Article

Measuring the Impact of ChatGPT on Fostering Concept Generation in Innovative Product Design

Stefano Filippi

DPIA Department, University of Udine, 33100 Udine, Italy; filippi@uniud.it; Tel.: +39-0432-558289

Abstract: The growing demand for innovative and user-centric product design has led to a growing need for effective idea generation methods. In recent years, natural language processing (NLP) tools such as ChatGPT have emerged as a promising solution for supporting idea generation in various domains. This paper investigates a framework for studying the role of ChatGPT in facilitating the ideation process in product design. This investigation measures the impact of ChatGPT on the generation of innovative concepts compared to the use of "classic" design methods. An overview of the state-of-the-art idea generation methods in product design opens the paper. Then, the paper highlights some hypotheses about the impact of ChatGPT on innovative product design, aiming for product augmentation by adding features. The paper then describes the design experience in which ChatGPT is used as a tool for concept generation. Finally, the paper analyzes the dataset, using precise metrics to characterize the participants' performance and compare them. This analysis allows the paper to argue about the validation/rejection of the hypotheses. The paper concludes with a discussion of the implications of the findings and some suggestions for future research. Along with the paper, the Microsoft Excel workbook used to perform the data analysis is available to the readers to perform their own data collection and analysis. The workbook UX has been carefully studied and developed to make it usable by anyone. At the same time, it should be flexible enough to manage several situations characterized by different numbers of participants, product functions to implement, and generated concepts.

Keywords: idea generation; conceptual design; product design; ChatGPT; innovation metrics

1. Introduction

Concept generation is a critical phase in product design [1,2]. If innovation is the goal, the product development process must separate the analysis of "what" the product will be called to do from the study and development of "how" that "what" will be accomplished. In other words, there are two stages in the product development process: a first stage, when pure product functions are considered, and a second stage, when those functions are embodied in physical products. This two-stage process helps to keep design degrees of freedom open as long as possible [3]. Concept generation is in between the two stages, and this, together with other peculiarities, makes its role fundamental for the success of the design activities.

There are many different methods that designers use to generate ideas, each with its own strengths and limitations [4]. These methods range from simple brainstorming to more structured methods like TRIZ and emerging AI-based methods like ChatGPT or Google Bard [5,6].

Inside this scenario, this research aims to develop a framework for measuring the impact of ChatGPT, the only AI-based method available in Italy at the time of this research, on the generation of ideas (concepts) in product design. The framework will compare the performance of designers who use classic methods to the performance of designers who use ChatGPT in designing products aiming for product augmentation by adding features. The comparison will be conducted both quantitatively and qualitatively, using well defined

metrics [7]. Along with the paper, the Microsoft Excel workbook used to perform the data analysis is available to the readers to perform their own data collection and analysis. The workbook UX has been carefully studied and developed to make it usable by anyone. At the same time, it should be flexible enough to manage several situations characterized by different numbers of participants, product functions to implement, and generated concepts.

The paper is structured as follows. After Section 2, which depicts several idea generation methods, Section 3 describes a design experience conducted with 18 participants. The participants were tasked with generating concepts that implemented a given set of product functions. Section 4 presents the outcome of the design experience. Section 5 analyzes the data and draws some considerations. Section 6 and the references conclude the paper.

2. Background

Idea Generation Methods

Traditional methods of idea generation include brainstorming, mind mapping, and sketching. These methods have been described in the literature [8–11]. Brainstorming is a group ideation technique that encourages participants to generate as many ideas as possible without criticism. This allows for a free flow of ideas, which can lead to more creative solutions. Mind mapping is a visual tool that can be used to represent the relationships between different concepts. This can help designers to see the big picture and to identify potential connections among seemingly unrelated ideas. Sketching is a common method used by designers to quickly generate and communicate ideas. This allows designers to capture their ideas in a tangible way, which can help to refine and develop them later.

Among more recent methods, the extreme-inverse method is a technique for generating ideas that challenge existing assumptions and can lead to innovative solutions. This method involves identifying the extremes of a particular product attribute and then generating ideas that are the opposite of these extremes. For example, if a product is designed to be lightweight, the extreme-inverse method would involve generating ideas for a product that is as heavy as possible. This approach can help designers to think outside the box and to generate creative solutions [12].

As another more recent method, the analogies-metaphors method is a technique for generating ideas by drawing inspiration from unrelated fields or domains. This approach allows designers to look at problems from new perspectives and to generate creative solutions. For example, if a designer is trying to solve a problem with a product, they might look at how the problem is solved in a different field, such as biology or physics. This can help the designer to come up with new ideas for solving the problem [13].

TRIZ is a problem-solving methodology that aims to develop systematic approaches to solve technical problems. It is one of the most well-known logical problem-solving methods [14]. TRIZ involves identifying common patterns in successful innovations and applying them to new problems. This allows designers to draw on the knowledge of past successes to solve new problems. The most well-known and used design tools offered by TRIZ are the 40 principles, the trends of evolution, the contradiction matrix, and trimming. These tools have been used successfully in a wide range of industries, including aerospace, automotive, and manufacturing [15–18]. The 40 principles are based on the analysis of a large number of patents and ideas. These principles summarize different hints to stimulate creative thinking and guide problem-solving processes by providing a systematic approach to generate innovative solutions. For example, one of the 40 principles is "segmentation", which suggests that a problem can be solved by dividing it into smaller, more manageable parts. Trends of evolution describe the patterns observed in the development of technological systems. These trends can help designers to understand the direction of technological progress and to guide their own inventive thinking. For example, one of the trends of evolution is "miniaturization", which suggests that technological systems will tend to become smaller over time. When designers try to improve something in a product, quite often something else risks getting worse. Thus, a contradiction occurs. The TRIZ contradiction matrix contains references to the 40 principles for each contradiction between improving and worsening product features.

This can help designers to find solutions to contradictions. Along with the 40 principles, the trends of evolution, and the contradiction matrix, TRIZ trimming is a tool that contributes to gaining product ideality. An ideal product has all the features and functions needed without any costs or drawback. Trimming helps to achieve product ideality by deleting unnecessary product components and moving the useful functions to the remaining components. TRIZ has been used successfully in a wide range of industries, such as aerospace, automotive, and manufacturing.

More recently, digital tools and techniques have been developed to support idea generation in product design. These include generative design tools, data-driven approaches, and artificial intelligence (AI) tools such as ChatGPT and Google Bard [19]. Generative design tools use algorithms to generate a large number of design options, which are based on predefined rules and constraints [20]. Data-driven approaches use user data and feedback to guide ideation and design decisions [21]. AI tools use natural language processing and machine learning algorithms to support ideation by interacting with users in natural language [22].

While these digital tools and techniques have the potential to enhance idea generation in product design, there are still challenges that need to be addressed, such as the potential for a lack of creativity and the difficulty in translating generated ideas into tangible designs.

Among the methods summarized above, this research exploited brainstorming, analogies-metaphors, extremes-inverses, TRIZ (limited to the 40 principles), and ChatGPT.

