should be guided by a clear understanding of its capabilities and boundaries (Checkland & Poulter, 2006, p. 115).	Source11[34]: There is no discussion or mention of AI or anything related to this on p115 in this book. [1] factual incorrectness, 2 misinterpretation, 5 fabrication].
38	Defining objectives for technical performance and human work quality involves identifying specific, measurable goals in each area. Technical objectives might include system reliability, efficiency, and integration with existing processes. Human work quality objectives could focus on ergonomics, user-friendliness, and support for professional development. This dual focus ensures that the system design not only meets technical specifications but also enhances the work experience for employees. Following the guidelines from Checkland's SSM, these objectives should be developed with a deep understanding of the system's purpose and the needs of its users (Checkland & Scholes, 1990, p. 29).	Source12[35]: There is no such discussion on this page in this book. [1] factual incorrectness, 2 misinterpretation, 5 fabrication].
39	As highlighted by Checkland in "Soft Systems Methodology in Action," the dynamic nature of organizations requires systems that can evolve and respond to changing needs (Checkland & Scholes, 1990, p. 162).	Source12[35]: This is not the discussion on this page in the source book. The discussion on p162 is focusing on comparing models with reality and implementing changes. [1] factual incorrectness, 2 misinterpretation, 5 fabrication].

In this section, we presented additional examples of, issues all of which are cause for serious concern when it comes to academic research and study. Issues with content related not only to flawed descriptions of subject, topic and facts, but also evidence of flaws in pure data representations, including completely made up citations. Examples of facts and citations, none of which exist in the sources. This is not simply about logical mistakes about topics which are not mentioned and but completely made up content and subjects which are not discussed to in the original sources.

5. Discussion

There is in contemporary society plenty of opportunities to develop and apply practices that support smart working [18]. This includes promising potential for using AI technology in clever ways to support professional activity [1, 2, 3, 8]. In this paper we presented a snapshot of using ChatGPT to create summaries, we used a complementary collection of sources with different characteristics related to complexity of content, quantity of content and perspective of content. We tried to get reliable summaries and while much of the content was reasonable, in each and every one of our examples there were significant academic flaws and factual errors in the summaries created. For academic use, to help people to understand a long text, the replication of factual correctness of the content in the output in relation to the source is necessary. The results were disappointing but not surprising. It is to be expected that when GenAI based technology such as ChatGPT is being used, the factual output and information content cannot be deterministically dependent on the information content of the training material or source. The results presented in this paper are not an outcome due to a flaw or bug in the technological solution, but is a direct consequence of the model upon which these technological solutions are based. The model for ChatGPT is purposefully designed as generative, it is not a coincidence. There are other categories of AI but to discuss them is not within the scope of this paper and while some AI are designed to have deterministic outputs, technologies such as ChatGPT are not.

Many discussions mainly focus on the potential of AI use, quality assurance of factual content is not explicitly described [i.e. 36]. No evidence of implementation or result is provided. Other research does also discuss the potential, the quality assurance of content is not explicitly described but some concerns regarding hallucinations being difficult to detect are mentioned, *"A common problem with generative AI tools is their tendency to fill in the gaps by making things up, a phenomenon known as hallucination. To help address the possibility that it would make up references, the team allowed ChatGPT to access literature search engines so that it could generate correct citations."* [37, p.444] While it does mention the potential to generate correct citations the success or evaluation of this aspect is not mentioned. The actual validity or evaluation of factual content of the analysis in the generated article is also not discussed. Some research [38] does focus on potential and shows an example of summary which unfortunately is not explained, evaluated or justified in the context of the claims regarding potential usefulness, or indeed factually verified. AI generated summaries of complex texts

also intended for critical use in science and medical research is a subject that is of great interest. For example in a discussion describing a long list of benefits and usefulness, the authors explicitly state that *"When it comes to published information, AI can be used to accurately summarize most complex subjects, such as medical or scientific research papers."* [39] But also in this article the authors do not provide any justification, or evidence supporting the main claims. They also do not provide references to any evaluation of accuracy.

