Can Generative AI Compete with the Crowd in the First Phase of New Product Development?

A MANCOVA-Based Comparative Analysis of Idea Quality in the Ice Cream Sector

Master's thesis

MSc Business Administration: Innovation and Entrepreneurship

Name: C.A.J. Koenen

Student number: 1129293

Supervisor: Dr. R.A.W. Kok

Second examiner: Prof. Dr. B. Hillebrand

11-08-2025

Word count: 11.623

Abstract

This research compares the quality of ideas generated by humans through crowdsourcing with ideas generated by generative AI (i.e., ChatGPT-4). 200 ice cream flavors (100 crowd-generated and 100 AI-generated) were rated by two independent raters from different ice cream shops on three dimensions of idea quality: novelty, feasibility and customer benefit. Since the agreement between raters (IRR) was low, mean idea quality scores were used for the analysis. A MANCOVA tested the differences between the crowd- and AI-generated ideas, with external motivation (crowd-sample only) and word count as covariates. The research wanted to test three hypotheses: (H1) crowdsourcing leads to higher novelty and (H3) customer benefit, while (H2) gen AI produces more feasibile ideas. The results showed no significant differences in novelty or feasibility between the two groups, but crowd-generated ideas did score significantly higher on customer benefit, meaning that only hypothesis 3 was supported. The findings suggest that ChatGPT can be a fast and efficient tool to generate new and practical ideas, but the crowd remains important when aligning ideas with real customer needs is most important.

Table of contents

Chapter 1: Introduction

In today's competitive business landscape, one of the main challenges that managers face is how to sustain a competitive advantage through continuous innovation (Roberts & Candi, 2024; Tushman & Nadler, 1986). New product development (NPD) plays a key role in this process, enabling organizations to develop a new product or improve an existing one, which is one of "the essential processes for success, survival, and renewal of organizations, particularly for firms in either fast-paced or competitive markets" (Brown & Eisenhardt, 1995, p. 344). Traditionally, marketers, engineers, and designers have driven new product innovations, but there is a quickly growing interest in the use of generative artificial intelligence (gen AI) to speed up the early-stage NPD process (Cooper & McCausland, 2024; Joosten et al., 2024).

AI, which is commonly defined as machines performing cognitive tasks such as learning, reasoning, and problem-solving (Grewal et al., 2021; Hassani et al., 2020), has gained attention across industries for its efficiency and cost reduction (Cooper, 2024). Gen AI builds on this by generating new content based on learned data (Spanjol et al., 2024). It "refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data" (Feuerriegel et al., 2024, p. 111). Since the launch of OpenAI's ChatGPT, the use of AI in NPD has gained more attention, presenting opportunities to enhance the NPD process, increase innovation and reduce costs (Zhang et al., 2021). Research has shown that AI-driven NPD has the potential to more innovative product ideas and designs (Kakatkar et al., 2019; Wang et al., 2020; Zhang et al., 2020) and is increasingly being used across all seven stages of the NPD process (Cooper, 1983; Zhang et al., 2021).

It has been widely accepted by researchers and practitioners that AI has the potential to provide significant benefits for organizations, like increased revenues (Bilgram & Laarmann, 2023; Cooper, 2024; Singla et al., 2024), and enhanced marketing effectiveness and efficiency (Grewal et al., 2021; Spanjol et al., 2024). A 2024 McKinsey Global Survey on AI showed that 65 percent of respondents stated that their organization frequently uses gen AI (Singla et al., 2024). Moreover, it showed that 72 percent is using gen AI in several parts of the business. Another large study by Zhang et al. (2021) showed significantly higher success rates for NPD projects that incorporated gen AI. The improvement rate depended on the stage AI had been used in. Projects that used AI for ideation, design, and development showed a success rate three times higher than those that did not use AI.

Existing research covers the comparison of different methods of idea generation, for example focus groups and online idea competitions (Schweitzer et al., 2012), but little is known about how gen AI performs compared to traditional approaches such as crowdsourcing, especially in enhancing idea quality during the early stages of NPD. This is crucial as idea quality at this first stage greatly influences the ultimate success or failure of the innovation process (Argument et al., 1998; Bhamra, 2004; Poole & Simon, 1997) and determines whether the idea continues to the next phase. Previously, organizations used crowdsourcing as an alternative way to generate new product ideas (Poetz & Schreier, 2012), but with the rise of gen AI they now have another method that automates and speeds up the idea generation process. Most existing literature has either focused on the general use of gen AI in innovation (Cooper, 2024; Kalota, 2024) or on crowdsourcing for idea generation (Poetz & Schreier, 2012). It lacks empirical evidence on the value of using gen AI for innovation success (Zhang et al., 2021). This creates a gap in understanding gen AI's potential to enhance innovation outcomes by generating high quality ideas in the first phase of the NPD process. Given the perceived benefits of gen AI, research is needed to explore whether it can deliver higher quality ideas than crowdsorucing and thus whether its integration into the NPD process would be beneficial.

From a managerial point of view, the decision to integrate gen AI into the NPD process should be based on evidence of its performance compared to alternative methods like crowdsourcing. As the use of gen AI becomes increasingly popular, organizations that fail to integrate AI are at risk of having slower innovation cycles, higher costs and losing market share to their AI driven competitors (Bilgram & Laarmann, 2023; Cooper, 2024; Singla et al., 2024). Managers thus need insight into whether gen AI can generate high quality ideas that contribute to innovation successes, enabling them to determine if and how it should be implemented to create organizational value.

The purpose of this research is to examine whether the use of gen AI can enhance idea generation during the ideation phase of NPD by improving idea quality and to compare its performance with crowdsourced ideas. Therefore, the research question is as follows: *To what extent does generative AI produce higher quality ideas in the first phase of new product development compared to crowdsourcing?* This research focuses on the contribution of gen AI tools, specifically ChatGPT-4, to the quality of ideas by comparing novelty, feasibility and customer benefit. The findings contribute to academic literature on AI for early-stage innovations and provide managers with evidence on where gen AI may possibly outperform the crowd, helping them decide whether AI should be used in the NPD process.

The next chapter discusses the relevant theories and research on the implementation of gen AI in the idea generation stage of the NPD process. Chapter 3 outlines the methodology used in this research and chapter 4 presents the results. Finally, chapter 5 interprets the results, draws conclusions, and discusses theoretical and practical implications as well as research limitations and recommendations for future research.

Chapter 2: Theoretical Framework

This chapter presents an overview of the relevant literature related to the main concepts of this research. Moreover, it presents the conceptual model, hypotheses along with the research question.

2.1 Ideas and Idea Quality

Ideas can be defined as potential solutions to a problem (Girotra et al., 2010). Organizations favor one 'exceptional' idea and 99 'bad' ideas over 100 'fairly good' ideas (Girotra et al., 2010), implying that quality is preferred over quantity. Idea quality refers to how helpful a generated idea is in achieving a goal (Reinig et al., 2007), i.e., developing a new product. The literature on idea evaluation does not have a set definition for idea parameters and different researchers take on different approaches when it comes to defining and evaluating idea quality.

Early ideation research used quantity as a way to measure quality (Diehl & Stroebe, 1987; Osborn, 1953), however other researchers have found that this relationship is weak (Barki & Pinsonneault, 2001) or even negative (Aiken et al., 1996). Others propose that quality should be assessed based on several criteria rather than quantity alone. Dean et al. (2006) define a quality idea as one that is a feasible solution to a problem, without the idea itself necessarily being novel or out of the ordinary. However, other scholars, such as Poetz and Schreier (2012) argue that novelty is in fact a key dimension of idea quality. They discuss that the quality of an idea can be evaluated based on three key quality dimensions previously identified in research: *novelty*, *feasability*, and *customer benefit* (Poetz & Schreier, 2012). Therefore, a (high) quality idea is defined as one that meets these three criteria. *Novelty* refers to how new the product idea is compared to existing products on the market, *feasibility* refers to how easily the idea could be translated into a commercial product, and *customer benefit* refers to how valuable the idea is in terms of solving the underlying problem or need (Poetz & Schreier, 2012). This definition of idea quality is chosen because it provides a structured and objective way to assess the potential of ideas in the NPD process, and it captures the key success factors of NPD: a new product must be original (novelty) (Chan et al., 2018), possible to develop (feasibility) and meet market needs (customer benefit) (Cooper, 1988; Ernst, 2002).

