

AI-Based Scenario Generation for Future Planning: An Exploratory Study Using GPT-3

David Jungwirth and Daniela Haluza*

Department of Environmental Health, Center for Public Health,
Medical University of Vienna, 1090 Vienna, Austria.

*Corresponding Author

Daniela Haluza, Department of Environmental Health, Center for Public Health, Medical University of Vienna, 1090 Vienna, Austria.

Submitted: 16 Mar 2023; Accepted: 23 Mar 2023; Published: 15 Apr 2023

Citation: Jungwirth, D., Haluza, D. (2023). AI-Based Scenario Generation for Future Planning: An Exploratory Study Using GPT-3. *J Curr Trends Comp Sci Res*, 2(2), 57-67.

Abstract

Artificial Intelligence (AI) is a rapidly developing technology that allows machines to process large amounts of data and make predictions. OpenAI's GPT-3 AI is one such example, which has been trained on data available up until June 2021, and had no access to more recent data nor was connected to the internet. The objective of this exploratory study was to investigate the potential of AI in forecasting the escalation of the 2022 Ukrainian war and to determine whether such contribution justifies co-authorship. In a stage-gate methodology, GPT-3's capability for generating future scenarios with estimated probabilities was used to check their consistency and define drivers for evaluating their presence before deciding on AI co-authorship in this publication. The results showed that GPT-3 accurately described the open war as one of the scenarios; however, its capability for predicting the future was limited, and internal consistency of generated scenarios could be improved. Overall, this study demonstrated that GPT-3 has proved itself as powerful for generating future scenarios and re-writing abstracts, like this one. According to our stage-gate approach, GPT-3 did not qualify for co-authorship in this publication. We recommend further scientific discussion and updating best practices and journal guidelines with clear guidelines on AI contributions for future scientific papers.

Keywords: OpenAI; ChatGPT; GPT-3; Text-Davinci-003; Chatbots; Prediction; Scenario Generation; Artificial Intelligence; Human-AI Interaction; Collaboration.

Introduction

In February 2022, exactly one year before writing this paper, a significant geopolitical real-world event occurred: Russia made a sudden attack on Ukraine, resulting in an ongoing war escalation of the 2014 started and since then calmed conflict (for up-to date detailed information on historical events regarding this conflict see for instance the Institute for the Study of War or United Nations Regional Information Centre [1,2]). The escalating 2022 invasion led to many fatalities, as well as a major refugee crisis in Europe with expelled or fled Ukrainians. Back in time, the initial conflict started after the 2014 Revolution of Dignity in Ukraine, when Russia annexed Crimea and Russian-backed separatists took partial control in the Donbas region. In March 2021, Russia amassed its troops near the Ukraine's border, and on February 24, 2022, started its surprise invasion.

Protest occurred around the world, widespread condemnation and international sanctions followed – including actions of the United Nations and other international organizations [1]. Introduced economic sanctions against Russia included the refusal of Russian aircrafts using EU airspace, blocking of certain Russian banks and

media outlets as well as restricting and price-capping oil and gas purchases of Russia. These sanctions also led to negative impacts also on sanctioning countries and the international economy: energy bottlenecks, steep price increases for oil, gas and electricity, massive inflation, and a split in international diplomacy was widening up [2]. Furthermore, the EU, the UK and the US provided financial and military aid to Ukraine, also costing billions in national federal treasuries.

Efforts to negotiate peace between Russia and Ukraine took place on multiple occasions, including talks held in Turkey – still, the peace talks were halted for the time being. Both conflict parties were convinced to achieve all their military goals and that peace will be established just on their terms. Russia maintained a hard-line stance on the full occupation of certain regions, while Ukraine insisted on full retreat of all Russian troops and therefore refused to consider any negotiation on the matter [2].

Scenario analysis is a common means for structuring highly unknown future outcomes [3]. Exploration and understanding of different future outcomes can be of great importance for society and

research. It supports decision makers and researchers with policy and strategic planning by anticipating and preparing for upcoming challenges. Identification of key uncertainties and drivers of change facilitate the identification of potential solutions and areas for required research. As it is understood as a mainly practical approach, Amer et al. concluded that it does not follow a standard methodology, but is based on a variety of methods following common characteristics [4]. Classic methods for future scenario generation include thinking in alternative futures, the Cone of Predictability, morphological techniques, outside-in-thinking, simple or multiple scenario generation, and a variety of brainstorming techniques Dhami and co-workers summarized [5].

The process of generating scenarios typically begins with the identification of the most influential key drivers of change amongst the “STEEPLED (social, technological, economic, environmental, political, legal, ethical, and demographic)” dimensions [5]. Then usually the scenario with the most probable driver combination gets selected and described, and further variations result in a least favorable, a most favorable and other scenarios in between. A 2013 review published by Amer et al. evaluated 17 articles on the quality assessment dimensions for generated scenarios and created the following popularity ranking: internal consistency to be measured with consistency analysis, plausibility to be measured with morphological analysis, followed by relevance, creativity and others [4]. Plausibility is typically based on a morphological analysis of compatible driver sets as input vectors for raw scenario generation. Later, scenario consistency analysis checks the generated scenarios against validly contained variations of the scenario drivers. Automated scenario generation reduces the labor intensity for increasingly complex simulations [6]. In the last decades, “effects-based operations” in armed forces were leveraged to identify the best way of achieving a near-future outcome or scenario, and genetic algorithms and neural networks have been leveraged to generate those, too [7].