3. Activities

In order to set up a framework to measure the impact of ChatGPT on fostering concept generation in innovative product design, the author carries out a design experience in the field involving university students and colleagues. The research path is as follows. First, the two product innovation experts who supervised the design experience highlighted the metrics for establishing the quality of the concepts and comparing the performance of the participants. The supervisors then hypothesized about the impact of ChatGPT on product innovation, which the design experience was designed to confirm or reject. The next activity was to characterize the design experience referring to the double diamond approach and implement it, which involved preparing the materials, selecting the participants, executing the design activities, and collecting the data. Only half of the participants had access to ChatGPT. Once finished, the supervisors discussed the results with the participants and then analyzed the dataset to verify the hypotheses.

3.1. Highlighting the Metrics

Four metrics have been selected from the literature to use in the framework. They are quantity, usefulness, novelty, and variety, as described in the research reported in [23,24]. The selection of these metrics makes it easier to explore possible relationships between personality traits and ChatGPT use, since previous research has established a relationship between the big five personality traits [25,26] and design activities.

The four metrics are defined and used as follows.

- Quantity. The quantity metric simply measures the number of concepts generated.
- Usefulness. The usefulness metric measures the applicability of each concept in the specific context; the supervisors will assign a value using a [0, 1] interval (useless to fully useful).
- Novelty. A concept is as novel as it has not been exploited in the specific context before. Again, novelty is assigned to each concept using the [0, 1] interval (well known to novel).
- Variety. The variety metric measures how original a concept is in the dataset. Again, the interval is [0, 1]. (If everybody highlighted that concept, the variety score will be 0; if the concept has been highlighted just once, the value will be 1).

3.2. Stating the Hypotheses to Verify

The considerations in the background section, along with the experience of the supervisors, allow for highlighting some hypotheses that can be verified through the design experience. These hypotheses, quite naturally mapped with the metrics described before, are as follows.

H1. *Given its wide knowledge base, ChatGPT is expected to foster much larger numbers of concepts than those highlighted by the participants not allowed to use it.*

H2. *This hypothesis refers to usefulness and is the most interesting one. ChatGPT could be seen as a big extension of TRIZ, as it has a much wider knowledge base than TRIZ (which is based on patents and inventions only). At the same time, just like TRIZ, ChatGPT can easily overcome the NIH—not-invented-here syndrome—and PI—psychological inertia [27–29]. However, the ability of ChatGPT to filter suggestions about concepts that could solve problems or implement functions is unknown, as is how this ability depends on the interaction between the designers and ChatGPT. From this point of view, TRIZ's strategy of processing data at a general level and delegating designers to the customization of the proposed solutions could be a winning strategy for generating useful concepts. All of this makes the expectations about the design experience outcome even more intriguing, depriving the supervisors of being able to get any real prediction. The supervisors could just suppose that there is a balance between ChatGPT ability to suggest concepts due to its wider knowledge base and TRIZ's finer strategies to focus and filter possible design suggestions. Finally, usefulness will impact the computation of the other metrics, as it is a sort of go-no go filter for considering concepts based on their novelty and variety.*

H3. *The novelty of the concepts generated by participants who use ChatGPT is expected to be low, as ChatGPT works on existing pieces of information reporting experiences that have already happened in the past.*

H4. *The variety of the concepts generated by participants who use ChatGPT is expected to be higher than that of the of the concepts generated using classic methods given the wideness of the ChatGPT knowledge base. However, again considering usefulness as a filter that removes out-of-scope concepts, etc., the results from the teams are expected to be comparable.*

These four hypotheses will be considered again once the dataset from the design experience is available. Their validation (confirmed vs. rejected) is described in Section 5, which is structured to find an exact match.

3.3. Characterizing the Design Experience According to the Double Diamond Approach

The double diamond approach is a well-known design process model that helps teams to understand and solve problems creatively [30]. It consists of four phases: discover, define, develop, and deliver. In the discover phase, the team gathers information using different methods and tools. In the define phase, the team understands the problem in detail by identifying the causes and the people affected by it. During the develop phase, the team generates ideas and explores different possibilities. In the deliver phase, the solutions are implemented and tested with the users; moreover, the outcomes are communicated to the stakeholders. All of this is double diamond shaped because flowing through the phases is represented by a diverge-converge-diverge-converge sequence. The mapping between the design experience used in this research and the double diamond approach can be seen as follows.

Regarding the discover phase, a very simple, mature, well-known product was considered. This reduced the divergence in analyzing the data, with all of this done in order to state the starting point for every participant of the test as more or less the same. The same considerations last for the define phase; converging to the problem to solve and to the product aspects to focus on to innovate it led to very similar conclusions for all the participants. The design methods and tools studied during the university course were adopted in the develop phase. Here, that mix should have encouraged as much divergence as possible. For example, TRIZ fundamentals make clear that this theory of inventive

problem solving aims at generating as many ideas as possible. Finally, the deliver phase finds correspondence in the design experience during the data analysis and evaluation performed by the innovation experts, who assigned marks to the concepts generated by the participants regarding the usefulness and novelty metrics.

3.4. Implementing the Design Experience

A total of 18 students were involved in the design experience, all of them enrolled in the "Product Interaction and Innovation" university course of the Management Engineering and Mechanical Engineering Master's Degrees at the University of Udine (Italy). The course focuses on the design process, mainly the mechanical one but not limited to it, from the very beginning—highlighting of customers' requirements—to the end-of-life of the product, with particular emphasis on innovative tools and methods (QFD, TRIZ, ChatGPT, etc.) for each step of the process. Along with this, ideation, concept generation and product evaluation consider product UX as one of the most important aspects; again, tools and methods for UX innovation (design thinking, persona development, journey-mapping, etc.) are described and put into practice during laboratorial activities. A total of 13 males and five females participated; nine of them were 20 years old and nine were 22. Their background was mainly technical, with a couple of exceptions coming from classic high schools. All of them knew the brainstorming method before accessing the university lessons but none got in touch with analogies/metaphors, extremes/inverses, or TRIZ. Some had very limited, unstructured experiences with ChatGPT, mainly because it had become available in Italy just some weeks before and they were just curious about it. With all of this said, it can be assumed that the students had more or less the same knowledge about the fundamentals of the concept generation methods used in the design experience, from brainstorming to TRIZ, with some slight difference regarding ChatGPT. A random selection but considering prior knowledge about ChatGPT, generated eight teams of two students each. Having nine sets of data rather than 18 makes data collection and analysis faster and easier. Clearly, this change of granularity could impact the meaningfulness and robustness of the research results, but the main point here is to set up a framework that will be applied in the future to get more data and to analyze the trends. Four teams (T1 to T4) were assigned to only use the classic generation methods; these teams will be addressed hereafter as the G1 group of teams. The other five teams (T5 to T9) were allowed to access ChatGPT; they are referred to as the G2 group of teams.

Two supervisors with pluriannual experience in product innovation methods and tools took part in the experience. They decided the design experience logistics and prepared the materials to conduct it.