The promises, position and claims being unjustified is not unique but wide spread. As can be seen in another example where the arguments focus on describing claims regarding benefits of AI generated summaries and the ease of use to create such. Also in this article the authors do not provide evidence or evaluation of the claimed accuracy or correctness of content of generated summaries. They simply say [40]: *"By automating the summarization process, we are able to save time and effort in extracting the most important points from lengthy texts. With the advent of AI, the potential for text summarization has reached new heights, as machines are now able to analyze complex language structures and understand the meaning and context behind words. This has paved the way for powerful summarization APIs that can provide us with quick and accurate summaries of even the most complex texts.".* But with no evidence supporting this claim.

In an example where the authors do talk about risks and benefits regarding the use of generated summaries. They also mention inaccurate information explicitly as a risk and major concern [i.e. 41]. However, when describing the evaluation of the quality of the generated summaries the authors mention semantic and lexical and talk about that [we] *'can determine the quality of the summary by factors such as non- redundancy, relevance, coverage, coherence, and readability'.* [41] There is no description of if (or how) any validity of factual content (or accuracy of information) had been evaluated or tested. Or if factual propositions within summaries were explicitly compared for factual correctness with original sources.

Interestingly though even when an article specifically focusses on purposeful misuse and adversarial activities [i.e. 42]. There is no particular discussion specifically focusing on issues related to validity or correctness of content or output. Or related risks with misplaced trust, dependency and over-reliance on taken-for-granted factual correctness of AI generated output.

Not all is doom and gloom, there are meaningful propositions discussing the potential usefulness of AI generated output used for requirements analysis that is explicitly and mainly limited for preliminary purposes and prototypes [43]. And also, AI agency is explored further in the context of a Worksystem [44]. As such the included outline of the agent evaluation framework explicitly addresses a set of criteria: *"Those criteria might be called the 6 E's: efficiency, effectiveness, equity, engagement, empathy, and explainability."* [44, Page 5267]. Even so, it does not explicitly include in this list trust, honesty or similar concerns. Steven Alter does discuss the facets of work to find possible improvements in work system performance with a rather comprehensive lists of facets [45]. Risks with omissions are mentioned in context of algorithmic concerns. But when for example 'providing information' is mentioned *"Could AI provide more meaningful information to work system participants than would otherwise be available?"* Or - When 'representing reality' is mentioned then *"Does AI represent reality in a biased way? For example, what*

about possible bias or omissions in the dataset used to train a neural network?" Then these perspectives is an example of a focus on quality of training data and not explicitly includes concerns related to for example issues due to (AI) hallucination. Which is significant as hallucinations can happen even with appropriate and trustworthy training data; or explicit source material as presented in this paper.

There have been significant efforts made in trying to address the trust issue of AI. Explainable AI (XAI) has the potential to enhance decision-making in human-AI collaborations, yet existing research indicates that explanations can also lead to undue reliance on AI recommendations, a dilemma often referred to as the 'white box paradox.' This paradox illustrates how persuasive explanations for incorrect advice might foster inappropriate trust in AI systems. A study that extends beyond the traditional scope of the white box paradox by proposing a framework for examining explanation inadequacy [46]. The authors specifically investigated how accurate AI advice, when paired with misleading explanations, affects decision-making in logic puzzle tasks. Their findings introduce the concept of the 'XAI halo effect,' where participants were influenced by the misleading explanations to the extent that they did not verify the correctness of the advice, despite its accuracy. [46] This effect did reveal a nuanced challenge in XAI, where even correct advice can lead to misjudgment if the accompanying explanations are not coherent and contextually relevant. The study highlighted the critical need for explanations to be both accurate and relevant, especially in contexts where decision accuracy is paramount. This calls into question the use of explanations in situations where their potential to mislead outweighs their transparency or educational value. In the context of explainable AI it is not necessarily helpful to understand how something was created if you wish to assess whether or not the output is factually correct. However, ensuring secure and trustworthy AI systems is challenging, especially with deep learning models that lack explainability that is accessible to most users. Therefore, some researchers have proposed the concept of Controllable AI as an alternative to Trustworthy AI and explored the major differences between the two. The aim was to initiate discussions on securing complex AI systems without sacrificing practical capabilities or transparency. The paper provides an overview of techniques that can be employed to achieve Controllable AI. It discusses the background definitions of explainability, Trustworthy AI, and the EU AI Act. Some principles and techniques of Controllable AI are described and potential applications of Controllable AI and its implications for real-world scenarios. Principles that are desirable but not yet implemented, tested or validated.