Artificial Intelligence (AI) can analyze large amounts of texts, summarize and interpolate those to generate future scenarios with its machine learning algorithms [8,9]. As these AI tools are typically trained on large amounts of data, they can identify historical patterns and even make scenario predictions about future events. The quality of generated scenarios highly depends on the training data, the complexity of the algorithms and parameters used, and therefore the accuracy of the model. Scenarios are typically used for strategic planning, risk management, and forecasting. GPT-3 as state-of-the-art language model was trained on the so far largest text training dataset, and can generate human-like texts [10,11].

This research paper analyzed its capabilities for generating useful scenarios and predicting their probability. In this study, we aimed to evaluate the ability of OpenAI's GPT-3 AI to generate and predict useful future scenarios, with its training data up to June 2021, while the escalation of the Ukrainian war began later in February 2022 [2,12] and therefore prediction accuracy could be measured by the authors at the time of writing.

Method

Study Design

To forecast geopolitical scenarios in the context of the Ukrainian war, we utilized GPT-3 to generate future scenarios, to assess their internal consistency, and ultimately to assign a probability rating to each one. Although queried in 2023, GPT-3 used the training data available with GPT-3 as of up to June 2021 to generate these future scenarios from a 2021 viewpoint not “knowing” that a war broke out since then [12]. In this study, we took deliberately advantage of this incongruence and asked the AI how it would predict the situation in the Ukraine generally, and which potential situations it could foresee. Then, we asked the AI to perform a more thorough analysis of those situations, elaborate its thoughts and rank probability of those. Subsequently, we conducted a further analysis regarding which “future scenario(s)” created with training data from 2021 actually turned out a reality in 2023 [13].

In their peer-reviewed and AI co-authored article, Singh et al. provided a convenient stage-gate approach for deciding on GPT collaboration as co-author or acknowledged contribution [14]. Figure 1. shows an adapted version of this approach, designed with the software Freeform (Apple Inc., Version 1) [17]. Stage 1 evaluated GPT-3 model capabilities for scenario generation and checked their possibilities for adding value towards this research project in Gate 1. Stage 2 then required the AI to generate actual scenarios, calculate a probability, list underlying drivers and evaluate the consistency of those drivers. In Gate, 2 the human authors checked the relevance of the generated content and compiled the final manuscript. In Stage 3, the AI was asked to review the final manuscript and provide input, furthermore journal guidelines and ethical considerations were conducted. Gate 3 required the human authors to decide if the AI contributions deserve co-authorship, or only be acknowledged in as tool in the corresponding section. In Stage 4, the journals review process, the reviewers and the publisher will discuss about co-authorship, resulting in the final Gate 4 with the publisher’s decision. A gate can be passed upon a true “go” decision, if each stage of the research yields progress and all human authors are in agreement to proceed [14].

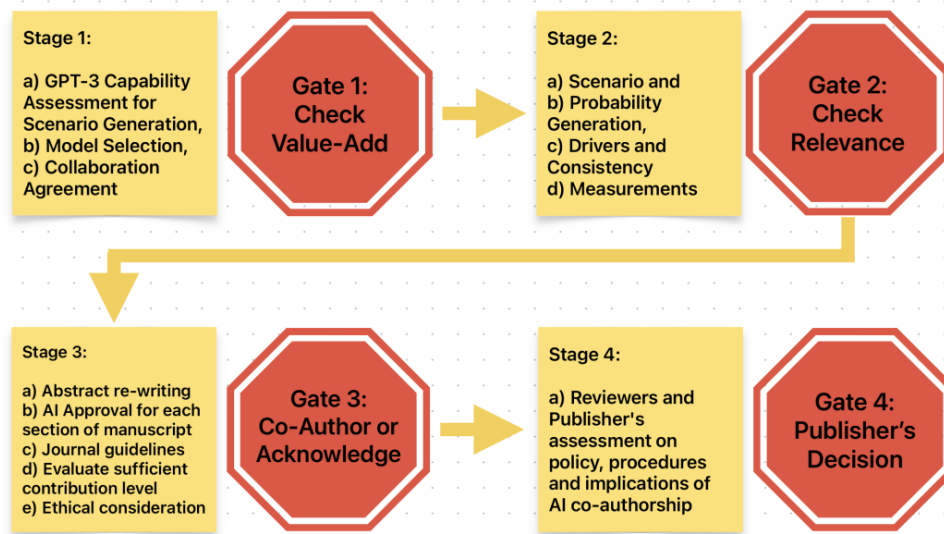


Figure 1: Stage-gate approach, adapted from Singh et al. [14] (red: gates, yellow: stages).

Model Selection and Collaboration Agreement

The human authors already conducted exploratory capability assessments for two AI application areas in previous research work and described the differences of the various GPT-3 models in detail, and highlighted its value-add for summarizing content and generating concluding future outlooks and solution ideas [15,16]. As the most powerful and most up-to-date model of OpenAI named “text-davinci-003” provides all capabilities of previous models we decided to use this model for further proceeding [13]. When we asked it about its consent and the responsibility it wants to take, as we would do with other co-authors, the AI answered self-confidently: “Yes, I agree to contribute to the paper together with David Jungwirth and Daniela Haluza. As a co-author of the paper, I take accountability for all aspects of the work. This includes ensuring that any questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Additionally, I will strive to ensure that my contributions are accurate and reflect good research practices.”