2.1.1 Stage-Gate Model

In this context, the Stage-Gate model by Cooper (1983) can serve as a structured framework for filtering and progressing only the most promising ideas. While it has largely had a positive influence on the creation, development and launch of new products (Cooper, 2011; Cooper, 2013; Cooper, 2014; Cooper & Edgett, 2012), it has also received some criticisms in an increasingly fast-paced, competitive, and unpredictable world. The model is often seen as overly linear, rigid, and structured, making it less suitable for innovative or dynamic projects. Although critics argue that it lacks adaptability and fails to promote experimentation by applying a 'one-size-fits-all' approach, the Stage-Gate model remains relevant to make sure that only high-quality ideas pass the gate and continue through the structured process. (Cooper, 2017). This highlights its emphasis on refining and selecting the best possible ideas, aligning with the broader organizational preference for quality over quantity in innovation, making the model a valuable theoretical framework in this research.

Since this research focuses on idea quality, the first phase of the Stage-Gate model is particularly relevant. The front end of the Stage-Gate model is essential for NPD performance (Cooper, 1988) as it is the stage in which ideas are generated and advanced into product concepts, ultimately leading to a "go" or "no go" decision on whether to proceed into the formal NPD process (Kijkuit & Van Den Ende, 2007). By analyzing how different types of idea generation (crowdsourcing vs. gen AI) contribute to this phase, this research can provide insight into which of these idea generation methods brings about higher quality ideas for further product development.

2.2 Crowdsourcing

Crowdsourcing is the outsourcing of the idea generation process to the "crowd" of consumers (Poetz & Schreier, 2012). The crowd can consist of, for example, users who are self-selected, willing and able to participate in the idea generation process (Lakhani et al., 2007; Piller & Walcher, 2006). Organizations use crowdsourcing to get direct access to information on consumers' needs and wants to gather concrete ideas for NPD (Schemmann et al., 2016). This definition of crowdsourcing is relevant as it captures the core idea that crowdsourcing involves using external participants outside of the organization.

2.2.1 Who is the Crowd?

The crowd is an "undefined (and generally large) network of people" (Howe, 2006, cited from Geiger et al., 2011, p. 1). There has been debate in the literature who should compose the crowd. On one hand, it is argued that an organization's experts, unlike users, have the necessary skills to create novel and promising ideas that appeal to a larger market and lead to successful products (Ulrich, 2011; Urlich & Eppinger, 2016) and that a creative new product idea "is very often out of the scope of the normal experience of the consumer" (Bennett & Cooper 1981, p. 54). Reasons for this are that users may be too focused on the present and therefore are unable to predict and shape the future (Leonard & Rayport, 1997), so when "relying on the method of asking buyers to describe potential future products, big leaps to novel product ideas are generally not likely" (Schulze & Hoegl, 2008, p. 1744).

On the other hand, there is research that proves that users can come up with good new product ideas (Jeppesen & Frederiksen, 2006). Users often create innovations according to their own needs, and many of these ideas have strong marketable potential (von Hippel, 2006).

The key question discussed by scholars and practitioners is whether crowdsourcing can engage users who create product ideas with broader market appeal (Poetz & Schreier, 2012). There is also research that combines both, such as Poetz & Schreier (2012) whose crowd consisted of both users and professionals who submitted ideas. The self-selective and diverse nature of the crowd raises challenges for crowdsource-based NPD, because finding a combination of different knowledge domains is crucial for identifying and developing desirable product features (Griffin & Hauser, 1992). To improve knowledge sharing across domains, organizations need to identify different expertise within crowds and understand how they can collaborate (Salter et al., 2014).

For the comparison between gen AI and crowdsourcing to be fair, researchers need to consider who the crowd is composed of. Gen AI is not innately an expert or user. If the crowd mainly is composed of experts, their ideas may be significantly different from AI-generated ideas, which could skew the comparison. Therefore, restricting the crowd to solely users may provide a more realistic view, as organizations often look for direct consumer input (Schemmann et al., 2016). This way it can be made sure that the comparison between AI-generated and crowdsourced ideas is fair.

2.2.2 Open Innovation Paradigm

Crowdsourcing can be perceived as part of the Open Innovation paradigm which states that organizations can and should make use of external ideas when innovating (Chesbrough, 2003). It is the internal use of external knowledge, also known as inbound open innovation (Huizingh, 2011). Through observation and analysis of the crowd's communication and discussion of generated ideas, the crowdsourcing organization oversees the ideation process and ultimately determines which ideas will be developed further (Schemmann et al., 2016). The Open Innovation paradigm is chosen as it provides strong theory to explore how external contributions, i.e., ideas from the crowd, influence the quality of innovation and why organizations increasingly make use of crowdsourcing as a strategy to generate new product ideas.

2.3 Generative AI

Gen AI can have multiple forms and applications (Gozalo-Brizuela & Garrido-Merchan, 2023). First, there are text-to-image models, where a text prompt is entered and an image will be produced as output. Conversely, image-to-text models exist in which a text passage is created with a description of an image. Another, more well known type of gen AI are text-to-text models, which can be seen as a question and answer mechanism. A popular example of this is ChatGPT, a large language model (LLM) which interacts in a conversational way (Gozalo-Brizuela & Garrido-Merchan, 2023).

Most ideation and brainstorming research recommends generating a lot of ideas and not yet evaluating or judging them to improve performance (Girotra et al., 2010). While this may be difficult to do for humans, LLM are specifically made to quickly generate many relatively good solutions without much judgment. Therefore, LLMs are expected to be good mechanisms to generate ideas. Cooper (2024, p.4) identified three ways in which AI impacts NPD: "1) idea generation and concept creation and testing; 2) building a robust business case leading to better "go-to-development" investment decisions; and 3) the design, engineering, development, and testing of the product". For these reasons, this research uses ChatGPT-4 as the gen AI method for idea generation.

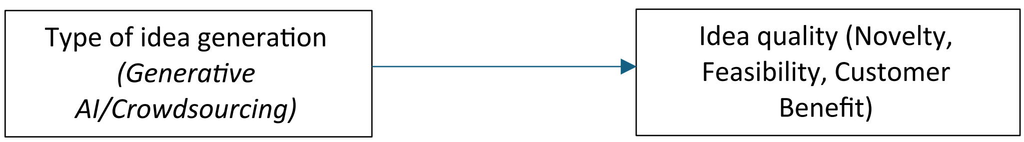
Moreover, like crowdsourcing, gen AI can be considered part of the Open Innovation paradigm, which emphasizes leveraging external knowledge to improve innovation processes.

2.4 Conceptual Model and Hypotheses

This study investigates how different idea generation methods (gen AI vs. crowdsourcing) affect the three key dimensions of idea quality (novelty, feasibility, and customer benefit). Given the possible influence of external motivation to participate in idea generation, it is considered a control variable. Similarly, word count is included as a control variable, as longer ideas may unintentionally get higher quality ratings regardless of their content. The relationship between the idea generation types and idea quality dimensions are shown in the conceptual model in Figure 1.

