Analyzing the Innovative Potential of Texts Generated by Large Language Models: An Empirical Evaluation

Oliver Krauss[1,2](✉), Michaela Jungwirth[1,2], Marius Elflein[1,2], Simone Sandler[1,2], Christian Altenhofer[1,2], and Andreas Stoeckl[1,3]

[1] University of Applied Sciences Upper Austria, Campus Hagenberg, Softwarepark 11, 4232 Hagenberg, Austria
{oliver.krauss,simone.sandler,andreas.stoeckl}@fh-hagenberg.at
[2] Advanced Information Systems and Technology, Hagenberg, Austria
[3] Digital Media Department, Hagenberg, Austria
http://aist.science

Abstract. As large language models (LLMs) revolutionize natural language processing tasks, it remains uncertain whether the text they generate can be perceived as innovative by human readers. This question holds significant implications for innovation management, where the generation of novel ideas from extensive text corpora is crucial. In this study, we conduct an empirical evaluation of 2170 generated idea texts, containing product and service ideas in current trends for specific companies, focusing on three key metrics: innovativeness, context, and text quality. Our findings show that, while not universally applicable, a substantial number of LLM-generated ideas exhibit a degree of innovativeness. Remarkably, only 97 texts within the entire corpus were identified as highly innovative. Moving forward, an automated evaluation and filtering system to assess innovativeness could greatly support innovation management by facilitating the pre-selection of generated ideas.

Keywords: Artificial Intelligence · Decision Support · Large Language Models · Data Quality

1 Introduction

We examine the perception of innovativeness in texts generated by Large Language Models (LLMs) among a human audience. To address this question, we perform an empirical evaluation on a comprehensive set of 2170 texts generated specifically for innovation management. Our evaluation focuses on assessing the degree of *innovativeness*, the *context*ual relevance of the innovations proposed, and the extent to which the generated *text*s resemble human-written content.

It is widely recognized that Large Language Models (LLMs) lack thought or reasoning capabilities and can be likened to "Stochastic Parrots" [2]. These models rely on predicting the next words in a sequence based on vast corpora of data, without true comprehension. However, recent advancements have integrated LLMs with online source access, enabling them to retrieve information [9].

G. Kotsis et al. (Eds.): DEXA 2023 Workshops, CCIS 1872, pp. 11–22, 2023.
https://doi.org/10.1007/978-3-031-39689-2_2

The integration of next word prediction with real-time data access has shown promising results in simulating novelty within various domains. This approach has been particularly valuable in innovation management, facilitating data analysis and idea generation [3]. Furthermore, it has proven effective in expanding, combining, and explaining novel ideas through creative processes [8].

Previous research has delved into the novelty of texts generated by LLMs, with a specific focus on programming exercise generation [16]. Although an empirical evaluation was not conducted, the study aimed to determine if the generated texts could be found on the internet. Remarkably, the findings revealed that 81.8% of all generated examples were deemed novel. While novelty does not directly imply innovativeness, these results indicate the ability of LLMs to produce new information. In smaller models such as GPT-2, instances of text duplication may occur, diminishing novelty and innovation [13].

Our objective is to determine the true innovativeness of texts generated by LLMs. Our evaluation encompassed a diverse range of domains, including new ideas, products, improvements, and novel customer service areas for a total of 31 companies. We aim to answer the following questions:

- Perception of Innovativeness: We explore whether ideas generated by LLMs are perceived as truly *innovative* by human evaluators. This investigation delves into the subjective assessment of the generated texts in terms of their novelty and creativity.
- Contextual Relevance: We examine the degree to which the generated ideas align with the *context* of the specific company they were generated for. This analysis assesses the appropriateness of the ideas in relation to the company's industry, objectives, and current trends.
- Human-Like Text Quality: We evaluate the *text* quality of LLM-generated passages and investigate whether they can convincingly pass as human-written content. This assessment takes into account factors such as coherence, grammar, and overall fluency.