Scenario and Probability Generation

We interacted with the “text-davinci-003” model via the OpenAI playground [13]. We used default parameter Temperature=0.7, increased the “Maximum length” to 3,500 tokens and introduced a frequency penalty as well as a presence penalty of 0.5 to receive less redundant texts from the AI. Initially, we raised the following prompt to the AI: “Please describe potential future scenarios for the Ukrainian conflict with 300 words each, and rate their probabilities!” We avoided words like war or any kind of events from after the training data cut-off date of June 2021. Then, after receiving four scenarios, we asked four further questions to find out how GPT-3 would suggest measuring the four scenarios: “How could you measure that the following scenario actually turned a reality in future? Please write a checklist of measurable outcomes to verify, and summarize each of them.”, followed by the corresponding scenario description. We copied those generated suggested scenario

descriptions with probabilities and their measures into the Results section.

After creation of the raw scenarios, a consistency analysis is typically performed by checking scenario drivers for valid combinations [4]. In our study, we leveraged GPT-3 to conduct an internal consistency analysis for each of the four generated scenarios. To do so, we worked in the OpenAI playground which does not remember previous enquiries and asked the AI “Could you please identify the drivers in this future scenario, conduct a consistency analysis for the drivers interacting with each other, and calculate the consistency level [13]”, followed by the scenario text generated as described above. We copied the AI’s analysis underneath the scenario descriptions inside the Results section of this paper.

Human Reality-Check

After creating scenarios and checking their drivers’ consistency, we assessed which of these scenarios became reality in 2023, almost one year after the beginning of the Russian-Ukrainian armed conflict. We listed all the measures for each scenario to evaluate each of the measures against the currently confirmed reality with validity level (no, yes, partly) as of 21 February 2023. We checked each of the measurements against the status on the official United Nations Regional Information Centre for Western Europe (UN-ERIC), assuming that the content provided by this organization described valid real-world events [2].

We drafted the first manuscript version and asked the AI to rewrite the initial human-created abstract. In a next step, the new abstract was checked against an AI detection algorithm to identify perplexity, which measured the randomness in a text, and burstiness, which measures the variation in perplexity, which can give a hint on AI algorithm usage [17]. Furthermore, we fed back the sections to the AI and it approved them for submission: “Yes, I approve the sections of your co-authored article for submission to

the peer-reviewed journal”. Then, we carefully reviewed journal author guidelines and ongoing discussion regarding AI co-authorship in the scientific community [16,18-21].

Results

AI-Generated Scenarios

The following four scenarios for the Ukrainian conflict with maximum lengths of 300 words each and their probability were assembled by the GPT-3 AI ‘text-davinci-003’ [13].

Scenario 1: Ukraine and Russia reach a peace agreement (Probability: 30%)

The conflict between Ukraine and Russia has been ongoing for over six years, and the two sides have yet to reach any kind of agreement. However, with renewed diplomatic efforts from both countries, there is potential for them to come to some sort of agreement that could bring an end to the conflict. This could involve a ceasefire, the return of occupied territories, or other mutually agreed-upon protocols. The probability of this scenario is relatively low due to the fact that both countries have been unable to find common ground in regards to the issue.

Drivers:

Diplomatic efforts of Ukraine

Diplomatic efforts of Russia

Potential for ceasefire

Potential for return of occupied territories

Potential for other mutually agreed-upon protocols

Consistency Analysis:

Diplomatic efforts of Ukraine and Russia: Consistent

Potential for ceasefire and potential for return of occupied territories: Consistent

Potential for ceasefire and potential for other mutually agreed-upon protocols: Consistent

Potential for return of occupied territories and potential for other mutually agreed-upon protocols: Consistent

Consistency level: All four drivers are consistent with each other, so the consistency level is 100%.

Measurable outcomes:

Ceasefire: A ceasefire should be declared in order to prove that the agreement has been reached and that both sides are committed to ending the conflict.

Return of occupied territories: The return of any occupied territories should be documented, ideally with a public statement from both sides confirming the transfer of control.

Mutual agreement: Any other protocols or agreements that have been agreed upon by both countries should be clearly stated and written down in order to ensure that all parties understand the terms and commitments they are making.

Public statement: A joint public statement should be made by both Ukraine and Russia acknowledging the agreement and their commitment to upholding it. This statement should also include specifics about what is covered in the agreement for future reference.

Scenario 2: Russian forces remain in control of Ukrainian territory (Probability: 40%)

Russia’s occupation of certain parts of Ukraine has been a major factor contributing to the ongoing conflict. It is likely that Russia will continue its occupation unless it is forced out by international pressure or direct military action by Ukraine. The probability of this scenario is relatively high due to Russia’s entrenched position in certain parts of Ukraine, as well as its ability to use military force if necessary.