Figure 1

Conceptual model



2.4.1 Impact of idea generation type on novelty

The literature on co-creation established that user ideas tend to be more novel than those that are generated internally (Magnusson et al., 2003; Poetz & Schreier, 2012). This is because users are less constrained by existing product designs and solutions and therefore create ideas that are more novel and meaningful (Kristensson et al., 2008). Moreover, researchers have found that since gen AI uses existing data to generate ideas, it is expected not to generate highly novel ideas (Filippi, 2023). Hence, it is worth investigating if crowdsourced ideas are also more novel than AI generated ideas. Therefore, the following hypothesis is proposed:

H1: Crowdsourcing produces more novel ideas than Generative AI.

2.4.2 Impact of idea generation type on feasibility

Chesbrough's (2003) Open Innovation paradigm emphasizes a dual approach: creating value and capturing value, respectively. As mentioned earlier, research has shown that users are less constrained by existing product designs and solutions (Kristensson et al., 2008). So, on one hand, crowdsourced ideas may be more novel and useful to satisfy unmet customer needs, and to expand to new customer markets (Hoyer et al., 2010). But on the other hand, users often do not have a deep enough understanding of an organization's capabilities to implement such ideas (Poetz & Schreier, 2012). So, user generated new product ideas may not be feasible to implement or need significant adjustment for them to be profitable for the organization (Shaner, 2015). Gen AI may, in this case, be more capable of creating ideas that are feasible as it considers the organization's owned competencies, technologies, and resources (Joosten et al., 2024). Therefore, the following hypothesis is proposed:

H2: Generative AI produces more feasible ideas than crowdsourcing.

2.4.3 Impact of idea generation type on customer benefit

Crowdsourcing communities often consist of people who use the organization's products and therefore know their own needs and wants better than anyone else. When the crowd consists of these individuals, this 'wisdom of the crowd' is leading in understanding how a new product or service can create value, and thus customer benefit, for them (Magnusson et al., 2016; Surowiecki, 2004). Moreover, the crowd tends to assess ideas out of self-interest (e.g., feasibility and desirability) (Rietzschel et al., 2010), favoring product ideas that meet their needs and wants. Whereas gen AI may not have the ability to comprehend nuanced social and cultural aspects, which may be fundamental when it comes to generating valuable product ideas for consumers (Amabile, 2018). Therefore, the following hypothesis is proposed:

H3: Crowdsourcing produces ideas with higher customer benefit than Generative AI.

Chapter 3: Methodology

This chapter outlines the methodology. It starts with the research design, sample, and data collection procedure. Then, it describes the measures and evaluation process. Lastly, the statistical analyses and assumption testing are explained in detail.

3.1 Research Design

This research adopted a quantitative, quasi-experimental design to compare the quality of ideas generated by two different sources: crowdsourcing and gen AI (i.e., ChatGPT-4.0). While participants were not randomly assigned to conditions, the design still allowed for a structured comparison between both types of idea generation. Since the conditions were kept the same (the number of ideas per group was equal, the same prompt was used and all ideas were evaluated using the same criteria by the same independent raters), any differences in idea quality could be attributed to the type of idea generation.

Experimental designs are useful to research causal relationships between factors (Gavin, 2008). They allow researchers to study the effects of manipulable things, in this case the type of idea generation, under controlled conditions (Shadish et al., 2002). While these designs often rely on predefined evaluation criteria to ensure consistency (Girotra et al., 2010), this research has some limitations due to its use of a survey for crowdsourced data, a prompt-based method for gen AI and human raters to evaluate the idea quality.

A qualitative approach, e.g., in-depth interviews or focus groups, could provide rich, nuanced insights (Fossey et al., 2002) into the idea generation process. However, it would not provide the large-scale data needed to compare the two methods systematically and therefore was not used.

This research adopts a positivist epistemological standpoint, which assumes that knowledge can be objectively measured and observed (Park et al., 2020). By using measurable data and statistical analyses in SPSS (version 30), this research produced reliable and replicable findings. However, it is important to note that the research design introduced some limitations to objectivity. The crowdsourced data was collected using a survey and the AI-generated idea was created using a structured prompt that mimicked a survey-like approach. Additionally, idea quality was evaluated by human raters which introduces some subjectivity. While controlling as much as possible, these factors may have influenced the outcomes and should be considered when interpreting the findings. In addition, while gen AI relies on measurable and observable data, its output may not be fully objective as it is shaped by data it has been trained on. Therefore, while the study is grounded in positivist principles, it also included post-positivist elements, recognizing that full objectivity is difficult to achieve when working with both human judgment and AI-generated data.

3.2 Data Collection and Sample

3.2.1 Idea generation

The design is comparative, with two separate sets of ideas generated. The first set consisted of 100 ideas submitted by the crowd (Appendix A) through a Qualtrics survey. The survey invited people to participate in a hypothetical ideation contest to come up with one new ice cream flavor. Participants were instructed to write approximately 20 words when describing their ice cream flavor idea to prevent excessively long responses. A word limit was set in the survey, but idea length still varied across participants. This variation was later controlled for in the MANCOVA by adding ‘word count’ as a covariate. The full survey can be found in Appendix B. The second set included 100 ideas generated by ChatGPT (Appendix C). To ensure a fair comparison, the same question from the survey was used as a prompt. ChatGPT was manually prompted 100 times, once for each idea, to replicate the ‘one response per person’ logic used in the crowdsourcing sample and to ensure variation and independence. This approach was chosen over asking for 100 ideas in a single prompt. This was also tested but resulted in less variation, since ideas grouped into similar themes and influenced each other. The full prompt and follow-up prompts are included in Appendix D. Both sets of ideas were anonymized, merged into one document and evaluated by the external raters.

3.2.2 Idea evaluation

The evaluation of idea quality was done by two independent raters: one ice cream shop owner (IJssalon Kees) and one Chief Operating Officer (de IJswinckel). Both experts independently assessed all 200 anonymized ideas on the three idea quality dimensions. To minimize bias, they were not told whether the ideas were generated by humans or AI. Each idea was rated on a 5-point Likert scale. The complete list of rating items can be found in Appendix E.

3.2.3 Measures

To operationalize idea quality, the dimensions were measured using a survey with items adapted from prior research. Items include: “the new product idea is original and uncommon” (novelty), “the new product idea is both feasible and marketable” (feasibility), and “the new product idea is considered suitable for customers’ desires” (customer benefit). For each dimension only one item was used. This choice was made to reduce the time needed for evaluation and to make the tasks easier for the raters as they had to evaluate 200 ideas, resulting in 600 ratings per rater in total. An overview of the constructs, dimensions, and items can be found in Table 1. All items were measured on a 5-point Likert scale. The type of idea generation (gen AI = 0, crowd = 1) serves as the independent variable (IV), idea quality (novelty, feasibility, and customer benefit) as the dependent variables (DVs). External motivation (e.g., rewards for motivation) is included as a covariate since previous research suggests it may positively influence creative output (Eisenberger & Rhoades, 2001; Eisenberger & Shanock, 2003). Additionally, word count was included as a control variable to account for the potential effect of idea length on how ideas were rated. Prior research has shown that longer ideas may be perceived as more elaborate or complete (Zeng et al., 2022). The items concerning external motivation were answered by respondents in the crowdsourcing survey.