2 Background

LLMs are large models trained on general data to solve a wide range of Natural Language Processing (NLP) tasks. In order to perform the desired task, a prompt is given to the LLM [14]. Examples for LLMs are GPT3 [4], which was used to generate the text corpus, BLOOM [17] and PaLM [5]. LLMs are able to achieve state-of-the-art results on various NLP tasks, including text generation [15].

When evaluating the innovativeness of texts generated by LLMs, it is crucial to recognize the specific task to which the text corpus was applied: ideation in the domain of innovation management. After the trend scouting phase, ideation is used to generate and subsequently filter ideas based on their potential. This stage serves as a foundation for further evaluation and analysis, allowing more promising ideas to be explored while discarding less viable ones [6].

The primary concept underlying the text corpus employed in our research was to leverage Language Models (LLMs) for analyzing extensive collections of data and generating fresh ideas. This approach aimed to provide innovation managers with one concise paragraph of text per idea that serve as valuable starting points. Manual sifting through such corpora would pose a significant challenge for humans; however, the utilization of LLMs holds promising potential for enhancing innovation management [7].

In our case the trends were mined from news articles, scientific publications from conferences and journals as well as arXiv documents. The trends were mined via topic modelling [10], of the title and abstract of each article.

We employed GPT-3, to create an idea that may be usable by a business to expand on their current portfolio, or to create new products or services. Our aim was to harness the capabilities of GPT-3 to generate ideas that businesses could potentially utilize to expand their existing portfolio or develop new products and services. The paramount goal of these generated ideas was to appear genuinely *innovative* to the readers. Additionally, the business provided us with a specific *context* to determine the innovation potential of the ideas. Furthermore, the LLM was equipped with knowledge of a specific trend, backed by relevant articles, to enhance its understanding of the current market landscape.

3 Methods

We present the experiment design, which involved conducting an empirical analysis of 2170 texts generated by GPT-3. Each of these texts corresponded to a unique idea generated based on a current industry trend for a specific company. To assess the quality and effectiveness of these ideas, we enlisted a group of 11 reviewers who evaluated text across 3 key metrics:*innovative*, *context*, and *text*. Each text underwent evaluation by 3 out of 11 reviewers to ensure a comprehensive and robust assessment process.

We selected 31 Austrian companies representing various industries, such as *IT*, *Law* or *Energy*. For each of these companies, we identified 10 current industry trends for which ideas were to be generated. Subsequently, we generated 7 distinct ideas for each of the identified trends. This results in a total of 2170 texts (31 companies x 10 trends x 7 ideas). This comprehensive approach ensured a diverse and extensive text corpus for our analysis.

For each of the 31 companies, a profile was created which has the following information, which is also presented in Table 1:

name of the company. The actual names of the companies were replaced with single words that broadly represent their core industry, ensuring anonymity throughout the work.
website of the company. Provided to the reviewers and not GPT-3, to assess if the generated text is fitting the context of the company.
keywords consisting of 5 keywords (e.g. "food") or keyword phrases (e.g. "food science"), representing the core industrial areas or products.
description summarizing a company was provided to the reviewers and GPT-3.

Table 1. Example profile of the company *Food* from which reviewers can assess the context of the idea.

Data	Example
name	*Food*
website	https://www.example-food.com/
keywords	food science, food processing, fruit, starch, sugar
description	*Food* is an Austrian company which produces a wide range of industrial products for the processing sector. In its business segments, Fruit, Starch and Sugar, *Food* supplies local producers and large international players, particularly those in the food processing industry

For each of the trends, a description (see Table 2) was formulated and was made available to GPT-3. Similarly, the reviewers had access to this description to assess the alignment of the generated idea texts with both the trend and the company's context. The trends were extracted from a collection of articles, and the trend titles and keywords were derived from the entire set of articles. The trend description incorporated the titles and summaries of the top 10 articles pertaining to that specific trend.