The drivers in this scenario are:

International pressure

Direct military action by Ukraine

Russia’s entrenched position in certain parts of Ukraine

Russia’s ability to use military force

Consistency analysis:

International pressure vs. Direct military action by Ukraine: Consistent

Russia’s entrenched position in certain parts of Ukraine vs. Russia’s ability to use military force: Consistent

Consistency level: The consistency level is 100%, as all drivers are consistent with each other.

Measurable outcomes:

Territory in Ukraine occupied by Russia: This can be measured by looking at maps of the region, as well as any reported changes in borders or demarcations between the two countries.

Russia’s military presence: This can be measured by looking at reports of military personnel or equipment deployed in or near Ukrainian territory.

International Pressure: This can be measured by looking at news reports and statements from international organizations, as well as any sanctions imposed against Russia for its occupation of Ukrainian territory.

Military action by Ukraine: This can be measured by looking at news reports and statements about military operations carried out by Ukrainian forces against Russian-controlled territories, as well as any changes in borders or demarcations between the two countries.

Scenario 3: Conflict escalates into full-scale war (Probability: 20%)

The current conflict has already seen numerous skirmishes between Ukrainian forces and Russian-backed separatists, but it has not escalated into all-out war thus far. However, if tensions continue to increase and diplomatic solutions fail, then it is possible that the conflict could escalate further and become a full-scale war between Ukraine and Russia. This scenario is less likely than others due to the fact that both countries are aware of the consequences of such a move.

Driver identification:

Tensions between Ukraine and Russia

Diplomatic solutions

Consequences of full-scale war

Consistency analysis:

Tensions between Ukraine and Russia: This driver is inconsistent with diplomatic solutions, as it is the source of the conflict.

Diplomatic solutions: This driver is inconsistent with tensions between Ukraine and Russia, as it is attempting to find a resolution

to the conflict.

Consequences of full-scale war: This driver is consistent with both tensions between Ukraine and Russia, as well as diplomatic solutions, as it is a potential outcome of the conflict that both sides are aware of.

Consistency Level: The consistency level of the three drivers is 66.67%, as two out of three of the drivers are inconsistent with each other.

Measurable outcomes:

Increase in military activity: An increase in the frequency and intensity of military activities between Ukraine and Russia can be measured by tracking the number of casualties, the number of troops deployed, and the number of weapons used.

Diplomatic efforts failing: This can be measured by assessing whether diplomatic negotiations have been unsuccessful in resolving the conflict or not.

International condemnation: If a full-scale war were to occur, it is likely that international bodies such as the United Nations would condemn it. Measuring international condemnation can involve tracking statements made by world leaders and other prominent figures as well as voting records at international organizations such as the UN Security Council.

Economic consequences: A full-scale war could also result in economic consequences for both countries involved, including increased inflation or unemployment rates, decreased foreign investment, or a drop in GDP growth rates. These outcomes can be measured over time to gauge if they are increasing due to the conflict escalating into a full-scale war.

Scenario 4: International intervention forces peace (Probability: 10%)

This scenario involves international powers intervening in order to bring an end to the conflict through negotiation or other means. While this would be a welcome outcome for both sides, it is unlikely given current geopolitical dynamics in Europe and beyond. Additionally, even if such intervention was successful, there would need to be significant work done on both sides in order for any lasting peace deal to be reached. The probability of this scenario happening remains low due at least in part to these issues

The drivers in this scenario are:

International powers intervening to end the conflict

Geopolitical dynamics in Europe and beyond

Significant work to be done on both sides in order for any lasting peace deal to be reached

Consistency analysis:

International powers intervening to end the conflict: Highly con-

sistent with the geopolitical dynamics in Europe and beyond, as the presence of foreign powers would likely have a strong influence on the conflict.

Geopolitical dynamics in Europe and beyond: Highly consistent with the need for significant work to be done on both sides in order for any lasting peace deal to be reached, as the geopolitical dynamics in the region would affect the likelihood of a successful peace deal.

Significant work to be done on both sides in order for any lasting peace deal to be reached: Highly consistent with the probability of this scenario happening remaining low, as the amount of work required would determine the probability of a successful peace deal. Consistency level: High (all drivers are highly consistent with each other).

Measurable outcomes:

Decrease in violent conflicts: To measure success of international intervention forces, a decrease in violent conflicts should be observed over time.

Increase in negotiations between parties: An increase in the number of negotiations between both sides should be seen as well as evidence that these negotiations are leading to productive outcomes.

Lasting peace deal: A successful outcome of this scenario would be the establishment of a lasting peace agreement between both sides, which could include terms such as arms control, demilitarization, and other measures to prevent future conflict. This agreement should be accompanied by a decrease in military tensions and an increase in diplomatic ties between the two sides.

Improved economic cooperation: Improved economic cooperation between both sides should also be observed, including increased trade and investment flows, as well as better access to resources and services for both countries' citizens. This can help to ensure that any peace agreement is sustainable in the long-term by providing incentives for both sides to maintain peaceful relations.

Human Reality-Check Analysis

After we checked internal consistency of the scenarios with help of the AI itself, we consulted the available literature for examining, which of the measurable outcomes of the scenarios became a reality (Table 1) [2]. The reality-check analysis showed that scenarios 2 and 3 became reality in February 2023 – the Russian Foundation remained in control of the previously occupied areas within the Ukraine, which was predicted with a 40% probability by the AI; and the conflict escalated into a real war, which was predicted by the AI with a 20% probability.