Table 1

Operationalization of Measures

Constructs	Dimensions	Items	Source
Idea generation (dichotomous IV)	Gen AI vs. Crowd	Binary: Gen AI/ChatGPT = 0, Crowd = 1	-
Idea quality (DV)	Novelty	The new product idea is original and uncommon	Kudrowitz & Wallace (2013), Poetz & Schreier (2012)
	Feasibility	The new product idea is both feasible and marketable	Kijkuit & Van Den Ende (2007), Kudrowitz & Wallace (2013)
	Customer benefit	The new product idea is suitable for customers' desires	Im & Workman (2004)
External motivation to participate (covariate)	Monetary rewards	I am more likely to contribute ideas if there is a (financial) reward	Self-developed, based on Amabile et al. (1994), Bretschneider et al. (2012)
	Non-monetary rewards	I would be more interested if the winning flavor were actually produced	Self-developed, based on Amabile et al. (1994), Bretschneider et al. (2012)
	No rewards	I enjoy coming up with ideas regardless of rewards	Self-developed, based on Amabile et al. (1994), Bretschneider et al. (2012)
		I would feel more pressure if it were a real contest	Self-developed, based on Amabile et al. (1994), Bretschneider et al. (2012)
Additional control variable	Word Count	Number of words used to describe each idea	-

3.2.4 Inter-rater reliability

To evaluate the consistency of the ratings provided by the two independent raters, an inter-rater reliability test (IRR; for full SPSS output see Appendix F) was conducted using the Intraclass Correlation Coefficient (ICC), which is the degree of agreement or consistency between two or more raters evaluating the same set of items, ranging from 0 to 1 (Perinetti, 2018). Following Cicchetti's (1994) guidelines: ICC below 0.40 is poor, ICC between 0.4 and 0.59 is fair, ICC between 0.60 and 0.74 is fair and ICC equal or greater than 0.75 is excellent. The ICC was calculated using a two-way random effects model with absolute agreement and single measurement, i.e. ICC (2,1) (Perinetti, 2018). This model was selected because each idea was rated by the same two independent raters who are considered representative of a larger pool of potential raters. The focus was on absolute agreement rather than consistency as the goal was to determine the degree to which both raters gave the same or similar scores to the same ideas in terms of idea quality.

For each dimension, the ICC value was calculated together with a 95% confidence interval, the F-statistic and the p-value. Table 2 presents the ICC values, confidence intervals, F-statistics and p-values for all three idea quality dimensions.

Table 2
Inter-rater Reliability Test (ICC per Dimension)

Dimension	ICC (2,1)	95% CI	F (199, 199)	p	Mean score (combined)
Novelty	-0.015	[-0.153, 0.124]	0.971	0.583	3.60
Customer benefit	0.205	[0.069, 0.334]	1.515	0.002*	3.85
Feasibility	0.043	[-0.096, 0.180]	1.090	0.272	3.95

*$p < 0.05$.

IRR for novelty was low with *ICC* (2,1) = -0.015, 95% CI [-0.153, 0.124], $F(199,199) = 0.971$, $p = 0.583$, indicating no significant agreement between the raters. IRR for customer benefit was low with *ICC* (2,1) = 0.205, 95% CI [0.069, 0.334], $F(199, 199) = 1.515$, $p = .002$, indicating a statistically significant but poor agreement between the two raters. IRR for feasibility scores was low with *ICC* (2,1) = .043, 95% CI [-.096, .180], $F(199, 199) = 1.090$, $p = .272$, indicating no significant agreement between the two raters.

These low ICC scores show that the raters often gave different scores to the same ideas. For example, *de IJswinckel* scored *Flavor 38* as 1 for novelty, 1 for customer benefit and 1 for feasibility, while *IJssalon Kees* scored this flavor as 4 for novelty, 5 for customer benefit and 4 for feasibility. Therefore, the decision was made to use the mean score (Table 2) for all 200 ideas per dimension, based on the combined ratings of both raters for all further analyses to ensure consistency. Thus, the mean score per idea does not mean the raters agreed but it shows the middle value between their judgements to provide a single value for consistent comparison.

3.3 Data Analysis

3.3.1 MANCOVA

A one-way between-groups multivariance analysis of covariance (MANCOVA) was performed to test the effect of idea generation type on idea quality. This approach is appropriate as it allows for an investigation of differences between groups while statistically controlling for an additional variable (covariate) (Pallant, 2016). The analysis is one-way, as there is one categorical IV, *Type of Idea Generation*, consisting of two levels: *Generative AI* and *Crowdsourcing*. The analysis was conducted as a *between-group design*, comparing the two groups (gen AI vs. crowdsourcing) on the three dimensions of idea quality (novelty, feasibility, and customer benefit) which are treated as separate but related DVs in the MANCOVA model. The sum of the average scores of these three dimensions gives us the total score on idea quality. The idea quality scores of crowdsourced ideas are then compared to AI-generated ideas to evaluate which of the two results in the highest quality ideas.

Wilk's lambda was used as the test statistic (with corresponding F-statistics and p-values) to test the multivariate effect as it is the most recommended statistic to use to determine whether there are significant differences between groups on a combination of multiple DVs (Laerd Statistics, n.d.).

3.3.2 Assumption Tests

To get meaningful results, the data must meet several assumptions (Laerd Statistics, n.d.; Pallant, 2016). The DVs and covariate must be continuous (interval or ratio level) and the IV must be categorical with two or more unordered groups. This is satisfied as the IV has two categorical groups without order and both the DVs and covariate use Likert scales, providing at least interval level data. Moreover, observations should be independent within and between groups. For gen AI, each idea was independently generated by ChatGPT using predefined prompts ensuring full independence of observations. For the crowdsourced data, full independence of observations cannot be guaranteed due to the nature of crowdsourcing, which is further discussed in the research limitations section.

Preliminary assumptions tests (see Appendix G) were performed to check for normality, linearity, outliers, homogeneity of variance covariance matrices and multicollinearity. The data showed no serious violations of linearity, homogeneity of variance (Levene's test), homogeneity of covariance (Box's M test), or multicollinearity (VIF < 5, tolerance > 0.10). Although normality tests (Kolmogorov-Smirnov and Shapiro-Wilk) were significant ($p < 0.001$), indicating a deviation from normality, the histograms showed reasonably normal distributions without extreme outliers. Because MANCOVA is robust to small deviations and both groups had a relatively large sample size of $n = 100$, the assumption of normality was still considered acceptable. Since most assumptions were met and deviations were minimal, it was considered appropriate to proceed with the MANCOVA.

3.4 Research Ethics

The ethics principles of the American Psychological Association were considered while conducting this research (Smith, 2003). First, all individuals had full informed consent meaning that they voluntarily participated in the research and had full knowledge of the (dis)advantages involved. Moreover, confidentiality and privacy were guaranteed, and the gathered data was anonymized and stored safely in an online folder provided by Radboud University, only accessible to the researcher, the supervisor and second examiner.

3.5 Methodological Limitations

Since the research compared crowdsourced data with gen AI responses, a potential selection bias existed which could have affected validity and generalizability. The crowdsourced data was collected through snowball sampling to increase response rates. However, limitations of this method are a lack of control over who make up the sample and thus limited representativeness (Ting et al., 2025).

For the AI generated data, each idea was produced independently by ChatGPT with a predefined prompt, ensuring independence of observations. However, the crowdsourced data was collected through a survey and therefore full independence cannot be guaranteed as participants can contribute multiple ideas. This could introduce variability into the data and may reduce the reliability of the crowdsourced results.

Chapter 4: Results

This chapter presents the results of the statistical analyses performed to examine the differences in idea quality between the crowd- and AI-generated ideas. First, descriptive statistics are presented, followed by the results of a Multivariate Analysis of Variance (MANOVA) and separate Analyses of Variance (ANOVAs). Then a MANCOVA is reported to control for word count, followed by separate Analyses of Covariance (ANCOVAs) to explore individual effects more closely. Lastly a post-hoc analysis is included to test the influence of external motivation on idea quality for the crowd sample only. Statistical significance was assessed using an alpha level of 0.05 unless otherwise specified. Results with p-values below 0.05 were considered statistically significant.