To ensure the integrity and quality of the generated text and maintain consistency in the evaluation process, certain information from the articles was intentionally excluded. This includes the source or news outlet and the full text of the articles. Both GPT-3 and the reviewers were provided with short summaries of the articles. By providing only summaries of the articles to GPT-3, the potential issue of verbatim copy-pasting from the articles was mitigated, which has been identified as a challenge in text generation using Language Models (LLMs) [13]. The trend descriptions included the following details:

title of the trend
keywords of the trend, mined from the entire article set available for a trend. The reviewers were presented with the top 5 words.
articles comprising summaries of the top 10 articles. Each article consisted of the following information:
  - **title** short title of the article.
  - **description** containing a short summary of the article.

The corpus of 2170 texts, derived from the company and trend descriptions, underwent analysis by a total of 11 reviewers. Each text was evaluated by 3 reviewers, who assigned individual scores for the following metrics. All metrics were assessed on a scale of 0 to 5:

innovative 0 indicates a lack of any innovative elements and may come across as generic or unoriginal. 5 suggests a novel and highly specific idea.
context 0 signifies that the text fails to align with either the trend or the company's context. 5 fits well within the given trend and the company.
text 0 implies that the text exhibits clear faults or incoherence, 5 is easily readable, and could convincingly pass as human-written.

Table 2. Example description of a trend for company *Food*. The articles 3-10 are not shown.

Data	Example
title	Plant-based sugar substitute
keywords	fruit quality, sugar content, ingredients, nutritional, fruit weight
article 1 title	Chocolate food product
article 1 summary	A chocolate food product that is low in fat, dairy free, soy and lecithin free, free of added sugar or ingredients that increase sugar content, substantially starch-free, Isomaltooligosaccharide-free, oligosaccharide-free, maltitol-free, sorbitol-free, xylitol-free, erythritol-free, and isomalt-free. The chocolate food product may comprise: a cocoa butter, an unsweetened cocoa powder, a glycerin, a coconut cream, an almond milk, a pectin, a salt, a monk fruit blend, and a coconut flour.
article 2 title	Agent for increasing sugar content in fruit
article 2 summary	Provided is a compound and composition capable of increasing a sugar content in a fruit by a simple method, without being restricted by a cultivation area of a plant or a climatic environment. The agent for increasing a sugar content in a fruit of a plant comprises a compound represented by the formula MX as an active ingredient, wherein M represents alkali metal ion or alkaline earth metal ion, and X represents carbonate ion, hydrogen carbonate ion, acetate ion, citrate ion, succinate ion, phosphate ion, hydrogen phosphate ion, or pyrophosphate ion

The setup involved a systematic process for evaluating the texts. A reviewer selected a company and was provided with the company profile, outlined in Table 1. The reviewer received the 10 trends associated with the company, and was required to read through all 10 short summary articles for each trend (see Table 2). The generated ideas for a particular company were presented to the reviewer, organized by trend. The reviewer was then tasked with ranking each idea based on the metrics of *innovative*, *context* and *text*. Reviewers had the flexibility to adjust their scores during the evaluation of a single company. Once the evaluation for a company was completed, scores were finalized and could no longer be modified.

A reviewer was assigned to review all trends for a company. This approach enabled an accurate analysis of ideas in terms of their *context*, since all ideas for a company were considered collectively. This setup may raise concerns about the diversity of reviewers' perspectives for a single company, potentially impacting the validity of the experiment. We conducted an analysis of the average scores assigned by our reviewers, as depicted in Fig. 1. The graphic illustrates that none of the reviewers produced outlier scores. Although the ranges of scores for all categories are relatively wide, the metric of "innovation" exhibited the lowest range of quartiles. Based on this observation, we conclude that the assignment of reviewers to companies had minimal influence on the overall results.

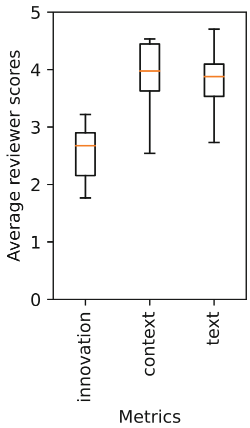

Fig. 1. Box plot of the average value that each reviewer assigned for the metrics.