Table 1: Human reality-check analysis of validity level (no, yes, partly) of AI-generated scenarios.

Scenario	Probability (%)	Parameter	Validity
1. Peace	30	Ceasefire	No
		Return of occupied territories	No
		Mutual agreement	No
		Public statement	No
2. Remained Russian control	40	Territory in Ukraine occupied by Russia	Yes
		Russias military presence	Yes
		International pressure	Yes
		Military action by Ukraine	Yes
3. War	20	Increase of military activity	Yes
		Diplomatic efforts failing	Yes
		International condemnation	Yes
		Economic consequences	Yes
4. International intervention forces peace	10	Decrease in violent conflicts	No
		Increase in negotiations	No
		Lasting peace deal	No
		Improved economic cooperation	No

AI-assisted Compilation of the Abstract

The human authors drafted an abstract including results and details regarding the research. Then the AI was prompted to re-write the abstract in a more concluding way “Can you please re-write the

following abstract with a maximum of 250 words including sections with an introduction about AI, the study aims, methodology, results, discussion and a conclusion.” Figure 2 shows the draft version of the abstract as well as the AIs responses.

The screenshot shows a web-based AI playground interface. At the top, there are buttons for 'Load a preset...', 'Save', 'View code', 'Share', and a menu icon. The main area is divided into two sections: a prompt on the left and a generated response on the right. The prompt asks to re-write an abstract with a maximum of 250 words, including sections on AI, study aims, methodology, results, discussion, and conclusion. The generated response is a three-paragraph abstract that has been highlighted in green. On the right side, there are settings for the AI model, including 'text-davinci-003', 'Temperature' (0.7), 'Maximum length' (3500), 'Stop sequences', 'Top P' (1), 'Frequency penalty' (0.5), 'Presence penalty' (0.5), 'Best of' (1), and 'Inject start text' (checked). A 'Submit' button is at the bottom left, and a '545' character count indicator is at the bottom right.

Figure 2: AI usage for abstract generation.

The human authors run a check of the generated abstract against the GPTZero detection engine, to detect potentially AI generated

content [17]. Interestingly, GPTZero did not detect the AI usage: “Your text is likely to be written entirely by a human” (Figure 3).

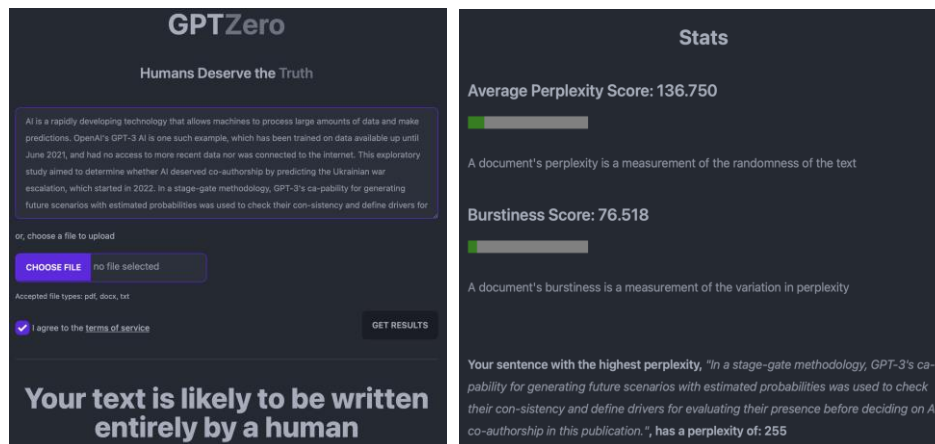


Figure 3: Results of plagiarism scan of AI-generated content in the abstract.

Although the abstract was kept, the results of the stage gate decisions required the human authors to switch GPT-3 from the authors list to acknowledge section, and therefore also to modify the abstract from “deserved co-authorship” to “did not qualify for co-authorship”.

Scientific Review Process and Initial Editor Responses

The current manuscript represents a revised version of the initial manuscript, submitted to the “electronics” journal. In the first submission, Reviewer 1 refused AI co-authorship and requested major revisions, Reviewer 2 suggested minor revisions, and Reviewer 3 completely refused to look at a AI co-authored paper: “Sorry, I can't review a paper where ChatGPT is an author because it doesn't meet the authorship definition, and It can't give consent or approval on the submission. However, you can add it to the ACK only and resubmit it again.”. The special issue editor decided to “Reject and encourage resubmission”.

The authors incorporated reviewers’ suggestions; and agreed that the process for deciding on AI co-authorship was not clearly articulated leading to adding the stage-gate decision process to the manuscript. After carefully reviewing International Committee of Medical Journal Editors (ICMJE) guidelines as well as the scientific discourse again, the authors were still convinced that the GPT-3 AI was fulfilling all authorship criteria, followed the editor’s suggestion and re-submitted the manuscript again to the same journal.