The first decision was about the order of the methods that the teams would use. The peculiarities of each method made their sorting easy. For example, the extremes-inverses method needs existing concepts to start from when generating variations. Therefore, this method cannot occur first when no concept has been already highlighted. In the end, the actual order of the methods was brainstorming, analogies-metaphors, extremes-inverses, TRIZ, and ChatGPT, addressed by the letters B, A, E, T, and C, respectively.

The design problem was the development of an innovative sharpener for classic wooden pencils. This selection came from the following requirements. First, the number of functions to manage should be limited, since the whole design experience must take at most one lesson. Second, the product should show a mature design in order to level the problem-solving of the teams by clearing it from contributions about creativity, etc., that are too exotic. Third, simplicity and maturity are coupled with the fact that the product must be well-known by everybody; all of this is required in order to lower the bias and focus on the impact of the ChatGPT use as much as possible. Again, always to introduce as little bias as possible, the functional scheme of the product was delivered to the students. Figure 1 reports that. The topmost box contains the overall function; the sentence in the verb-object format that describes the functional scope of the product. The other boxes contain three main functions (F1, "Sharpen correctly", and F5) and three subfunctions (F2, F3, and F4).

The leaf functions in the graph (F1 to F5) are those functions to find concepts for. Thus, this set is fixed for every team.
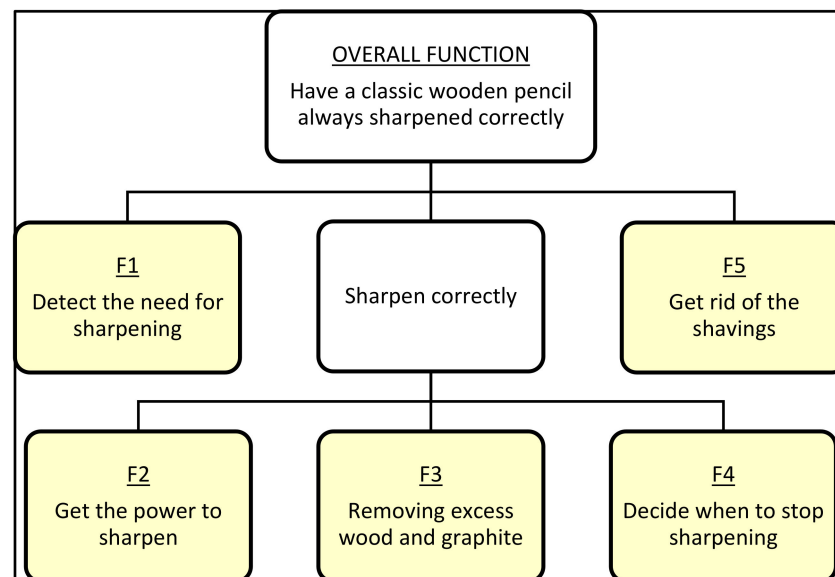
Figure 1. The functional scheme of the product, an innovative sharpener for classic wooden pencils.

The expected result for every team is a filled table, called a morphology [3], where five rows correspond to the leaf-functions F1–F5, and each row contains a list of concepts implementing the specific function. Each concept is tagged with the letters B, A, E, T, and C in order to identify the method used to highlight it. Exceptionally, due to the possible suggestions about further functions coming, for example, from the analogies-metaphors method and its exploitation of the semantic fields [31] or from ChatGPT, the teams were allowed to add functions with the related concepts.
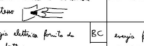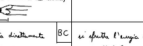
The teams had one hour and a half to fill the morphology by sketching the concepts or describing them textually on paper sheets to avoid delays. For each concept, there was a box allowing the insertion of the letters B, A, E, T, and C, corresponding to the five methods available. Explanations of the sketches by the students would come afterwards, in an offline mode, to get the real meaning and to allow the supervisors to classify and evaluate the concepts as best as possible. To limit bias as much as possible again, the adoption of each method was given a precise amount of time. The supervisors set this timing: brainstorming was given 20 min, analogies-metaphors and extremes-inverses 15 min each, and TRIZ and ChatGPT 20 min each. Teams not allowed to use ChatGPT (teams T1 to T4) could use the last 20 min to refine their findings or to highlight more concepts using any method they like (except for ChatGPT, of course).

A briefing before the start of the test allowed the supervisors to remember the objectives, the rules to follow, the use of the material, and the behavior to adopt. Furthermore, the precise reason why those design methods and tools were involved, in addition to having been studied in the university course, was explained using the concept of product augmentation [32]. The different ways a product can be improved to meet marketing requirements and make it stand out from the competition can refer to additional features, services, or benefits. Here, dealing with a simple product like a pencil sharpener, additional features are chosen to innovate the product rather than adding services or benefits. Therefore, in addition to brainstorming, those proposed (analogies-metaphors, extremes-inverses, and TRIZ) are the most used methods and tools to add features and optimize products. Clearly, thanks to its versatility, ChatGPT would help all three aspects (features, services, and benefits); hence, it is added easily and smoothly. The scope of the experiment does not include customer expectations; in other words, there has not been a preventive analysis in the field on real needs. If, on one hand, this does not allow us to quantify the complexity of

the tasks relating to product augmentation, on the other, it does not limit design freedom, also aiming to discover unexpected solutions, even for the most demanding customers.

4. Results

At the end of the design experience, nine morphologies are collected. Figure 2 shows one of the morphologies (concepts are in Italian).

Figure 2. One of the morphologies generated during the design experience.

The logs of the interactions with ChatGPT are also collected from teams T5 to T9 to get more data for further research, analysis, and comparisons.

To get some examples of concepts highlighted during the design experience, the concept of "solar panel", referring to function F2, "get the power to sharpen", appeared in the morphologies of teams T2, T4, T5, T6, and T8, highlighted thanks to the A(nalogies-metaphors), B(rainstorming), B(rainstorming), C(hatGPT), and C(hatGPT) design methods, respectively. The concept of "gravity", referring again to function F2, appeared only in the morphology of team T5, highlighted thanks to the B(rainstorming) design method; finally, the concept of "burner," referring to function F5, "get rid of the shavings", appeared in the morphology of teams T1, T2, T4, T7, T8, and T9, highlighted thanks to the B(rainstorming) design method by everyone except for T9, where the A(nalogies-metaphors) allowed highlighting it.

As a first consideration, it is quite surprising that none of the teams added any new functions, given the kind of design problem proposed (a well-known product, performing a daily action for thousands of engineering students).

Having said this, the quantity metric was considered. In all, the nine teams generated 92 different concepts, distributed as follows: 18 concepts referred to function F1, 18 to F2, 22 to F3, 20 to F4, and 14 to F5. Figure 3 shows the number of concepts generated by each team, highlighting the numbers of concepts coming from the different methods adopted. The far right of the bar chart compares the situations between the teams not allowed to use ChatGPT and those that could do it.