Table 3. Average metric scores for all 70 ideas per company.

	Innovativeness	Context	Text
Average	**2.64**	**3.94**	**3.81**
Food	3.35	3.73	3.99
Automation	3.21	3.67	3.94
Hospitality	2.85	4.20	4.08
Non-profit	2.98	3.93	3.98
Spacecraft	3.19	3.95	4.02
Chemicals	2.49	4.20	3.86
IT	2.29	4.37	3.83
Law	2.34	4.14	3.88
Banking	1.77	3.58	3.47
Engineering	2.67	4.34	4.18
Legal	2.86	4.30	3.84
Education	2.82	3.57	3.62
Firearms	2.52	4.28	3.94
Laser	2.77	4.23	3.95
Nuclear	2.63	3.58	3.81
Recruitment	2.78	4.09	4.14
Entertainment	2.18	3.46	3.48
Healthcare	2.05	4.28	4.21
Social	2.43	3.63	3.49
Pharmaceuticals	2.44	3.74	3.32
Hospital	2.56	3.73	3.89
Media	2.44	4.14	3.93
Forestry	3.06	4.06	4.13
Sports	1.90	4.09	3.78
Dating	2.76	3.52	3.61
Software	2.44	3.81	3.54
Logistics	2.91	4.04	3.74
Insurance	2.61	3.55	3.50
Energy	2.77	4.14	3.73
Steel	2.90	4.03	3.82
Furniture	2.81	3.60	3.48

4 Results

We conducted an analysis of the ideas based on their average review scores per company, as presented in Table 3. The results show that the innovativeness is almost exactly the average, but the context and text quality generated via LLMs is rather high.

Table 4. Best and worst ranked *innovative* ideas.

Least innovative idea in ***Firearms - Trend: Purpose-driven Marketing***: using the power of purpose to drive marketing decisions and strategies. Through this approach, companies can create a strong bond with their customers by understanding what motivates them and creating unique experiences that will meet those needs. The goal is to use purpose-driven marketing to build loyalty, trust and engagement with customers
Most innovative idea in ***Firearms - Trend: Distributed Ledger***: creating a blockchain-based system that will enable users to securely store, track and transfer ownership of firearms. The Distributed Ledger technology created by *Firearms* is aimed at creating an immutable record of firearm transactions in order to improve safety and security for all involved parties. This new system will give law enforcement agencies the ability to trace weapons across their entire lifecycle, from production to purchase, sale, transport and storage. Additionally, it will provide real-time access to this data so that they can quickly identify suspicious activities or potential illegal activities related to firearms

Perception of Innovativeness: Was ranked the least high of all metrics. With an average score of 2.64 for all ideas, it suggests that the generated ideas fell in the middle range between not being innovative at all and being highly innovative. Certain industries or sectors, such as Banking and Sports, may inherently pose challenges in terms of generating innovative ideas. Banking is an established sector with limited scope for innovation, it is not surprising to observe lower scores. The sports industry is known for its continuous innovation, making it more difficult to identify ideas that have not already been explored.

Table 4 presents a selection of exemplary ideas, representing both the most promising and least innovative concepts overall. While several ideas received the same average score, we have chosen a single representative for each category. The least innovative idea related to *Firearms* (score of 0) is clearly lacking innovation, as it fails to offer any substantial new concept. In contrast, the notion of utilizing blockchain technology for weapons tracing stands out as a genuinely innovative proposal (score of 4.6). This idea not only highlights the potential applications of the technology, but also elucidates the resulting advantages it can yield.

The most innovative idea concerning *Firearms* was perceived as innovative by our reviewers. However, it is not entirely novel. Lokre et al. [12] describe a gun tracking system using blockchain technology in detail, and so do other publications [1,18]. Despite this, the idea itself was still perceived as novel.