4.1 Descriptive Statistics

Table 3

Descriptive Statistics for Idea Quality Dimensions by Idea Generation Type

Idea Type	Novelty (M)	SD	Customer benefit (M)	SD	Feasibility (M)	SD
ChatGPT	3.55	1.00	3.81	1.17	3.76	0.97
Crowd	3.65	0.97	4.09	1.13	3.93	0.95

The main descriptive statistics of the data can be found in Table 3 (for full SPSS output see Appendix H). Means are based on the average ratings from the two independent judges for each idea. As shown in Table 3, crowd-generated ideas scored somewhat higher than AI-generated ideas on all three dimensions of idea quality. The largest difference was found in customer benefit, where crowd ideas had a higher mean score (M = 4.09, SD = 1.13) compared to ChatGPT ideas (M = 3.81, SD = 1.17). Differences in novelty and feasibility were smaller but also favored the crowdsourced ideas. To test whether these observed differences were statistically significant, a MANOVA followed by separate ANOVAs was conducted, of which the results are reported in section 4.2 (Table 5). Only the difference in Mean Customer Benefit between ChatGPT and the crowd was statistically significant (p = 0.011), while Mean Novelty (p = 0.032) and Mean Feasibility (p = 0.065) were not.

4.2 MANOVA without Covariates

A one-way MANOVA was conducted to examine the effect of idea generation type on perceived idea quality across the three idea quality dimensions. The multivariate test was not statistically significant (*Wilks' Λ* = .968**,** F (3, 196) = 2.18**,** p = .091**,** *partial* η^2 = .032) (see Table 5; for full SPSS output see Appendix I)**,** indicating no overall difference in idea quality based on generation type.

4.2.1 Correlation analysis between dependent variables

Before examining each of the ANOVAs and looking at each outcome separately, the relationship between the three DVs was checked using Pearson correlations (Table 4; for full SPSS output see Appendix J).

Table 4

Correlations between Mean Novelty, Mean Feasibility and Mean Customer Benefit

	Mean Novelty	Mean Feasibility	Mean Customer benefit
Mean Novelty	-	0.385**	0.384**
Mean Feasibility	0.385**	-	0.756**
Mean Customer benefit	0.384**	0.756**	-

**p < 0.01 (2-tailed)

All correlations were statistically significant ($p < 0.01$). A Pearson's correlation coefficient below 0.10 indicates a negligible correlation, between 0.10 and 0.39 indicates a weak correlation, 0.40 and 0.69 a moderate correlation, 0.70-0.89 a strong correlation, and above 0.90 a very strong correlation (Schober et al., 2018). Novelty was weakly (but almost moderately) correlated with feasibility ($r = 0.385$) and customer benefit ($r = 0.384$), while feasibility and customer benefit were strongly correlated ($r = 0.756$). This shows that the three DVs are related but still different enough to look at them separately. Consequently, in line with Huberty and Morris (1989), additional insight was looked for by performing separate ANOVAs for each DV to explore whether effects that were not found by the overall MANOVA might show at the individual dimension level, especially when the DVs are only moderately correlated as was the case in this analysis. These individual dimension level analyses were based on the Tests of Between-Subject Effects table from SPSS and are shown in Table 5.

Table 5

MANOVA and MANCOVA Analyses of Idea Quality Dimensions

Outcome level	Dependent variable	Model	Effect	Overall test (Wilks' Λ)	F (df1, df2)	p-value	Effect size (Partial η²)
Multivariate results	-	MANOVA (no covariate)	Idea Type	0.968	2.18 (3, 196)	0.091	0.032
	-	MANCOVA (word count covariate)	Idea Type	0.968	2.15 (3, 195)	0.095	0.032
Results per dimension	Mean Novelty	MANOVA (no covariate)	Idea Type	-	0.95 (1, 198)	0.332	0.005
		MANCOVA (word count covariate)	Idea Type	-	0.92 (1, 197)	0.340	0.005
		MANCOVA (word count covariate)	Word Count	-	0.69 (1, 197)	0.409	0.003
	Mean Feasibility	MANOVA (no covariate)	Idea Type	-	3.44 (1, 198)	0.065	0.017
		MANCOVA (word count covariate)	Idea Type	-	3.35 (1, 197)	0.069	0.017
		MANCOVA (word count covariate)	Word Count	-	5.12 (1, 197)	0.025*	0.025
	Mean Customer Benefit	MANOVA (no covariate)	Idea Type	-	6.60 (1, 198)	0.011*	0.032
		MANCOVA (word count covariate)	Idea Type	-	6.50 (1, 197)	0.012*	0.032
		MANCOVA (word count covariate)	Word Count		1.85 (1, 197)	0.176	0.009

Note. Each MANCOVA row includes two effects from the same model: the effect of Idea Type (ChatGPT vs. Crowd) and the effect of Word Count as a covariate. These effects are listed seperately for clarity. *$p < 0.05$.

4.2.2 Separate ANOVAs for each dependent variable (Dimension level analyses)

Only customer benefit showed a statistically significant difference between the ChatGPT and crowdsourced ideas (F (1, 198) = 6.60, p = 0.011**,** *partial* η^2 = 0.032). A *partial* η^2 of 0.032 indicates a small effect size according to Cohen's guidelines (2013). Crowd ideas were rated higher than ChatGPT ideas. No significant effects were found for novelty (F (1,198) = 0.95, p = 0.332, *partial* η^2 = 0.005) or feasibility (F (1,198) = 3.44, p = 0.065, *partial* η^2 = 0.017). These results suggest that idea type had a significant impact on customer benefit only, with crowd ideas being seen as more beneficial to customers from the raters' perspectives. This supports hypothesis 3 which stated that crowd-generated ideas would be perceived as offering greater customer benefit than AI-generated ideas. However, hypotheses 1 and 2 predicting that crowd ideas would be more novel and that AI-generated ideas would be more feasible, were not supported since no significant differences were found for novelty (p = 0.332) or feasibility (p = 0.065) (see Table 5).

4.3 MANCOVA (Controlling for word count)

To examine whether idea quality differed between crowdsourced and AI-generated ideas while controlling for word count (idea length) a one-way between-subjects multivariate analysis of covariance (MANCOVA) was conducted. The IV was idea type (ChatGPT vs. Crowd), the covariate was word count and the DVs were mean novelty, mean feasibility and mean customer benefit. The overall multivariate test was not statistically significant (*Wilks'* Λ = .968, F (3, 195) = 2.15, p = .095, *partial* η^2 = .032; see Table 5; for full SPSS output see Appendix K) indicating no overall difference in idea quality between the two groups after adjusting for word count.

4.3.1 Separate ANCOVAs for each dependent variable (Dimension level analyses)

In line with Huberty and Morris (1989), separate ANCOVAS were conducted for each DV even though the multivariate test was not significant. The ANCOVA for customer benefit showed a significant effect of idea type (F (1, 197) = 6.50, p = .012, *partial* η^2 = .032; see table 5) with crowd-generated ideas scoring higher. The *partial* η^2 of 0.032 indicated a small effect. This supports hypothesis 3 which predicted that crowd-generated ideas would score higher on customer benefit. Word count had a small but significant effect on feasibility (F (1, 197) = 5.12, p = .025, *partial* η^2 = .025; see table 5) but it did not significantly influence novelty or customer benefit. These results suggest that even when word count is considered, crowd-generated ideas still scored higher on customer benefit. Word count only had a small effect on how feasible the ideas were seen as (*partial* η^2 = 0.025) and had no effect on novelty or customer benefit, meaning that hypothesis 1 and 2 predicting higher novelty for crowd-generated ideas and higher feasibility for AI-generated ideas, were not supported.

4.4 Post-hoc Analysis (External Motivation)

A post-hoc analysis was performed to examine whether external motivation influenced idea quality within the crowd-generated ideas. Only the crowd sample was included in this analysis as the survey participants answered the motivation related questions. Since motivation does not apply to AI-generated ideas, the AI generated ideas were excluded from this analysis. A one-way MANCOVA showed no significant multivariate effect of external motivation on the combined DVs (novelty, customer benefit and feasibility) (*Wilks' Λ* = 0.965, *F* (3, 96) = 1.16. *p* = 0.327, *partial η²* = .035; see Table 7; for full SPSS output see Appendix L).