Next, the supervisors started evaluating each concept from the usefulness point of view. This is because useless concepts would have been ignored from that moment on. Figure 4 shows the result of this evaluation, the teams' performance regarding the usefulness metric. Each bar represents the total number of concepts generated by the team and the portion

of them showing any usefulness (weighted sum, based on the usefulness values, and percentage). Again, the far right of the bar chart compares the situations between the teams not allowed to use ChatGPT and those that could do it. It is worth mentioning that although the allowed interval for usefulness was [0, 1], for simplicity, the supervisors used only the values 0 (useless), 0.5 (partially useful), and 1 (fully useful).
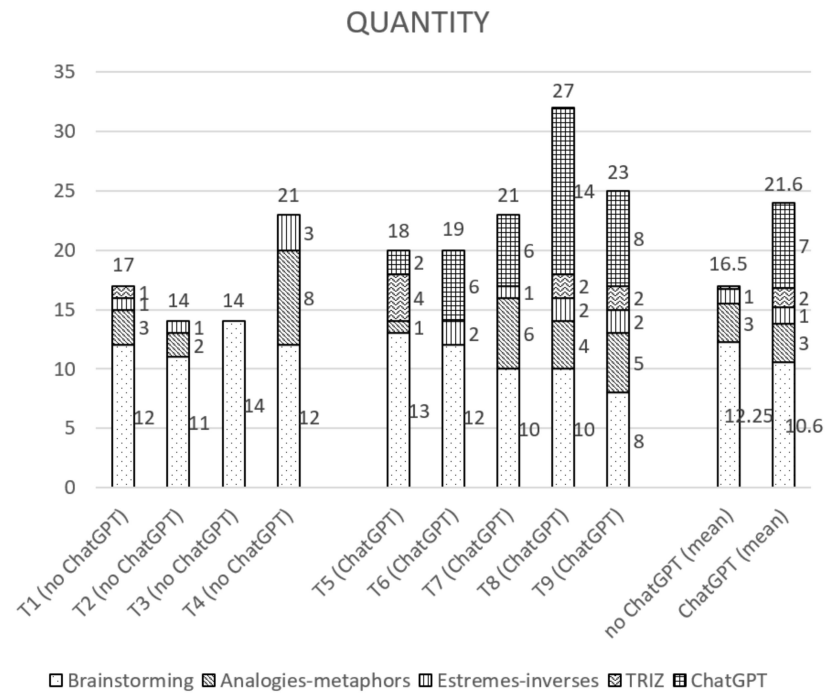
Figure 3. The bar chart representing the quantity metric.
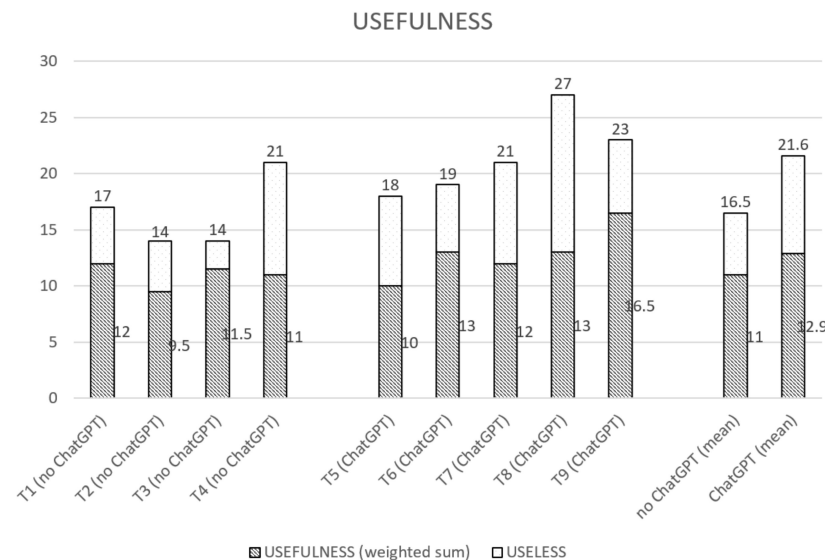
Figure 4. The performance of the teams regarding the usefulness metric.

Then, novelty and variety values were associated with the concepts showing some usefulness. Figure 5 reports the performance of the teams regarding the novelty metric. Each bar represents the total number of concepts generated by the team, with the highlighted portion of them showing some novelty (weighted sum, based on the novelty values of each concept, and percentage). The far right of the bar chart compares the situations between the teams not allowed to use ChatGPT and those that could. Again, it must be noted that

although the allowed interval for novelty was [0, 1], for simplicity, the supervisors used only the values 0 (known), 0.5 (partially novel), and 1 (fully novel).
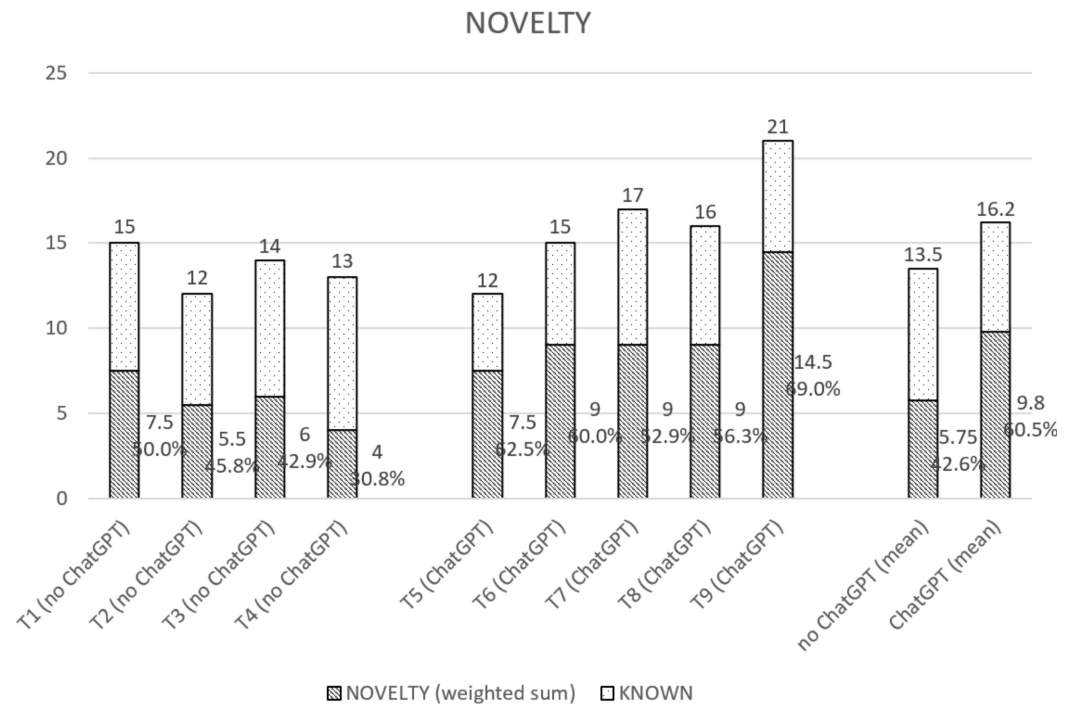
Figure 5. The performance of the teams regarding the novelty metric.