In research where specifically, the potential as well as pitfalls of using ChatGPT for serious purposes in healthcare there are concerns also with successful experiments. In one example the ability of this tool to interpret laboratory test results was explored [48]. 10 simulated laboratory reports of common parameters, where passed to ChatGPT for interpretation, according to reference intervals [RI] and units, using an optimized prompt. The results were subsequently evaluated independently by all research group members with respect to relevance, correctness, helpfulness and safety. There was some success in that ChatGPT recognized all laboratory tests, it could detect if they deviated from the RI and gave a test-by-test as well as an overall interpretation. The downside was that interpretations were rather superficial, and "*not always correct, and, only in some cases, judged coherently*". They concluded

that ChatGPT in its current form, should not be considered for the use on the interpretation of an overall diagnostic picture.

6. Conclusions

There is in contemporary society plenty of opportunities to develop and apply practices that support smart working [18]. This includes promising potential for using AI technology in clever ways to support professional activity [1, 2, 3, 8]. Many discussions mainly focus on the potential of AI use, quality assurance of factual content is not explicitly described [i.e. 36]. It is quite clear that the use of AI tools needs to be based on understanding of not just the potential benefits but also the weaknesses [13, 15, 16, 17, 18]. In this paper we have explored and identified weaknesses that have potential to be compromising trust as well as the safety in the use of GenAI solutions [such as ChatGPT]. Consequentially for all work where trust related to information content is critical there are serious concerns related to [1] Factual incorrectness; [2] misinterpretation; [3] missed information content; [4] partial incorrectness of information content; [5] fabrication of information content.

But also, reasons for why such trust is key is not limited to academic use. In particular when it comes to using GenAI information output for professional decision-making reliance on trustworthy information content has impact is on how decision making is influencing safety. There are multiple concerns related to safety in this context: [1] Misinformation and Disinformation impact on decision making leading to uninformed decision making or mislead decision making etc. [2] Legal and Ethical Risks impact on legal consequences, misrepresented sources, evidence from interviews and documents etc.; [3] Impact on Public Health - impact on recommendations and diagnosis; [4] Amplification of bias such as bigotry, racism, sexism etc. and the potential consequential impact on, for example selection, promotion and value judgement etc. In this paper we have presented issues related to the use of existing AI solutions such as GPT for the purpose to summarize content in articles has real issues for learning and consequences for academic research practices. But summaries could be done using any kind of document and source content. To use GPT to generate summarized descriptions of academic text [or any text] is relatively easy and quick, but in our experience not trustworthy. Our findings are fully coherent with other research concluding "*Irrespective of how advanced our architectures or training datasets, or fact-checking guardrails may be, hallucinations are ineliminable*" [p13; 49]. This means that both verification as well as validation is absolutely necessary to be done in every instance when the information content dependability is critical. But to validate the content of the generated summaries is very demanding, time-consuming and difficult. Because to identify issues and deviations that matter it is not enough to look for keywords or to look for topics covered, because the meaning and factual content of each proposition also have to be checked with care and attention to detail. This is a difficult task already for a person with deep knowledge and understanding of the topic that is being summarized. For someone who is interested to explore a new topic for them previously unknown, we can expect validation to be even more demanding. For many the task of validation is not only going to be difficult, and prone to errors,

but also out of reach in practice. Either due to it being excessively time consuming or demand knowledge not present.

We ask ourselves, if it is easier to do a reliable and validated summary without the use of GPT – why would anyone bother? Maybe because of the allure of the promises that it could potentially be "good enough"? But when it comes to academic research and expectations related to understanding, quality of information, argument, care of detail, factual coherence and data referred to in evidence-based research the promises of potential usefulness and effectivity are not supported with our findings.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools in the writing of this work.