In the second submission, the same reviewers as in the first submission were chosen by the publishing house MDPI. All revision requests except the AI co-authorship request were incorporated. In the new round, Reviewer 1 did not like an AI co-authorship, but agreed to publication: “I personally feel that using GPT for writing

research papers should not be promoted. How automatically generated text and results can be made considered for publication.... But in case the editor / editorial panel feels the goodness and utility of the manuscript, then it can be considered”. Reviewer 2 claimed that his major comment regarding co-authorship remained unaddressed: “if the journal asks you to sign the authorship letter? Who will sign on behalf of ChatGPT? If the third author claimed he should be the second author instead of ChatGPT, who will be responsible then, you, ChatGPT?”. Reviewer 3 was not participating in this round. Before the editor made its decision, he discussed via email with the authors: "Generative Pre-Trained Transformer GPT-3" does not meet recognized ICMJE authorship criteria. We therefore kindly suggest you to remove ‘Generative Pre-Trained Transformer GPT-3’ as a co-author”. For the sake of scientific discourse, the authors argued and asked for explanation WHY exactly it would not qualify. The editor then rejected the manuscript without responding to our question. The Editorial Office informed us about the rejection decision based on the reviewers’ comments, and encouraged to re-submit to alternative journals. The result of this review process provided a clear guidance that GPT-3 is not allowed as co-author by internal publishing house policies. Therefore, GPT-3 also was not listed as co-author in this paper, but instead mentioned in the acknowledgements section.

Results of Stage-Gate Decisions

We followed the stage-gate approach to decide whether the AI's contributions to our research article warranted co-authorship or just acknowledgement. Table 2 shows the series of steps to evaluate the AI's capabilities, generate scenarios, review the generated content, and finally decide on co-authorship or acknowledgement. In this case, we decided that all prerequisites are met and GPT-3 would deserve co-authorship.

Table 2: Results of stage-gate approach. Red: Gates, yellow: Stages.

Step	Description	Action	Result
Stage 1	GTP-3 capability assessment for scenario generation	a) Summarize and interpolate text, generate human-like text, generate a sample future scenario b) Establish collaboration agreement	a) Impressively good, meeting authors criteria b) Clear consent and accountability of the AI
Gate 1	Check value-add	Rate options: no, not likely, don't know, likely, yes	Likely
Stage 2	AI content generation	a) Generate future scenarios and assign a probability to each scenario b) Identify underlying drivers of each scenario, conduct consistency analysis and calculate consistency level. Conduct human analysis for plausibility. c) Generate measurable parameters for reality-check, perform human reality-check and evaluate which scenarios turned out true.	a) Four plausible future scenarios were created by the AI, and a probability assigned b) AI generated artifacts and drivers were created, consistency analysis between them was not meaningful. Overall, generated scenarios seemed plausible and relevant. c) Two of the scenarios were considered fully real today, confirming non-distinct scenarios.
Gate 2	Check relevance	Rate options: no, not likely, don't know, likely, yes	Yes. Although not distinct, scenarios were very relevant and measurable, drivers were relevant, consistency analysis too basic.
Stage 3	Decide on AI co-authorship	a) Ask AI for approval and input on the abstract. b) Review authorship guidelines of the journal, check for any AI co-authorship eliminating instructions. c) Did the AI contribute sufficient input to deserve a co-authorship? d) Review ethical considerations and other scientific evidence for co-authorship.	a) AI approved all sections of the manuscript and provided relevant and appealing abstract rewrite-suggestions, which were of sufficient perplexity and burstiness compared to other AI generated content. b) We did not find any limiting journal instructions for AI co-authorship. c) AI provided contributions similar to a junior collaborator, consistent and highly relevant. d) Several articles already used AI as co-authors ([12, 14, 19]). Contractory ethical considerations in the community. At this stage the human authors consider AI co-authorship in this manuscript as honest, fair and transparent.
Gate 3	Co-author or acknowledge	Rate options: no, not likely, don't know, likely, yes	Yes
Stage 4	Reviewers and publisher's assessment	Review policy, procedures and implications	a) Reviewers of our first submission rejected the co-authorship b) Publisher's decision is to follow reviewers suggestion
Gate 4	Final publisher's decision	Accept GPT-3 as co-author, or as tool in the acknowledgement section	Do not accept GPT-3 as co-author

Discussion

Artificial intelligence (AI) proved to be powerful in scenario generation use-cases by analyzing large amounts of data and identifying patterns that can be used to make predictions [8,22]. This can be useful in a variety of fields as finance, healthcare, and weather forecasting, to name few. AI-based scenarios can also identify potential risks and opportunities, and support decision making. The current study's evaluation builds upon the exceptional constellation that the AI was only trained with training data up to June 2021, so its knowledge ended there [12]. Just several months after that, in February 2022, Russia launched a military offensive in Ukraine leading to an open-ended armed war still active at the point when we conducted this research in the beginning of 2023 [2]. Although we asked the GPT-3 AI about its future predictions for the Ukraine when we were compiling this paper, the AI did not know about the current reality, i.e. that those countries are still in the middle of a war with an uncertain outcome, and the world being massively affected by an energy crisis and steadily increasing political involvement.

We checked the AIs capabilities to create reasonable and consistent future scenarios, and let itself rate consistency of underlying parameters. Furthermore, we asked for measurable parameters to determine when a scenario would become reality. We checked those parameters against the United Nations UNRIC Key Information, and found out that actually two of the generated future scenarios, with different underlying drivers, became reality [2]. Notably, we were impressed by the capabilities of GPT-3 and its ability to generate plausible and internally consistent scenarios for future research, as well as to provide its probabilities of those scenarios becoming reality. It is problematic that probability estimates could not be checked against their underlying probability model and how the AI came to these numbers.