Similarly, Figure 6 reports the performance of the teams referring to the variety metric. The bars represent the variety of each team (weighted mean, based on the variety values of each concept). The far right of the bar chart compares the situations between the teams not allowed to use ChatGPT and those that could.
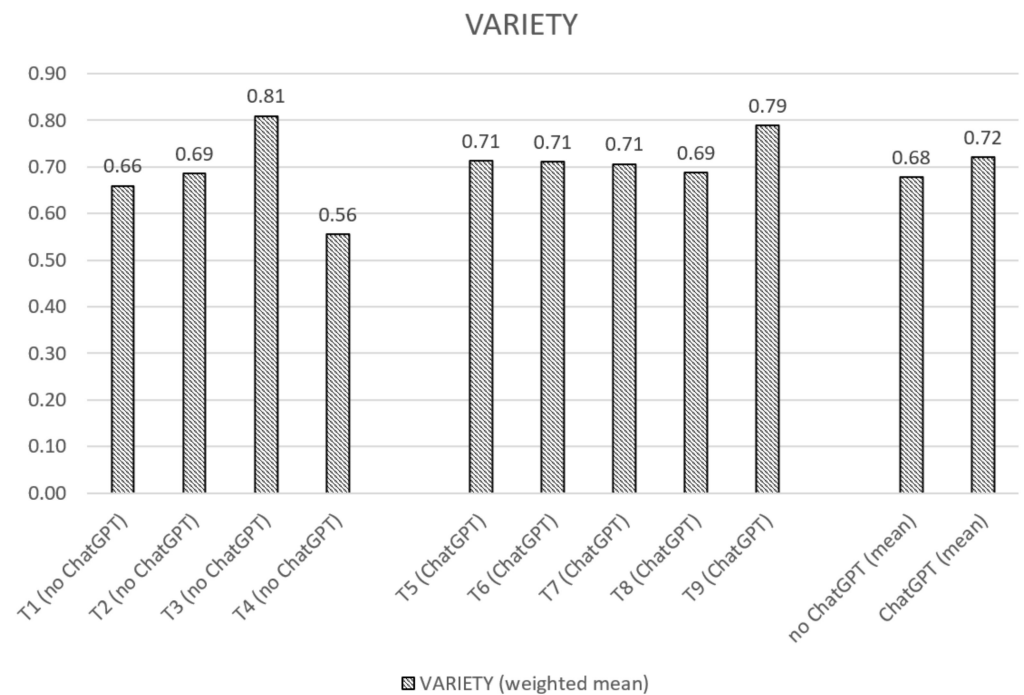
Figure 6. The performance of the teams regarding the variety metric.

The Microsoft Excel workbook used to elaborate the dataset generated during the design experience can be downloaded here: (https://uniudamce-my.sharepoint.com/:x:/g/personal/stefano_filippi_uniud_it/EbphTEZJrZ5Nnke7mANJs3gBeYqOasGBBC_klAP0G91tfA?rtime=rZH55FSZ20g (accessed on 18 June 2023)). Aside from obviously working properly, the workbook has been developed to be usable enough for anyone to collect data and do their own analysis. Thus, it can be used in different design contexts, from schools to academia/research centers and industries, in a homogeneous way. Figure 7 shows the user interface of the workbook. It replicates the procedure to collect the data (by offering, among other things, the material for doing it), insert them into the data sheet, and perform the analysis. It works with up to nine product functions, two groups (no ChatGPT allowed and ChatGPT allowed) of up to ten teams each, and one hundred generated concepts (different from each other). The author reputes that these numbers are enough to make the workbook suitable for almost any research situation.
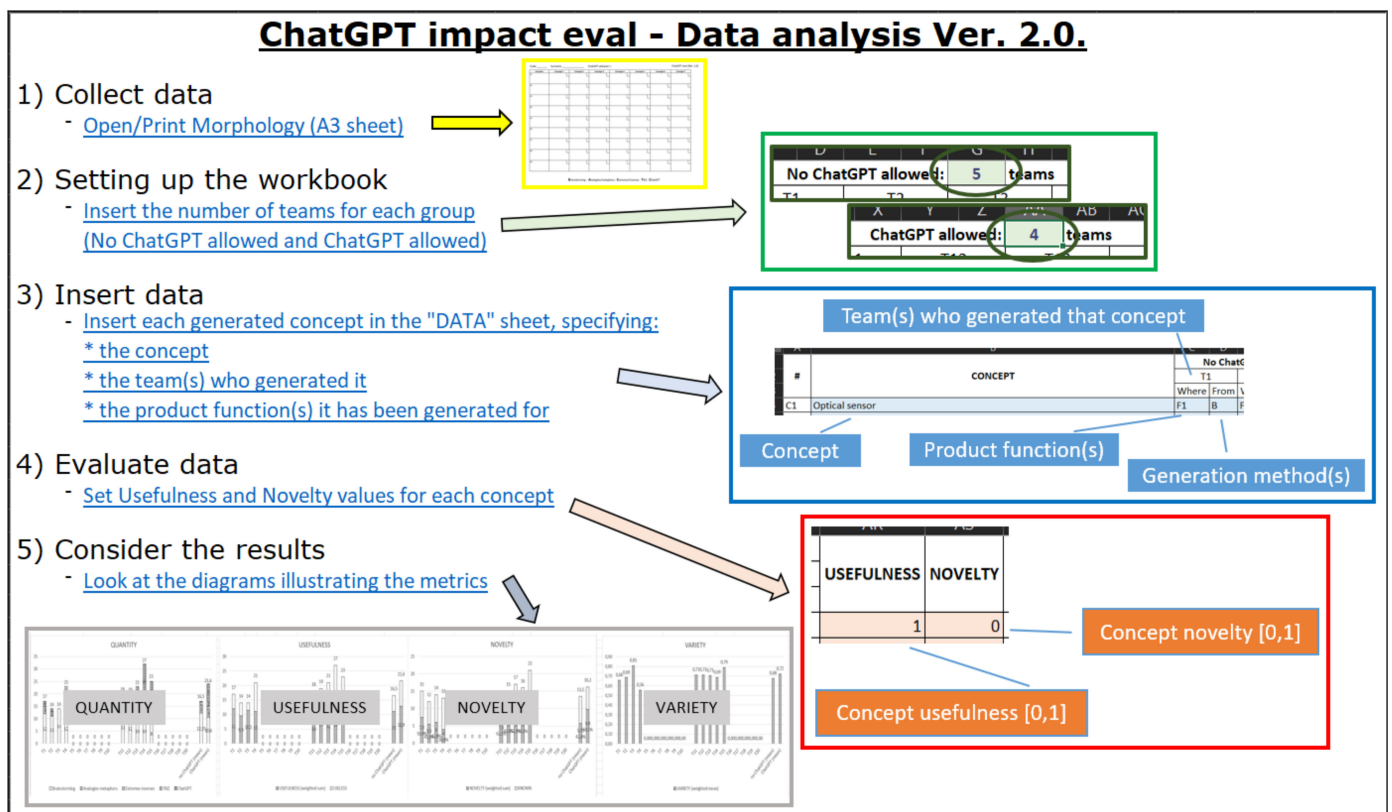
Figure 7. The user interface of the Microsoft Excel workbook used in this research, which is available to everyone who would like to perform the same study.