A total of 7 ideas stood out as highly innovative, receiving an average score surpassing 4.5. Additionally, 97 ideas garnered a score of 4 or higher, making up approximately 4.47% of the corpus. Conversely, only 49 ideas were ranked with an innovativeness below 1, and a mere 3 ideas ranked below 0.5. Considering the average score of 2.64 across all ideas, it is evident that while Language Models (LLMs) like ours are not yet capable of true innovation, some ideas can be perceived as innovative by human readers who may not necessarily be domain experts. This was demonstrated in the case of the *Firearms* example.

Table 5. Best and worst fitting ideas in *context*.

Least contextual idea in ***Banking - Trend: Innovation transfer***: to create a new type of venture capital fund. This fund would focus on investing in innovative companies with the goal of transferring their technology and knowledge to other markets or sectors. The fund could be structured as an equity-based investment, debt financing, convertible notes, or any combination thereof. The fund would provide capital for early stage startups and established businesses that are working on cutting edge technologies that can be transferred to different markets and industries. Through this transfer of knowledge and technology, these businesses can gain access to larger customer bases and expand their reach into new areas of business opportunity
Most context-fitting idea in ***Energy - Trend: Energy planning as a school subject***: Introducing energy planning as a school subject. The goal of this project would be to educate the next generation on energy efficiency, renewable energy sources and their importance in our daily lives. This project would involve creating lesson plans for teachers around the topics of electricity, sustainability, hydropower, renewable energies and other related topics. *Energy* could provide resources such as technical information about their activities or experts from different areas to help create these lesson plans

Contextual Relevance: The *context* of the ideas received the highest average score of all 3 metrics, with a rating of 3.94. No company received a significantly negative score in this regard. This finding aligns with the known capabilities of LLMs, as they are designed to learn from and adapt to the provided context, enabling them to generate appropriate responses [20]. Research has shown that instructions on how to answer play a vital role in ensuring contextually faithful responses from LLMs [21,22]. Ideas ranked lower in terms of context, primarily focused on either the company for which the idea was intended or the specific trend the idea addressed, while neglecting the other aspect.

One of the ideas that ranked low in terms of context (see Table 5), pertains to the *banking* sector. Although the idea was generated within the context of *Innovation Transfer*, it primarily revolves around the concept of establishing a venture capital fund. While this idea is still relevant to the banking sector, it is more focused on investment funding rather than innovation transfer. In comparison, other ideas in the same sector proposed the creation of an innovation lab or a collaboration platform with universities, which align more closely with the intended context.

On the other hand, the most contextual idea, was specifically generated within the context of *Energy planning as a school subject*. This idea seamlessly integrates with the given context as it addresses the development of a lesson plan. Moreover, it aligns with the company's focus on providing sustainable energy solutions to its customers, further enhancing its contextual relevance.

Impressively, none of the ideas received a context ranking below 1. The poorest performing ideas obtained a ranking of 1.66. Out of the entire text corpus, a substantial 61.9% (1343 ideas) received a ranking of 4 or higher in terms of

Table 6. Best and worst *text* generated by the LLM.

Idea with bad text ranking in ***Insurance - Trend: Cybersecurity***: a comprehensive Cybersecurity solution that provides cutting-edge technology and services to protect businesses from cyber threats. The solution consists of four main components: 1) Security Awareness Training: This component helps organizations understand the risks associated with cyber threats and how to mitigate them. It also provides employees with training on best practices for safe online behavior, such as secure password management, secure data storage, and application security
Idea with good text ranking in ***Healthcare - Trend: Care home for seniors***: One concrete development that could raise sales numbers for *Healthcare* is the implementation of innovative technologies such as virtual reality (VR) and augmented reality (AR) to provide seniors with a more immersive care experience. VR and AR can be used to create interactive environments, allowing caregivers to provide better care while also providing seniors with an engaging way to stay connected with their families, friends, and activities outside of their homes. Additionally, these technologies can help reduce anxiety levels among elderly residents by creating calming visualizations or simulations

context. Furthermore, among these highly contextually relevant ideas, 354 ideas achieved a score of 4.5 or higher and were recognized as exceptionally relevant to the context for which the ideas were generated. This highlights the overall success in maintaining contextual coherence within the generated ideas.