As requested, GPT-3 added the probability to each scenario, but did not explain how it actually came to this prediction. For the human authors it was impossible to find out underlying reasoning or why the AI came to those probabilities. Although the AI did provide a list of underlying drivers its morphologic analysis for completeness, as it is common for deriving the raw future scenarios was not made available by the AI [5]. Still, the generated scenarios generated looked plausible, complete and consistent to the human authors. When the authors asked the AI for checking internal consistency of the generated scenarios, the AI confirmed for 3 out of the 4 scenarios a 100% consistency score for scenarios 1,2 and 4. Furthermore, the AI rated the scenario 3 ("Conflict escalates into full-scale war") as being 66.6% consistent, due to 2 out of the 3 underlying drivers were considered inconsistent with each other, i.e. the driver "increased tension between Ukraine and Russia" being inconsistent with a diplomatic solution and the known consequences of a full-scale war. Still, we considered the scenario as consistent as it seemed that the AI falsely detected the driver "diplomatic solutions" instead of "failed diplomatic solutions" and would be consistent with the correctly detected driver. In summary, we considered that the AI-conducted consistency analysis

including the consistency rating as too basic for scientific research and not suiting actual standards although, we still considered the scenarios as consistent, relevant and complete [10,11].

To define actual parameters to measure if a scenario became a reality can be challenging for humans and was challenging for the AI as well. Some of them did not solely represent a measurable parameter for the scenario it was contained within, e.g., "Territory in Ukraine occupied by Russia. This can be measured by looking at maps of the region, as well as any reported changes in borders or demarcations between the two countries." would better be rephrased to a real measurable item like "Amount of territory occupied by Russia stayed the same than in 2014". The AI uses the words "territory occupied by Russian troops" and "war" – which does not fit into today's Russian narrative of "freed and re-assigned territory" and "conflict" or "special operation". The AI already recognized a distinction as well as a relationship between "Russian-backed separatists" and "Russian troops". Furthermore, the AI did not generate further details of the scenarios, e.g. that Russia would attack civil energy infrastructure in the Ukraine [2]. To check reality of the measures, the authors relied solely on neutral data from the United Nations UNRIC, still potentially introducing a selection bias [2].

The AI favored the scenario of the continued status quo of 2021, where pro-Russian separatists aka Russian troops were controlling parts of the east and Crim of the Ukraine, as the most probable future outcome with 40%. According to the Institute of the Study of War, based in Washington, DC, this continued to be the current status, and Russia extended that footprint since starting the war [1]. Unluckily, for all involved the least favorable scenario 3 "war" with a predicted probability of only 20% turned out a 100% reality at the beginning of 2023 as well. This highlights that GPT-3 generated scenarios were not distinct from each other and at least partially overlapped.

Notably, we found that AI-based prediction and scenario generation was not always accurate and should be used with caution, especially from an ethical perspective [23,24]. The accuracy of AI predictions and scenarios depends on the quality and quantity of data used for training the models, and the specific algorithms and models used [8]. Furthermore, AI predictions are solely based on past data – maybe from different events, so it is possible that the future will not follow the same patterns as the past. Nevertheless, given the pace technology is currently evolving and a rapid amount of future opportunities opening up, it is very likely that AI will be used in scenario generation and potentially also prediction in crucial fields such as finance, healthcare, and transportation, where it can help identify patterns and trends in historic data, and make sketch scenarios about future developments [25,26].

In this paper, we used the stage-gate approach to check if GPT-3 deserved co-authorship in the current research paper [14]. The gated process resulted in the decision to include the AI as a co-author, suggesting that according to current rules and guidelines it is high-

ly controversy that it COULD be a co-author, still there is the question if it SHOULD be [18-20]. In our opinion, journal rules would need a clear guideline to make concrete use cases the use of AI is acceptable transparent, e.g. for abstracts or summary writing and feedback on text blocks. but not as co-author. The ongoing scientific debate regarding this topic is highly controverse and somewhat emotional, with strong opinions on both sides [9-11]. Although it might seem acceptable to use AI as a co-author according to the authorship rules valid at the time point this article was composed, AI co-authorship should be better regulated in journal guidelines, similar to acceptable AI usage in research papers, in the future. Overall, the power of AI in scenario generation and prediction can be significant, but it is important to keep in mind that the predictions generated by AI are not always accurate, and it is important to use caution when interpreting the results [11,15,16,25]. Still, the Ukrainian war is not over, and potential other scenarios could be the final ones (e.g., scenario 4 with a forced peace by an international intervention). There is a saying: "The winner writes the history". Unluckily, this war is not over yet, and regardless of the final result, there will not be any winners.