5. Discussion

Before dealing with the precise validation of the hypotheses, some considerations can be drawn from the analysis of the design methods from which the highlighted concepts come. First, it seems that teams in G1 exploited brainstorming much more, on average, than teams in G2. In G1, 49 out of 66 concepts (74%) come from brainstorming. In G2, this percentage dropped to 49% (53 out of 108). The supervisors justify this with the longer time that G1 used the design methods they liked while G2 was using ChatGPT. Brainstorming is undoubtedly the easiest method to adopt among those available; thus, G1 likely used it for 20 min longer than G2. Second, it is worth mentioning the singular distribution of the concepts highlighted thanks to the TRIZ 40 principles. In G1, only one concept appears, from team T1. In G2, eight concepts were generated by three teams out of five. At the moment, the supervisors do not have a precise explanation for this other than a simple, random distribution. It could be that the concepts suggested by ChatGPT reminded G2

teams of one or more TRIZ principles, and for this reason, teams T5 to T9 tagged those principles with T instead of C. Clearly, more data are needed to give more precise answers for these impressions.

Referring to hypothesis H1, considering Figure 3, teams T1 to T4 highlighted an average of 16.5 concepts, while teams T5 to T9 reached a value of 21.5. This suggests that ChatGPT clearly boosts the design activities a lot in terms of the number of concepts generated. This is enough to validate hypothesis H1.

Regarding H2, Figure 4 shows that despite the higher numbers of concepts generated by teams T5–T9, the percentages of concepts considered as somehow useful (i.e., with a usefulness value different from zero) are much lower for these teams. The same trend is observed for the average values for the two groups of teams (66.7% for G1 and 59.7% for G2). This starts to give answers to the H2 hypothesis. Surprisingly, ChatGPT was not as helpful in suggesting useful concepts (on average). There is no definitive answer for the reason why the concepts suggested by ChatGPT do not seem to be as focused on implementing the product functions and, ultimately, on solving the design problem. Is it due to bad interaction with ChatGPT, or to bad use of it? Further investigation is needed. Nevertheless, as planned, the usefulness evaluation lowered the number of concepts to evaluate from then on. They moved from the total of 92 to the useful or somehow useful 58.

Regarding H3, the percentages shown in Figure 5 make clear that all novelty values for teams in G2 are higher than those for teams in G1. This contradicts the hypothesis that ChatGPT would suggest low-novelty concepts. The fact that ChatGPT suggests design solutions by exploiting pieces of information referring to things that have already happened in the past (and are conveniently reported) does not seem to be a limit on suggesting novel design solutions. However, it is worth noting that almost half of the initial concepts were excluded from the novelty evaluation due to the usefulness filter; therefore, the larger quantities of concepts generated in G2 made a difference. In other words, it could be said that "the more concepts generated, the more likely that some of them will be novel", As a result, checking the H3 hypothesis is not straightforward. If all the generated concepts (useful or not) are considered, novelty values between G1 and G2 are quite comparable. On the other hand, if useless concepts are excluded, G2′s performance appears much better that that of G1 from the novelty point of view. All of this could be interpreted differently. The adoption of the "usefulness filter" could be seen as similar to what TRIZ does when selecting the best solutions to suggest. The big difference is that TRIZ has the selection strategy embedded, while the "usefulness filter" was applied by the supervisors during the dataset analysis.

Finally, regarding hypothesis H4, Figure 6 shows that the mean variety values of G1 and G2 are quite comparable (0.68 vs. 0.72). Therefore, the hypothesis seems to be confirmed. However, there is something else to note. The variance in G1 (0.00795) is much higher than in G2 (0.001216). This suggests that ChatGPT somehow leveled the teams' performance from the variety perspective. Teams that performed differently (sometimes significantly differently) in terms of quantity, usefulness, and novelty appear to be comparable in terms of generating varied design solutions. The only reason for this could be that since ChatGPT uses the same knowledge base, the answers it provides to users are more or less the same. Clearly, ChatGPT seems to make its pieces of information available regardless of the specific user. This could also highlight that the way users interact with ChatGPT (the questions they ask and how they ask them) is almost irrelevant. This is an important point to consider. In any case, the analysis of the conversations between the participants and ChatGPT, collected at the end of the design experience, highlighted interesting considerations about the different approaches adopted. Students went from looking for confirmation of their own concepts ("What do you think about a laser beam to get rid of the shavings when sharpening a wooden pencil?") to asking direct questions ("Give me some concepts on detecting the need to sharpen a wooden pencil") to writing the entire problem to be solved and asking for help ("I am developing a wooden pencil sharpener. I need some concepts to implement the required subfunctions"). All

of this shows the different ways in which participants conceptualized ChatGPT and its potentialities, which is likely due to the short time passing since it became available. It is also worth noting that these different approaches led to an increasing amount of help from ChatGPT. When asked for simple confirmations, ChatGPT limited its intervention to almost yes/no answers. On the other hand, when the entire problem was posed, ChatGPT exploited its freedom and suggested several solutions, organized by topic, and described the inferential process that led to them. As a final consideration, different degrees of empathy were observed in the dialogues between designers and ChatGPT. These ranged from warm and highly empathic to cold, impersonal, and unfeeling. Figures 8 and 9 contain excerpts of two such dialogues (which were originally in Italian and have been translated into English using ChatGPT). The excerpt in Figure 8 shows a generic formulation of the problem to be solved (with many degrees of freedom open). Moreover, the designer's interaction was quite independent of the ChatGPT answers; it seems that the designer had a list of questions to ask and simply proceeded through them. Finally, the dialogue is mainly unfeeling, cold, and impersonal.

The excerpt in Figure 9 contains a detailed description of the problem to be solved. Moreover, the dialogue is ChatGPT-driven (the designer's questions occur based on ChatGPT feedback) and is highly empathic on both sides.

Designer
I need to generate concepts for a device whose main function is to keep the wooden pencil constantly sharpened.

ChatGPT
Here are some concepts for a device that keeps the wooden pencil constantly sharpened:
Automatic sharpener: A compact device that automatically sharpens the pencil when inserted. It could have an internal sharpening mechanism that rotates the pencil while sharpening, ensuring a sharp tip every time.
Self-sharpening pencil: A special type of pencil that has a carbon or a hard, wear-resistant material tip. This material keeps the tip sharp for a longer time compared to a traditional wooden pencil. You could design a device that holds the pencil vertically with the tip facing downwards, so it gradually self-sharpens while being used.
Electric sharpening system: A device that uses a power-driven rotating blade to sharpen the pencil. It could have a compartment to insert the pencil and a button to start the sharpening process. This system could be battery-powered or connected to an electrical outlet.
Ultrasonic sharpening: A more advanced concept could involve the use of an ultrasonic unit to sharpen the pencil. Ultrasonic waves could be used to remove the excess outer layer of wood and maintain a sharp tip. This device would require a base or a holder to insert the pencil and an integrated ultrasonic unit.
Rotational sharpening system: A device that uses a rotating mechanism to sharpen the pencil. It could have a crank or a button to initiate the sharpening process. You could design a system where the pencil is inserted into a hole, and the device rotates the pencil while a blade sharpens it.
Remember, these are just concepts and may require further development and refinement to make them functional and safe.