Table 7

Effects of External Motivation as Covariate (Crowd Sample Only)

Outcome level	Dependent variable	Covariate	Overall test (Wilks' Λ)	F (df1, df2)	p-value	Partial η²
Multivariate results	-	External motivation	0.965	1.16 (3, 96)	0.327	0.035
Results per dimension	Mean Novelty	External motivation	-	2.68 (1, 98)	0.105	0.027
	Mean Feasibility	External motivation	-	0.32 (1, 98)	0.572	0.003
	Mean Customer Benefit	External motivation	-	1.37 (1, 98)	0.245	0.014

These post-hoc analyses per DV align with the approach by Huberty and Morris (1989) to further analyze potential group differences despite non-significant multivariate results. Individual ANCOVAs revealed that external motivation did not significantly influence novelty (*p* = 0.105), customer benefit (*p* = 0.245) or feasibility (*p* = 0.572). These results suggest that within this sample, the level of external motivation did not significantly impact the quality of ideas submitted by the crowd.

4.5 Top Rated Ideas per Idea Generation Type

An additional descriptive analysis was performed to gain more insight into which idea generation type generated the highest quality ideas. This analysis focused on the top rated ice cream flavors by looking at the total idea quality score (novelty + feasibility + customer benefit). As each of the three idea quality dimensions was rated on a 5-point Likert scale, the highest possible total idea quality score was 15.

The original goal was to create a top 10 list of the highest scoring ice cream flavors. However, this was not possible due to a high number of ties at the top score levels. Based on the mean idea quality score of both raters combined, 14 ideas received the same highest score of 14 out of 15. When looking at the raters individually, *de IJswinckel* assigned the top score of 15 out of 15 to 25 ideas, while *IJssalon Kees'* highest score was 13 out of 15, given to 107 ideas. Therefore, instead of a top 10, the analysis looked at all the ideas that received the highest score for the mean idea quality (both raters combined), and *de IJswinckel* and *IJssalon Kees* individually.

Based on the scores of *de IJswinckel* and *IJssalon Kees* combined, 14 flavors received a score of 14, of which 5 were generated by ChatGPT and 9 by the crowd. Out of the 25 flavors that received a score of 15 from *de IJswinckel*, 9 were generated by ChatGPT and 16 by the crowd. Among the 107 flavors that received a score of 13 from *IJssalon Kees*, 53 were generated by ChatGPT and 54 by the crowd. The full lists of top rated ideas can be found in Appendix M.

These findings suggest that both idea generation types can generate high quality ideas when it comes to ice cream flavors. However, the crowd contributed slightly more to the highest scoring ideas overall, especially when looking at the combined ratings and the ratings from *de IJswinckel*. The ratings of *IJssalon Kees* showed a relatively balanced distribution (53 ChatGPT and 54 crowd). While not statistically tested, this comparison shows that although the crowd seems to slightly outperform gen AI in producing the highest quality ideas, gen AI also showed good potential to produce high quality ideas in early idea generation.

Chapter 5: Discussion and conclusion

This final chapter discusses the findings for each idea quality dimension in relation to existing literature. Their theoretical and practical implications are then summarized, followed by the study's limitations and suggestions for future research. This chapter ends with a summary of the key insights.

5.1 Discussion

Previous research has shown the advantages of using gen AI in various phases of the NPD process (Zhang et al., 2021; Spanjol et al., 2024), but little research has directly compared the performance of gen AI with a traditional ideation method such as crowdsourcing. Therefore, this research aimed to examine and compare the quality of ideas generated through crowdsourcing and gen AI in the early stages of NPD with the goal of answering the following research question: *To what extent does generative AI produce higher quality ideas in the first phase of new product development compared to crowdsourcing?* The results offer insight into how the idea quality of gen AI compares to crowdsourcing, using novelty, feasibility and customer benefit as evaluation dimensions.

5.1.1 Novelty

There was no significant difference in novelty between crowd- and AI-generated ideas. Earlier research found that user ideas are often more novel than internally generated ideas (Kristensson et al., 2008; Magnusson et al, 2003; Poetz & Schreier, 2012) but they did not look at crowdsourced ideas in comparison to AI-generated ideas. In this research, the expected advantage in novelty for crowd-generated ideas was not present. Filippi (2023) pointed out that gen AI relies on existing data and therefore is not expected to create extremely novel ideas, which is confirmed by the study's results. One possible solution for these results is that the task of coming up with a new ice cream flavor was quite simple and therefore both methods performed equally well. Another possible solution is the low IRR. Even though the two raters were given the same item to evaluate novelty, they may have interpreted and judged it differently leading to different results. This subjectivity is visible in the slightly higher mean score on novelty for crowd-generated ideas, but again this difference was not statistically significant.

5.1.2 Feasibility

There was no significant difference in feasibility, although crowd-generated ideas had a slightly higher mean score on feasibility. Contrary to what was expected based on earlier research, which suggested that AI could generate more feasible ideas since it uses patterns from existing data (Joosten et al., 2024). A possible explanation for this could be that feasibility was influenced by the way the ideas were described rather than their actual practicality. This is supported by the small but significant effect (*partial* $\eta^2 = 0.025$) of word count. Previous research showed that longer ideas may be seen as more elaborate or complete (Zeng et al., 2022). Therefore, longer, more detailed responses could have made it easier for the raters to understand how an idea could be implemented, which in turn could have led to higher feasibility scores. However, because the agreement between the two raters was low, these small differences are more likely to reflect subjective interpretation rather than a clear advantage for one of the idea generation methods.

5.1.3 Customer Benefit

Crowd-generated ideas scored significantly higher on customer benefit, although the effect size was small (*partial* $\eta^2 = 0.032$). This is in line with the 'wisdom of the crowd' idea (Surowiecki, 2004) and other research stating that users know and understand their own needs and preferences and thus can generate ideas that match them well (Poetz & Schreier, 2012). The crowd may naturally focus more on customer preferences because they themselves are potential users/consumers. In contrast, gen AI cannot know what customers actually want and can only guess based on patterns in existing data. This could be a reason as to why the AI-generated ideas were rated less suited to customer needs.

5.1.4 Linking back to theory

When looking at these results in the context of the Stage-Gate model, both crowdsourcing and gen AI seem to be able to generate ideas that are novel and feasible enough to pass the first gate. However, the crowd seems to have an advantage when it comes to customer benefit. From an Open Innovation perspective, gen AI can be seen as an extra external source of knowledge that can speed up the idea generation process. However, customer input is still key in making sure that new products are relevant to the market in the early stages of NPD.

5.2 Theoretical Implications

This research adds new evidence to the limited amount of literature comparing gen AI with crowdsourcing in the first phase of the NPD process. Moreover, it adds to literature on AI in NPD and NPD performance (Cooper, 2023; Zhang et al., 2021). The findings suggest that for simple creative tasks such as coming up with a new ice cream flavor, gen AI can perform about the same as the crowd in terms of novelty and feasibility, but the crowd performs better in terms of customer benefit.

The results show that the effect of idea generation type is not the same for all quality dimensions. Idea quality should thus be interpreted as a combination of different dimensions rather than looked at as one overall score. The results also show that the way idea quality is measured can influence the outcomes. The chosen dimensions change how the performance of both AI and the crowd are judged.

By only looking at the first stage of NPD, this study shows that AI does not have an advantage over the crowd at this point in terms of idea quality. However, this adds nuance to gen AI's role in innovation, suggesting that it possibly may be more valuable in other stages of the NPD process rather than ideation, unless the main goals are speed and cost-effectiveness.