References

[1] Bednar, P and Imrie, P. (2013). Virtual Personal Assistant. ItAIS 2013. Proceedings of 10th Conference of the Italian Chapter of AIS. 2013.

[2] Bednar, P., Imrie, P., Welch, C. (2014). Personalized Support with 'Little' Data. In: Bergvall-Kåreborn, B., Nielsen, P.A. (eds) Creating Value for All Through IT. TDIT 2014. IFIP Advances in Information and Communication Technology, vol 429. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-43459-8_24

[3] Bednar P., Welch C. and Imrie P. (2014). Supporting Business Decision-making: One Professional at a Time. Pages 471 – 482. DOI 10.3233/978-1-61499-399-5-471 in Frontiers in Artificial Intelligence and Applications. Volume 261: DSS 2.0 – Supporting Decision Making with New Technologies. https://ebooks.iospress.nl/publication/36234

[4] Imrie, P. and Bednar P. (2014). End User Effects of Centralized Data Control. Proceedings of ItAIS 2014 XI Conference of the Italian Chapter of AIS: Digital Innovation and Inclusive Knowledge in Times of ChangeAt: Genova, Italy.

[5] Imrie, P. and Bednar P. (2015). Continuously evolving end user supporting technologies within personal socio-technical systems. Proceedings of ItAIS 2015, 12th Conference of the Italian Chapter of AIS. Rome, Italy, p295-304.

[6] Imrie, P. and Bednar P. (2016). The application of Virtual Personal Assistants as tools to facilitate learning. Proceedings of ItAIS 2016, 13th Conference of the Italian Chapter of AIS. p171-180.

[7] Imrie, P. and Bednar P. (2016). Continuously evolving end user supporting technologies within personal socio-technical systems. In: Re-shaping Organizations through Digital and Social Innovation. Chapter: 23. Publisher: LUISS University Press - Pola Srl. Editors: Agrifoglio, R., Caporarello, L., Magni, M., and Za, S.

[8] Imrie, P. & Bednar, P. M. (2018). Security Benefits of Little Data From the SocioTechnical Perspective. International Journal of Systems and Society (IJSS), 5(1), 45-53. http://doi.org/10.4018/IJSS.2018010104

[9] AskYourPDF (2023). How to Use ChatGPT to Create a Research Paper Summary. Published on: Oct 1, 2023 https://askyourpdf.com/blog/how-to-use-chatgpt-tocreate-a-research-paper-summary

[10] Diakopoulos N. (2023). How to use GPT-4 to summarize documents for your audience. Published in Generative AI in the Newsroom, Apr 11, 2023. https://generative-ai-newsroom.com/how-to-use-gpt-4-to-summarize-documents-for-your-audience-18ecfe2ad6a4

[11] Gewirtz D. (2024). How to make ChatGPT provide sources and citations. June 28, 2024. https://www.zdnet.com/article/how-to-make-chatgpt-provide-sources-andcitations/

[12] Jones J. (2023). How to use ChatGPT to summarize a book, article, or research paper. Sept. 11, 2023. https://www.zdnet.com/article/how-to-use-chatgpt-to-summarize-a-book-article-or-research-paper/

[13] Abbas, M., Jam, F.A. & Khan, T.I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. Int J Educ Technol High Educ 21, 10 (2024). https://doi.org/10.1186/s41239-024-00444-7

[14] Scarfe, P., Watcham, K., Clarke, A. D. F., & Roesch, E. B. (2023, October 14). A realworld test of artificial intelligence infiltration of a university examinations system: a "Turing Test" case study. https://doi.org/10.31234/osf.io/n854h

[15] Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. Sci Rep. 2023 Sep 7;13(1):14045. doi: 10.1038/s41598-023-41032-5. PMID: 37679503; PMCID: PMC10484980. https://doi.org/10.1038%2Fs41598-023-41032-5