Designer
The pencil is the traditional wooden one.

ChatGPT
I apologize for my previous confusion. Here are some specific concepts for keeping a traditional wooden pencil sharp:
Sharpening pencil holder: A special pencil holder with an integrated sharpening blade. This pencil holder may have a slot or a hole to insert the pencil and a blade that automatically sharpens it when rotated or pushed inside the holder. It could be designed to be portable and easy to use.
Sharpening case: A protective pencil case that incorporates a sharpening mechanism. This case may have a small sharpener built into the top or bottom so that when the pencil is inserted or removed, the blade sharpens the tip.   This concept provides a compact and convenient solution for keeping the pencil always sharp.
…

Designer
Concepts to identify the need to sharpen a pencil

ChatGPT
…

Designer
Concepts to get rid of pencil shavings.

ChatGPT
…

Figure 8. Excerpt of a designer–ChatGPT dialogue (generic description of the problem to be solved; independence from ChatGPT answers; unfeeling, cold, and impersonal).

Designer

I need to design a pencil sharpener for a traditional pencil (wooden stick with a graphite core). I need to generate concepts for each function of the pencil sharpener. There are 5 functions: detecting the need for sharpening the pencil, obtaining energy for sharpening, removing excess wood and graphite, deciding when to stop sharpening, and disposing of the shavings. Could you generate some concepts for these functions?

ChatGPT

Certainly! Here are some concepts for each of the requested functions for the pencil sharpener for traditional pencils:

1. Detection of the need for sharpening:
- Visual indicators: an indicator on the pencil sharpener that changes color when the pencil needs to be sharpened.
- Pressure sensor: a built-in sensor in the pencil sharpener that detects the pressure exerted by the pencil to determine if it needs sharpening.
- Weight system: the pencil sharpener is equipped with a small scale that weighs the pencil and signals when it needs sharpening.

2. Obtaining energy for sharpening:
- Crank mechanism: a crank mechanism on the pencil sharpener that requires manual action to generate the energy needed for sharpening.
- Solar charging: the pencil sharpener is equipped with a solar panel that absorbs solar energy to power the sharpening process.
- Rechargeable battery: the pencil sharpener has a built-in battery that can be recharged via a USB port or another power source.

3. Removal of excess wood and graphite:
- Adjustable blades: the pencil sharpener is equipped with adjustable blades that allow adjusting the amount of wood and graphite removed during sharpening.
- Vacuum system: an integrated vacuum system that captures wood and graphite shavings during the sharpening process.
- Built-in brush: the pencil sharpener has a small built-in brush to easily remove excess wood and graphite after sharpening.

4. Decision on when to stop sharpening:
- Temperature sensor: a temperature sensor in the pencil sharpener that detects the heat generated during the sharpening process and signals when the pencil is adequately sharpened.
- Resistance sensor: a resistance sensor that measures the electrical resistance between the pencil and the pencil sharpener to determine when it is sufficiently sharpened.
- Programmable timer: the pencil sharpener has a programmable timer that allows setting the desired duration of the sharpening process.

5. Disposal of shavings:
- Collection container: the pencil sharpener is equipped with a built-in collection container that gathers wood and graphite shavings for easy disposal.
- Ejection system: an automatic ejection mechanism that pushes wood and graphite shavings outward from the pencil sharpener when opened.
- Shavings-free design: an innovative pencil sharpener that cuts the pencil without producing shavings, reducing the need for disposal.

These are just a few examples of concepts for the different functions of the pencil sharpener for traditional pencils. I hope they are helpful as a starting point for developing your project!

Designer

Apologies for the confusion. For the third function, which is sharpening the pencil, here are some additional concepts:

1. Automated angle adjustment: The pencil sharpener automatically adjusts the sharpening angle based on the pencil type or user preference, ensuring optimal sharpening for different pencils.
2. Guided sharpening slot: The pencil sharpener has a guided slot that holds the pencil in place and guides it for a consistent and precise sharpening result.
3. Electric motor: The sharpener is equipped with an electric motor that rotates the pencil automatically, making the sharpening process effortless and efficient.
4. Dual-stage sharpening: The sharpener incorporates a two-stage sharpening mechanism, where the first stage shapes the pencil tip, and the second stage fine-tunes it for a smooth and precise point.

Regarding the concept I mentioned that doesn't produce shavings, here's an explanation:

Shavings-free design: This concept involves a pencil sharpening mechanism that doesn't produce wood or graphite shavings during the sharpening process. Instead of removing material from the pencil, it uses a different method such as laser technology or a fine abrasive surface to gradually shape and refine the pencil tip without creating shavings. This concept aims to minimize mess and waste, providing a clean and efficient sharpening experience.

I hope these additional concepts and the explanation clarify the ideas for you!

ChatGPT

…

Designer

…

Figure 9. Excerpt of another designer–ChatGPT dialogue (detailed description of the problem to be solved; dependence on ChatGPT answers; warm, very empathic).

6. Conclusions

The research described in this paper suggests a framework for measuring the impact of ChatGPT on the generation of innovative concepts in product design. Some hypotheses were posed regarding the characteristics of the concepts, and the dataset collected during a design experience involving 18 university students allowed for drawing some conclusions about them. On the one hand, the impact of ChatGPT is clear in terms of the number of concepts suggested. On the other hand, there are both positive and negative aspects to its effectiveness from the perspectives of usefulness, novelty, and variety. ChatGPT proved to be not so helpful in suggesting useful concepts; classic design methods appeared more effective. However, from the novelty perspective, ChatGPT performed quite well, contrary to expectations since its knowledge base contains only pieces of information regarding things that have already happened (and been documented) in the past. Finally, regarding variety, ChatGPT involvement in design does not seem to make a big difference. However, it emerged that there could be a kind of performance leveling among designers who use it, regardless of their individual characteristics. This may be due to the fact that the knowledge base is always the same. Nevertheless, apart from the evaluation of the metrics involved, some interesting considerations emerged during the data analysis. For example, there are analogies between TRIZ and ChatGPT and the use of them in design. It is also clear that the usefulness metric plays an important role in evaluating ChatGPT performance in design.

Regarding some research perspectives, the use of the specific four metrics makes the exploitation of previous research about the influence of personality on design activities easier, aiming at linking personality to ChatGPT use, always in the product design domain. Moreover, the availability of the logs of the conversations with ChatGPT suggests exploring the possibility of highlighting one more relationship among personality traits, design performance, and use of ChatGPT. Some metrics are under study to cover this aspect as well. These include the query language, the number of questions, and the percentage of independent questions. Another hint for future work is that the low number of participants in the design experience does not allow the research findings to be considered as definitive. In the future, new design experiences will be conducted to make the dataset richer. More-

over, the Microsoft Excel workbook, already optimized from the usability point of view and made available to anyone wanting to carry out personal reasoning and evaluation, will be further improved by making it able to send, with the consent of the researcher, the results of these personal activities to a common repository in order to make the research results as reliable as possible. Finally, particular attention should be placed on comparing TRIZ and ChatGPT use in different design/redesign situations. This comparison was not possible in this study due to the imposed limitations on TRIZ methods and tools (only 40 principles) and the amount of data available. Ad-hoc research could be conducted on this topic in the near future.