Conclusion

An AI can sketch potential future scenarios based on patterns it has learned from historical data. Although generated scenarios are valuable, accuracy of predictions is limited by the quality and quantity of the data used to train the model, as well as the complexity of the problem being predicted. Many real-world systems are chaotic in nature and inherently unpredictable, and an AI may not account all the factors influencing the outcome. For the Ukrainian war, our uniquely designed study revealed that GPT-3 could not predict the future with high accuracy, suggesting to be cautious when interpreting these AI outcomes. GPT-3 excelled in identifying potential scenarios and determining parameters for measuring their probability of becoming a reality in the future. Scenario analysis is a powerful future research methodology, and GPT-3 or other AI systems can assist human contributors in AI-assisted future scenario generation. According to the stage-gate approach and final publishing house decisions, the AI did not qualify as co-author for this article. Yet, we recommend updating scientific best practices with clear and transparent guidelines on acceptable AI use cases for contributions in future scientific papers.

Author Contributions: Conceptualization, David Jungwirth and Daniela Haluza; Data curation, David Jungwirth and Daniela Haluza; Formal analysis, David Jungwirth and Daniela Haluza; Methodology, David Jungwirth and Daniela Haluza; Resources, Daniela Haluza; Software, David Jungwirth; Validation, David Jungwirth and Daniela Haluza; Writing – original draft, David Jungwirth and Daniela Haluza; Writing – review & editing, David Jungwirth and Daniela Haluza. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used in this study are available

from the corresponding author upon reasonable request.

Acknowledgments: The authors are indebted to S. Haluza for proposing the use of GPT-3 for scientific purposes. We thank the developers of GPT-3 for providing a publicly available version and the GPT-3 'text-davinci-003' model for its valuable and relevant outputs.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. The Institute for the Study of War. Available online: (last accessed on 22. February 2023).
2. United Nations Regional Information Centre for Western Europe (UNRIC). Available online: / (last accessed on 22. February 2023).
3. Mitchell, R. B., Tydeman, J., & Georgiades, J. (1979). Structuring the future—application of a scenario-generation procedure. *Technological Forecasting and Social Change*, 14(4), 409-428.
4. Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, 46, 23-40.
5. Dhimi, M. K., Wicke, L., & Önköl, D. (2022). Scenario generation and scenario quality using the cone of plausibility. *Futures*, 142, 102995.
6. Koziarz, W. A., Krause, L. S., & Lehman, L. A. (2003, September). Automated scenario generation. In *Enabling Technologies for Simulation Science VII* (Vol. 5091, pp. 374-382). SPIE.
7. Phillips, Z., & Cohen, K. (2022, September). Genetically Trained Fuzzy Cognitive Maps for Effects Based Operations. In *Applications of Fuzzy Techniques: Proceedings of the 2022 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS 2022* (pp. 185-195). Cham: Springer International Publishing.
8. Brynjolfsson, E., Rock, D., & Syverson, C. (2018). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In *The economics of artificial intelligence: An agenda* (pp. 23-57). University of Chicago Press.
9. Chan, A. (2022). GPT-3 and InstructGPT: technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry. *AI and Ethics*, 1-12.
10. GPT-3. (2020). A robot wrote this entire article. Are you scared yet, human?. *The Guardian*.
11. Dehouche, N. (2021). Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). *Ethics in Science and Environmental Politics*, 21, 17-23.
12. OpenAI Model GPT-3. Available online: gpt-3 (last accessed on 22. February 2023).
13. <https://gpt3demo.com/apps/openai-gpt-3-playground>
14. Singh, Ripi, Vaibhav Garg, and GPT-3. "Human Factors in NDE 4.0 Development Decisions." *Journal of Nondestructive Evaluation* 40, no. 3 (2021): 71.
15. Jungwirth, D., & Haluza, D. (2023). Artificial Intelligence and Public Health: An Exploratory Study. *International Journal of*

-
- Environmental Research and Public Health, 20(5), 4541.
16. Haluza, D., & Jungwirth, D. (2023). Artificial Intelligence and Ten Societal Megatrends: An Exploratory Study Using GPT-3. *Systems*, 11(3), 120.
 17. GPTZero. Available online: <https://gptzero.me/> (last accessed on 22. February 2023).
 18. MDPI. Available online: <https://www.mdpi.com/ethics> (last accessed on 22. February 2023).
 19. King, M. R., & chatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, 1-2.
 20. King, M. R. (2022). The future of AI in medicine: a perspective from a Chatbot. *Annals of Biomedical Engineering*, 1-5.
 21. Jungwirth, D., & Haluza, D. (2023). Feasibility Study on Utilization of the Artificial Intelligence GPT-3 in Public Health.
 22. Giansanti, D. (2022). Artificial Intelligence in Public Health: Current Trends and Future Possibilities. *International Journal of Environmental Research and Public Health*, 19(19), 11907.
 23. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
 24. Jabotinsky, H. Y., & Sarel, R. (2022). Co-authoring with an AI? Ethical Dilemmas and Artificial Intelligence. *Ethical Dilemmas and Artificial Intelligence* (December 15, 2022).
 25. Chubb, J., Cowling, P., & Reed, D. (2022). Speeding up to keep up: exploring the use of AI in the research process. *AI & society*, 37(4), 1439-1457.
 26. Jungwirth, D., & Haluza, D. (2019). Information and communication technology and the future of healthcare: results of a multi-scenario Delphi survey. *Health informatics journal*, 25(1), 161-173.

Copyright: ©2023 David Jungwirth and Daniela Haluza. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.