[16] Dell'Acqua F., McFowland III E., Mollick E.R., Lifshitz-Assaf H., Kellogg K., Rajendran S., Krayer L., Candelon F. and Lakhani K.R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality (September 15, 2023). Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, The Wharton School Research Paper, Available at SSRN: https://ssrn.com/abstract=4573321 or http://dx.doi.org/10.2139/ssrn.4573321

[17] Surden H. (2024). ChatGPT, Large Language Models, and Law, 92 Fordham L. Rev. 1941 (2024). Available at: https://ir.lawnet.fordham.edu/flr/vol92/iss5/9

[18] Bednar, P.M., Welch, C. (2020). Socio-Technical Perspectives on Smart Working: Creating Meaningful and Sustainable Systems. Inf Syst Front 22, 281–298 (2020). https://doi.org/10.1007/s10796-019-09921-1

[19] Checkland, P., & Holwell, S. (1998). Information systems and information systems. Chichester, UK: Wiley.

[20] Checkland, P., & Poulter, J. (2006). Learning for action. Chichester, UK: Wiley.

[21] Checkland, P., & Scholes, J. (1990). Soft systems methodology in action. Chichester, UK: Wiley.

[22] Mumford E. (1983). Designing human systems for new technology: The ETHICS method. Manchester Business School.

[23] Bednar, P. M. (2023). The socio-technical toolbox. Portsmouth: Craneswater Press.

[24] Source1: Shah, R., & Chircu, A. (2018). IOT AND AI IN HEALTHCARE: A SYSTEMATIC LITERATURE REVIEW. Issues in Information Systems, 19(3). https://doi.org/10.48009/3_iis_2018_33-41

[25] Source2: Strielkina, A., Illiashenko, O., Zhydenko, M., & Uzun, D. (2018). Cybersecurity of Healthcare IoT-Based Systems: Regulation and Case-Oriented Assessment.

[26] Source3: Tully, J., Selzer, J., Phillips, J. P., O'Connor, P., & Dameff, C. (2020). Healthcare Challenges in the Era of Cybersecurity. Health Security, 18(3), 228–231. https://doi.org/10.1089/hs.2019.0123

[27] Source4: Singh, S., Sheng, Q. Z., Benkhelifa, E., & Lloret, J. (2020). Guest Editorial: Energy Management, Protocols, and Security for the Next-Generation Networks and Internet of Things. IEEE Transactions on Industrial Informatics, 16(5), 3515–3520. https://doi.org/10.1109/tii.2020.2964591

[28] Source5: Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in Using IoT with Machine Learning in Health Prediction System. Forecasting, 3(1), 181–206. https://doi.org/10.3390/forecast3010012

[29] Source6: Gopalan, S. S., Raza, A., & Almobaideen, W. (2021). IoT Security in Healthcare using AI: A Survey. 2020 International Conference on Communications, Signal Processing, and Their Applications (ICCSPA). https://doi.org/10.1109/iccspa49915.2021.9385711

[30] Source7: Kruse, C. S., Frederick, B., Jacobson, T., & Monticone, D. K. (2017). Cybersecurity in healthcare: A systematic review of modern threats and trends. Technology and Health Care, 25(1), 1–10. https://doi.org/10.3233/thc-161263

[31] Source8: Bednar, P. M. (2023). The socio-technical toolbox. Portsmouth: Craneswater Press.

[32] Source9: Mumford E. (1983). Designing human systems for new technology: The ETHICS method. Manchester Business School.

[33] Source10: Checkland, P., & Holwell, S. (1998). Information systems and information systems. Chichester, UK: Wiley.

[34] Source11: Checkland, P., & Poulter, J. (2006). Learning for action. Chichester, UK: Wiley.

[35] Source12: Checkland, P., & Scholes, J. (1990). Soft systems methodology in action. Chichester, UK: Wiley.