Human-Like Text Quality: The evaluation of *text* quality yielded a slightly lower average score of 3.81 compared to the *context* ranking. No company received a low ranking in this aspect. Common shortcomings in the generated ideas included the lack of complete sentences at the beginning or end of the text. Worse-ranked ideas often exhibited more severe flaws, such as abruptly ending in the middle of a sentence or containing single-item lists, which compromised the overall cohesiveness of the text.

One of the worst ranked ideas with a value of 1.66 was identified in *Insurance* (see Table 6). The text starts with a lowercase character. The idea comprises four components, but only one of them is explained. While this limitation can be attributed to the token constraints during text generation, it clearly indicates that the text was generated rather than authored by a human.

One of several ideas ranked the maximum of 5 in *text*, is from *Healthcare*. The text exhibits grammatical correctness, introduces, and subsequently employs these abbreviations. The idea reads as a complete thought, without abruptly stopping in the middle of an explanation. This level of coherence and fluency further emphasizes the high-quality nature of the idea text.

Similarly to the *context* metric, it is notable that no ideas received a ranking below 1 in terms of *text* quality. 1343 ideas were ranked 4 or higher on average, indicating a generally good quality of generated text. 360 ideas (16.5%) received a score of 4.5 or higher, slightly more than in the context metric. This finding aligns with the advancements in language models, as it is becoming increasingly challenging to detect differences between human-generated and LLM-generated texts [19].

5 Conclusion and Outlook

Our analysis of 2170 texts generated by an LLM conclusively shows that an LLM is capable of most often generating *text* that is perceived as human equivalent, mostly fits a given *context*, and is capable of producing ideas that are only sometimes perceived *innovative*. These findings highlight the strengths and limitations of LLMs, showcasing their proficiency in mimicking human-like text while presenting the need for further advancements in generating innovative ideas.

Perception of Innovativeness: The study findings indicate that approximately 97 ideas, accounting for around 4.5% of the entire text corpus, were deemed highly innovative by human reviewers. Conversely, only 49 ideas were ranked as generic or lacking innovativeness. Considering the context of innovation management and the reduced effort required for an innovation manager to review and evaluate paragraph-long ideas, this level of innovativeness demonstrated by the LLM justifies its use case.

Contextual Relevance: A majority of 1343 ideas, or 61.9% of the total, were ranked as fitting the context, encompassing both the company and the trend for which the ideas were generated. Interestingly, in some instances, the LLM overlooked either the context of the given company or the context of the trend, resulting in lower scores. No idea was completely mismatched with the provided context.

Human-Like Text Quality: The evaluation revealed that 61.9% of the entire text corpus received rankings indicative of human-like text quality, with no text being deemed completely faulty. This outcome aligns with expectations, as LLMs have demonstrated their ability to generate text that closely resembles human-written content.

Our study on the innovativeness of LLM-generated texts within a given context has provided conclusive insights. However, it is important to note that there are still many aspects to consider when evaluating LLMs. Existing benchmark frameworks like the Holistic Evaluation of Language Model (HELM) Benchmark [11] encompass various metrics relevant metrics such as fairness and toxicity, but they may not cover task-specific metrics such as *innovativeness*, which is highly relevant in innovation management.

In the field of innovation management, there is still much work to be done. One potential avenue is exploring unsupervised methods for guiding the idea space using LLMs to conduct idea ranking, similar to the process undertaken by our human reviewers. This approach could enable the automatic ranking or filtering of ideas based on their scores, thereby enhancing the presentation of novel ideas to human innovation managers.

Acknowledgements and Funding. Funding was provided by the Austrian Research Promotion Agency (FFG) under the Project Explainable Creativity (EACI, project number 892004).

We thank AnyIdea (https://anyidea.ai/) for their provision of data sets used in this work. We thank the project partner Cloudflight (https://www.cloudflight.io/), especially Michael Weissenböck, Anna Hausberger and Rine Rajendran, for their valuable contributions to this work. We thank Michaela Jungwirth and Marius Elflein for the conceptualization and the organization of the evaluation, and the reviewers of the texts for their valuable contribution to this work.