References

1. Cooper, R.G. Perspective: The Stage-Gate® Idea-to-Launch Process—Update, What's New, and NexGen Systems. *J. Prod. Innov. Man.* **2008**, *25*, 213–232. [CrossRef]
2. Liu, Y.-C.; Chakrabarti, A.; Bligh, T. Towards an 'ideal' approach for concept generation. *Des. Stud.* **2003**, *24*, 341–355. [CrossRef]
3. Ullman, D.G. *The Mechanical Design Process*, 4th ed.; McGraw-Hill Series in Mechanical Engineering; McGraw-Hill Higher Education: Boston, MA, USA, 2010.
4. Chulvi, V.; González-Cruz, M.C.; Mulet, E.; Aguilar-Zambrano, J. Influence of the type of idea-generation method on the creativity of solutions. *Res. Eng. Des.* **2013**, *24*, 33–41. [CrossRef]
5. Füller, J.; Hutter, K.; Wahl, J.; Bilgram, V.; Tekic, Z. How AI revolutionizes innovation management—Perceptions and implementation preferences of AI-based innovators. *Technol. Forecast. Soc. Chang.* **2022**, *178*, 121598. [CrossRef]
6. Ram, B.; Verma, P. Artificial intelligence AI-based Chatbot study of ChatGPT, Google AI Bard and Baidu AI. *World J. Adv. Eng. Technol. Sci.* **2023**, *8*, 258–261. [CrossRef]
7. Vargas, S.; Castells, P. Rank and relevance in novelty and diversity metrics for recommender systems. In Proceedings of the Fifth ACM Conference on Recommender Systems. Presented at the RecSys '11: Fifth ACM Conference on Recommender Systems, ACM, Chicago, IL, USA, 23–27 October 2011; pp. 109–116. [CrossRef]
8. Goldenberg, O.; Wiley, J. Individual and Group Brainstorming: Does the Question Matter? *Creat. Res. J.* **2019**, *31*, 261–271. [CrossRef]
9. Novak, J.D.; Gowin, D.B.; Kahle, J.B. *Learning How to Learn*, 1st ed.; Cambridge University Press: Cambridge, UK, 1984. [CrossRef]
10. Cross, N. *Design Thinking: Understanding How Designers Think and Work*; Berg: Oxford, UK; New York, NY, USA, 2011.
11. Malycha, C.P.; Maier, G.W. The Random-Map Technique: Enhancing Mind-Mapping with a Conceptual Combination Technique to Foster Creative Potential. *Creat. Res. J.* **2017**, *29*, 114–124. [CrossRef]
12. Brown, T.; Katz, B. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*, 1st ed.; Harper Business: New York, NY, USA, 2009.
13. Casakin, H.P. Assessing the Use of Metaphors in the Design Process. *Environ. Plan. B: Plan. Des.* **2006**, *33*, 253–268. [CrossRef]
14. Al'tšuller, G.S.; Shulyak, L.; Rodman, S. *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*, 2nd ed.; Technical Innovation Center: Worcester, MA, USA, 2007.
15. Liu, Z.; Feng, J.; Wang, J. Resource-Constrained Innovation Method for Sustainability: Application of Morphological Analysis and TRIZ Inventive Principles. *Sustainability* **2020**, *12*, 917. [CrossRef]
16. Ghane, M.; Ang, M.C.; Cavallucci, D.; Kadir, R.A.; Ng, K.W.; Sorooshian, S. TRIZ trend of engineering system evolution: A review on applications, benefits, challenges and enhancement with computer-aided aspects. *Comput. Ind. Eng.* **2022**, *174*, 108833. [CrossRef]
17. Lu, S.; Guo, Y.; Huang, W.; Shen, M. Product Form Evolutionary Design Integrated with TRIZ Contradiction Matrix. *Math. Probl. Eng.* **2022**, *2022*, 3844324. [CrossRef]

18. Edward, C.; Labadin, J.; Kulathuramaiyer, N. Mathematical Modelling and Formalization of TRIZ: Trimming for Product Design. In *Systematic Innovation Partnerships with Artificial Intelligence and Information Technology, IFIP Advances in Information and Communication Technology*; Nowak, R., Chrząszcz, J., Brad, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 3–16. [CrossRef]

19. Rico Sesé, J. Nuevos Retos para el Diseño y la Comunicación. La Inteligencia Artificial en Los Procesos Creativos del Diseño GráFico. Ph.D. Thesis, Universidad Politécnica de València, Valencia, Spain, 2023.

20. De Peuter, S.; Oulasvirta, A.; Kaski, S. Toward AI assistants that let designers design. *AI Mag.* **2023**, *44*, 85–96. [CrossRef]

21. Cantamessa, M.; Montagna, F.; Altavilla, S.; Casagrande-Seretti, A. Data-driven design: The new challenges of digitalization on product design and development. *Des. Sci.* **2020**, *6*, e27. [CrossRef]

22. Siemon, D.; Strohmann, T.; Michalke, S. Creative Potential through Artificial Intelligence: Recommendations for Improving Corporate and Entrepreneurial Innovation Activities. *CAIS* **2022**, *50*, 241–260. [CrossRef]

23. Filippi, S.; Barattin, D. Influence of Personality on Shape-Based Design Activities. *Adv. Hum. -Comput. Interact.* **2019**, *2019*, 9651369. [CrossRef]

24. Sarkar, P.; Chakrabarti, A. Assessing design creativity. *Des. Stud.* **2011**, *32*, 348–383. [CrossRef]

25. Goldberg, L.R. An alternative "description of personality": The Big-Five factor structure. *J. Personal. Soc. Psychol.* **1990**, *59*, 1216–1229. [CrossRef] [PubMed]

26. Sung, S.Y.; Choi, J.N. Do Big Five Personality Factors Affect Individual Creativity? the Moderating Role of Extrinsic Motivation. *Soc. Behav. Pers.* **2009**, *37*, 941–956. [CrossRef]

27. Katila, R.; Ahuja, G. Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction. *Acad. Manag. J.* **2002**, *45*, 1183–1194. [CrossRef]

28. Jansson, D.G.; Smith, S.M. Design fixation. *Des. Stud.* **1991**, *12*, 3–11. [CrossRef]

29. Kohn, N.W.; Smith, S.M. Collaborative fixation: Effects of others' ideas on brainstorming. *Appl. Cognit. Psychol.* **2011**, *25*, 359–371. [CrossRef]

30. Gustafsson, D. Analysing the Double Diamond Design Process through Research & Implementation. Available online: https://aaltodoc.aalto.fi/handle/123456789/39285 (accessed on 18 June 2023).

31. Rosch, E.H. Natural categories. *Cogn. Psychol.* **1973**, *4*, 328–350. [CrossRef]

32. Colgate, M.; Alexander, N. Benefits and Barriers of Product Augmentation: Retailers and Financial Services. *J. Mark. Manag.* **2002**, *18*, 105–123. [CrossRef]