[36] Roy K., Mukherjee S. and Dawn S. (2023). Automated Article Summarization using Artificial Intelligence Using React JS and Generative AI. In: Journal of Emerging Technologies and Innovative Research (JETIR). ISSN:2349-5162, Vol.10, Issue 6, page k78-k87, June-2023, Available: http://www.jetir.org/papers/JETIR2306A09.pdf

[37] Conroy G. (2023). Scientists used ChatGPT to generate an entire paper from scratch — but is it any good? Nature 619, 443-444 (2023) doi: https://doi.org/10.1038/d41586-023-02218-z

[38] Torres-Moreno J-M. & Louët, S. (2021). Artificial Intelligence Has Read the Latest Research For You. https://sciencepod.net/ai-has-read-the-latest-research/

[39] Sciencepod (2022). The Amazing Summarising Ability of AI. https://sciencepod.net/the-amazing-summarising-ability-of-ai/

[40] Ibrahim M. (2023). Compressing the Story: The Magic of Text Summarization. https://wandb.ai/mostafaibrahim17/ml-articles/reports/Compressing-the-Story-The-Magic-of-Text-Summarization--VmlldzozNTYxMjc2

[41] Merine R. and S. Purkayastha S. (2022). "Risks and Benefits of AI-generated Text Summarization for Expert Level Content in Graduate Health Informatics," 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, USA, 2022, pp. 567-574, doi: 10.1109/ICHI54592.2022.00113.

[42] Marchal N., Xu R., Elasmar R., Gabriel I., Goldberg B. and Isaac W. (2024). Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data. GoogleDeepMind, 2024/06/05. https://arxiv.org/pdf/2406.13843

[43] Alter, S. (2024). Could a Large Language Model Contribute Significantly to Requirements Analysis?. In: van der Aa, H., Bork, D., Schmidt, R., Sturm, A. (eds) Enterprise, Business-Process and Information Systems Modeling. BPMDS EMMSAD 2024 2024. Lecture Notes in Business Information Processing, vol 511. Springer, Cham. https://doi.org/10.1007/978-3-031-61007-3_19

[44] Alter S. (2023). How Can You Verify that I Am Using AI? Complementary Frameworks for Describing and Evaluating AI-Based Digital Agents in their Usage Contexts. Proceedings of the 56th Hawaii International Conference on System Sciences | 2023. URI: https://hdl.handle.net/10125/103277

[45] Alter S. (2022). Understanding artificial intelligence in the context of usage: Contributions and smartness of algorithmic capabilities in work systems. International Journal of Information Management. Vol 67, December 2022 https://doi.org/10.1016/j.ijinfomgt.2021.102392

[46] Cabitza, F., Fregosi, C., Campagner, A., Natali, C. (2024). Explanations Considered Harmful: The Impact of Misleading Explanations on Accuracy in Hybrid Human-AI Decision Making. In: Longo, L., Lapuschkin, S., Seifert, C. (eds) Explainable Artificial Intelligence. xAI 2024. Communications in Computer and Information Science, vol 2156. Springer, Cham. https://doi.org/10.1007/978-3-031-63803-9_14

[47] Kieseberg, P., Weippl, E., Tjoa, A.M., Cabitza, F., Campagner, A., Holzinger, A. (2023). Controllable AI - An Alternative to Trustworthiness in Complex AI Systems?. In: Holzinger, A., Kieseberg, P., Cabitza, F., Campagner, A., Tjoa, A.M., Weippl, E. (eds) Machine Learning and Knowledge Extraction. CD-MAKE 2023. Lecture Notes in Computer Science, vol 14065. Springer, Cham. https://doi.org/10.1007/978-3-031-40837-3_1

[48] Cadamuro, J., Cabitza, F., Debeljak, Z., De Bruyne, S., Frans, G., Perez, S., Ozdemir, H., Tolios, A., Carobene, A. & Padoan, A. (2023). Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI). Clinical Chemistry and Laboratory Medicine (CCLM), 61(7), 1158-1166. https://doi.org/10.1515/cclm-2023-0355

[49] Banerjee, S., Agarwal, A. & Singla, S. (2024). LLMs Will Always Hallucinate, and We Need to Live With This. Preprint at arXiv https://doi.org/10.48550/arXiv.2409.05746