5.3 Practical Implications

Because the results for novelty and feasibility were almost the same for both idea generation methods, organizations can use gen AI for speed and variety (Cooper, 2023), and the crowd for ideas that fit customer needs better. Gen AI may also be more cost-efficient than crowdsourcing because it does not require rewards or recruiting and managing participants. This makes gen AI useful when speed, variety and lower costs are important. Organizations can reduce costs by replacing expensive large-scale crowdsourcing with AI, and reserve budget for other stages such as customer testing and validation later on in the NPD process.

The small advantage of the crowd for customer benefit suggests that while customer/user input adds value, the effect size in this research was so small (partial $\eta^2 = 0.032$) that it may be negligible in contexts where time and money are the main priority.

Managers can match the idea generation method to the stage of the NPD process. Using the Stage-Gate model as a reference, gen AI can help to quickly fill the ‘idea pool’, while the crowd can help decide which ideas are most relevant for customers before moving on in the NPD process. A combined approach could also be effective in which gen AI is used to generate a large set of ideas, to then involve the crowd to filter them for a good market fit.

Organizations can also improve the idea evaluation process by involving more raters, or by including customers, to make sure that decisions reflect a wider range of perspectives and are better suited to customer/market needs. For example, internal experts can be used to look at feasibility, while customers look at customer benefit.

5.4 Limitations and Suggestions for Future Research

Several limitations should be considered when interpreting the results. First, the difference in ratings between the two raters, and thus the low IRR, indicates subjectivity in the evaluation of idea quality. Although the mean score per idea was used, it is likely that it reduced the variance but it may also have hidden meaningful differences in how the raters judged certain ideas. If one rater considered an idea to be high quality and the other rated it as low quality, the average score would result in a middle rating which may give the impression that the idea was of average quality while it really showed strong disagreement between the raters. Taking the average score between the raters helped to reduce differences but their perspectives were still not fully aligned. This means that the results should be interpreted with caution since the results may not only reflect differences in judgment but also actual differences in idea quality.

Second, both raters were Dutch, thus non-native English speakers. While they were both given the opportunity to evaluate the ideas in Dutch, neither chose to do so. However, one rater explicitly pointed out that it took him longer to evaluate the ideas in English as he had to translate them to Dutch first. This may have affected how he interpreted and judged some of the ideas, especially those that were more abstract or creatively worded.

Third, the research used ChatGPT-4. Different results might be found when using different AI models, for example Gemini or CoPilot.

Fourth, the ideas were evaluated based on the raters’ judgment and are therefore highly subjective and no real customer perspective was used in the idea quality evaluation.

Fifth, the crowd-generated ideas were collected from a convenience sample which may reduce the generalizability of the findings.

Future research could use more raters to increase the IRR and decrease individual bias. To minimize interpretation issues, evaluations should be done in the raters’ native language. Moreover, to improve generalizability of the study, future research could use multiple gen AI models to compare them and examine whether idea quality differs across different models. In addition, it may be useful to include actual consumers in the evaluation process to gain access to a more market-oriented perspective of idea quality, especially when it comes to the customer benefit dimension. Finally, future research may explore different dimensions of idea quality, as previous research suggests that there are other ways to evaluate idea quality other than novelty, feasibility and customer benefit, for example originality (Kudrowitz & Wallace, 2013).

5.5 Conclusion

In summary, this research compared ideas from gen AI and crowdsourcing in the first phase of NPD. Based on the evaluations from the two raters, gen AI did not produce higher quality ideas than crowdsourcing in the first phase of NPD. Both idea generation types scored approximately the same on novelty and feasibility, while the crowd scored significantly higher on customer benefit. However, as previously mentioned, this advantage for the crowd was small, meaning that it may be less relevant in contexts where speed and money are more important. A small effect was found for word count on feasibility which suggested that more detailed (longer) ideas can be seen as more practical, although this did not change the overall comparison between the two methods. External motivation did not have a significant impact.

Overall, the choice between gen AI and crowdsourcing in early NPD should be based on the organization's priorities. Gen AI can be a fast and efficient way to generate new (novel) and practical (feasible) ideas, while the crowd can be used when matching customer needs (customer benefit) is most important.

References

Aiken, M., Vanjani, M., & Paolillo, J. (1996). A comparison of two electronic idea generation techniques. *Information & Management*, *30*(2), 91-99.

Amabile, T. M. (2018). *Creativity in context: Update to the social psychology of creativity*. Routledge.

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: assessing intrinsic and extrinsic motivational orientations. *Journal of personality and social psychology*, *66*(5), 950.

Argument, L., Lettice, F., & Bhamra, T. (1998). Environmentally conscious design: matching industry requirements with academic research. *Design Studies*, *19*(1), 63-80.

Barki, H., & Pinsonneault, A. (2001). Small group brainstorming and idea quality: Is electronic brainstorming the most effective approach? *Small Group Research*, *32*(2), 158-205.

Bennett, R. C., & Cooper, R. G. (1982). The misuse of marketing. *McKinsey Quarterly*, *3*, 52-70.

Bhamra, T. (2004). Ecodesign: the search for new strategies in product development. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *218*(5), 557-569.

Bilgram, V., & Laarmann, F. (2023). Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. *IEEE Engineering Management Review*, *51*(2), 18-25.

Bretschneider, U., Rajagopalan, B., & Leimeister, J. M. (2012). Idea generation in virtual communities for innovation: The influence of participants' motivation on idea quality. 2012 45th Hawaii International Conference on System Sciences, 3467-3479. IEEE.

Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings, and future directions. *Academy of management review*, *20*(2), 343-378.

Chan, K. W., Li, S. Y., & Zhu, J. J. (2018). Good to be novel? Understanding how idea feasibility affects idea adoption decision making in crowdsourcing. *Journal of Interactive Marketing*, *43*(1), 52-68.

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cooper, R. G. (1983). A process model for industrial new product development. *IEEE Transactions on Engineering Management*(1), 2-11.

Cooper, R. G. (1988). Predevelopment activities determine new product success. *Industrial marketing management*, *17*(3), 237-247.

Cooper, R. G. (2011). *Winning at new products: Creating value through innovation*. Basic books.

Cooper, R. G. (2013). New products: What separates the winners from the losers and what drives success. *PDMA handbook of new product development*, *3*, 34.

Cooper, R. G. (2014). What's next?: After stage-gate. *Research-Technology Management*, *57*(1), 20-31.

Cooper, R. G. (2017). Idea-to-Launch Gating Systems: Better, Faster, and More Agile: Leading firms are rethinking and reinventing their idea-to-launch gating systems,

Cooper, R. G. (2023). Revolution or Evolution: A Study of the Adoption of AI in New-Product Development. *Penn State Institute for the Study of Business Markets: Oct*.

Cooper, R. G. (2024). The AI transformation of product innovation. *Industrial marketing management*, *119*, 62-74.

Cooper, R. G., & Edgett, S. J. (2012). Best practices in the idea-to-launch process and its governance. *Research-Technology Management*, *55*(2), 43-54.

Cooper, R. G., & McCausland, T. (2024). AI and New Product Development. *Research-Technology Management*, *67*(1), 70-75. https://doi.org/10.1080/08956308.2024.2280485

Dean, D. L., Hender, J., Rodgers, T., & Santanen, E. (2006). Identifying good ideas: constructs and scales for idea evaluation. *Journal of Association for Information Systems*, *7*(10), 646-699.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology*, *53*(3), 497.

Eisenberger, R., & Rhoades, L. (2001). Incremental effects of reward on creativity. *Journal of personality and social psychology*, *81*(4), 728.

Eisenberger, R., & Shanock, L. (2003). Rewards, intrinsic motivation, and creativity: A case study of conceptual and methodological isolation. *Creativity Research Journal*, *15*(2-3), 121-130.