References

1. Akello, P., Vemprala, N., Lang Beebe, N., Raymond Choo, K.K.: Blockchain use case in ballistics and crime gun tracing and intelligence: toward overcoming gun violence. ACM Trans. Manage. Inf. Syst. **14**(1), 1–26 (2022). https://doi.org/10.1145/3571290
2. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610–623. ACM (2021). https://doi.org/10.1145/3442188.3445922
3. Bouschery, S.G., Blazevic, V., Piller, F.T.: Augmenting human innovation teams with artificial intelligence: Exploring Transformer-Based Language Models. J. Product Innov. Manage. **40**(2), 139–153 (2023). https://doi.org/10.1111/jpim.12656
4. Brown, T.B., et al.: Language models are few-shot learners. CoRR arXiv:2005.14165 (2020)
5. Chowdhery, A., et al.: PaLM: scaling language modeling with pathways. arXiv:2204.02311 (2022)
6. Cooper, R.G., Edgett, S.: Ideation for product innovation. What are the Best Methods, pp. 12–17 (2008)
7. Del Vecchio, P., Di Minin, A., Petruzzelli, A.M., Panniello, U., Pirri, S.: Big data for open innovation in SMEs and large corporations: trends, opportunities, and challenges. Creat. Innov. Manage. **27**(1), 6–22 (2018). https://doi.org/10.1111/caim.12224
8. Di Fede, G., Rocchesso, D., Dow, S.P., Andolina, S.: The idea machine: LLM-based expansion, rewriting, combination, and suggestion of ideas. In: Creativity and Cognition, pp. 623–627. C&C 2022, Association for Computing Machinery. https://doi.org/10.1145/3527927.3535197
9. Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., Fritz, M.: More than you've asked for: a comprehensive analysis of novel prompt injection threats to application-integrated large language models (2023). arXiv:2302.12173
10. Krauss, O., Aschauer, A., Stöckl, A.: Modelling shifting trends over time via topic analysis of text documents. https://doi.org/10.46354/i3m.2022.mas.009. https://www.cal-tek.eu/proceedings/i3m/2022/mas/009
11. Liang, P., et al.: Holistic evaluation of language models (2022)

12. Lokre, S.S., Naman, V., Priya, S., Panda, S.K.: Gun tracking system using blockchain technology. In: Panda, S.K., Jena, A.K., Swain, S.K., Satapathy, S.C. (eds.) Blockchain Technology: Applications and Challenges. ISRL, vol. 203, pp. 285–300. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-69395-4_16
13. McCoy, R.T., Smolensky, P., Linzen, T., Gao, J., Celikyilmaz, A.: How much do language models copy from their training data? Evaluating linguistic novelty in text generation using RAVEN (2021). arXiv:2111.09509
14. Min, B., et al.: Recent advances in natural language processing via large pre-trained language models: a survey. CoRR arXiv:2111.01243 (2021)
15. Radford, A., et al.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)
16. Sarsa, S., Denny, P., Hellas, A., Leinonen, J.: Automatic generation of programming exercises and code explanations using large language models. In: Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1. ICER 2022, vol. 1, pp. 27–43. Association for Computing Machinery. https://doi.org/10.1145/3501385.3543957
17. Scao, T.L., et al.: Bloom: A 176b-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100 (2022)
18. Soni, D.B., Mahler, M.L.: Blockchain technology for a firearm registry. New Zealand J. Business Technol. **2**, 34–42 (2020)
19. Tang, R., Chuang, Y.N., Hu, X.: The science of detecting LLM-generated texts. arXiv:2303.07205
20. Wang, X., Zhu, W., Wang, W.Y.: Large language models are implicitly topic models: explaining and finding good demonstrations for in-context learning (2023)
21. White, J., et al.: A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv:2302.11382
22. Zhou, W., Zhang, S., Poon, H., Chen, M.: Context-faithful Prompting for Large Language Models. arXiv:2303.11315