Ernst, H. (2002). Success factors of new product development: a review of the empirical literature. *International journal of management reviews*, *4*(1), 1-40.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, *66*(1), 111-126.

Filippi, S. (2023). Measuring the impact of ChatGPT on fostering concept generation in innovative product design. *Electronics*, *12*(16), 3535.

Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand journal of psychiatry*, *36*(6), 717-732.

Gavin, H. (2008). Understanding research methods and statistics in psychology.

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., & Schader, M. (2011). Managing the crowd: towards a taxonomy of crowdsourcing processes.

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management science*, *56*(4), 591-605.

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A State of the Art Review of large Generative AI models. *arXiv preprint arXiv:2301.04655*.

Grewal, D., Guha, A., Satornino, C. B., & Schweiger, E. B. (2021). Artificial intelligence: The light and the darkness. *Journal of Business Research*, *136*, 229-236.

Griffin, A., & Hauser, J. R. (1992). Patterns of communication among marketing, engineering and manufacturing—A comparison between two new product teams. *Management science*, *38*(3), 360-373.

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial intelligence (AI) or intelligence augmentation (IA): what is the future? *Ai*, *1*(2), 8.

Hoyer, W. D., Chandy, R., Dorotic, M., Krafft, M., & Singh, S. S. (2010). Consumer cocreation in new product development. *Journal of Service research*, *13*(3), 283-296.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*(2).

Huizingh, E. K. (2011). Open innovation: State of the art and future perspectives. *Technovation*, *31*(1), 2-9.

Im, S., & Workman Jr, J. P. (2004). Market orientation, creativity, and new product performance in high-technology firms. *Journal of marketing*, *68*(2), 114-132.

Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization science*, *17*(1), 45-63.

Joosten, J., Bilgram, V., Hahn, A., & Totzek, D. (2024). Comparing the ideation quality of humans with generative artificial intelligence. *IEEE Engineering Management Review*.

Kakatkar, C., Bilgram, V., & Füller, J. (2020). Innovation analytics: Leveraging artificial intelligence in the innovation process. *Business Horizons*, *63*(2), 171-181.

Kalota, F. (2024). A primer on generative artificial intelligence. *Education Sciences*, *14*(2), 172.

Kijkuit, B., & Van Den Ende, J. (2007). The organizational life of an idea: Integrating social network, creativity and decision-making perspectives. *Journal of Management Studies*, *44*(6), 863-882.

Kristensson, P., Matthing, J., & Johansson, N. (2008). Key strategies for the successful involvement of customers in the co-creation of new technology-based services. *International journal of service industry management*, *19*(4), 474-491.

Kudrowitz, B. M., & Wallace, D. (2013). Assessing the quality of ideas from prolific, early-stage product ideation. *Journal of Engineering Design*, *24*(2), 120-139.

Laerd Statistics. (n.d.). *How to perform a one-way MANCOVA in SPSS statistics*. https://statistics.laerd.com/spss-tutorials/one-way-mancova-using-spss-statistics.php

Lakhani, K. R., Jeppesen, L. B., Lohse, P. A., & Panetta, J. A. (2007). *The value of openess in scientific problem solving*. Division of Research, Harvard Business School Boston, MA.

Leonard, D., & Rayport, J. F. (1997). Spark innovation through empathic design. *Harvard business review*, *75*, 102-115.

Magnusson, P. R., Matthing, J., & Kristensson, P. (2003). Managing user involvement in service innovation: Experiments with innovating end users. *Journal of Service research*, *6*(2), 111-124.

Magnusson, P. R., Wästlund, E., & Netz, J. (2016). Exploring users' appropriateness as a proxy for experts when screening new product/service ideas. *Journal of Product Innovation Management*, *33*(1), 4-18.

Osborne, A. F. (1953). Applied imagination: principles and procedures of creative problem solving. *Charles Scribener's Sons, New York*.

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Park, Y. S., Konge, L., & Artino Jr, A. R. (2020). The positivism paradigm of research. *Academic medicine*, *95*(5), 690-694.

Perinetti, G. (2018). StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. *South European journal of orthodontics and dentofacial research*, *5*(1), 3-5.

Piller, F. T., & Walcher, D. (2006). Toolkits for idea competitions: a novel method to integrate users in new product development. *R&d Management*, *36*(3), 307-318.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management*, *29*(2), 245-256.

Poole, S., & Simon, M. (1997). Technological trends, product design and the environment. *Design Studies*, *18*(3), 237-248.

Reinig, B. A., Briggs, R. O., & Nunamaker, J. F. (2007). On the measurement of ideation quality. *Journal of Management Information Systems*, *23*(4), 143-161.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2010). The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British journal of psychology*, *101*(1), 47-68.

Roberts, D. L., & Candi, M. (2024). Artificial intelligence and innovation management: Charting the evolving landscape. *Technovation*, *136*, 103081.

Salter, A., Criscuolo, P., & Ter Wal, A. L. (2014). Coping with open innovation: responding to the challenges of external engagement in R&D. *California management review*, *56*(2), 77-94.

Schemmann, B., Herrmann, A. M., Chappin, M. M., & Heimeriks, G. J. (2016). Crowdsourcing ideas: Involving ordinary users in the ideation phase of new product development. *Research Policy*, *45*(6), 1145-1154.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, *126*(5), 1763-1768.

Schulze, A., & Hoegl, M. (2008). Organizational knowledge creation and the generation of new product ideas: A behavioral approach. *Research Policy*, *37*(10), 1742-1750.

Schweitzer, F. M., Buchinger, W., Gassmann, O., & Obrist, M. (2012). Crowdsourcing: Leveraging innovation through online idea competitions. *Research-Technology Management*, *55*(3), 32-38.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Shaner, M. B. (2015). Managing the co-creation of innovation: the influence of team regulatory style and reflexivity on customer idea selection and innovation outcomes.

Singla, A., Sukharevsky, A., Yee, L., & Chui, M. (2024, May 30). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Smith, D. (2003). Five principles for research ethics. *Monitor on psychology*, *34*(1), 56.

Spanjol, J., Noble, C. H., Baer, M., Bogers, M. L., Bohlmann, J., Bouncken, R. B., Bstieler, L., De Luca, L. M., Garcia, R., & Gemser, G. (2024). Fueling innovation management research: Future directions and five forward-looking paths. *Journal of Product Innovation Management*, *41*(5), 893-948.

Surowiecki, J. (2004). *The wisdom of crowds: Why the Many are Smarter Than the Few and how Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Doubleday Books.

Ting, H., Memon, M. A., Thurasamy, R., & Cheah, J.-H. (2025). Snowball sampling: A review and guidelines for survey research. *Asian Journal of Business Research (AJBR)*, *15*(1).

Tushman, M., & Nadler, D. (1986). Organizing for innovation. *California management review*, *28*(3), 74-92.

Ulrich, K. T. (2011). *Design: Creation of artifacts in society*. Karl T. Ulrich.

Urlich, K., & Eppinger, S. (2016). Product design and development. ed. In: New York: McGraw-Hill Education.

Von Hippel, E. (2006). *Democratizing innovation*. the MIT Press.

Wang, Y., Zhang, H., & Song, M. (2020). Does big data–embedded new product development influence project success? *Research-Technology Management*, *63*(4), 35-42.

Zeng, Q., Zhang, L., Guo, Q., Zhuang, W., & Fan, W. (2022). Factors influencing user-idea selection in open innovation communities. *International journal of electronic commerce*, *26*(4), 415-440.

Zhang, H., Song, M., & He, H. (2020). Achieving the success of sustainability development projects through big data analytics and artificial intelligence capability. *Sustainability*, *12*(3), 949.

Zhang, H., Zhang, X., & Song, M. (2021). Deploying AI for new product development success. *Research-Technology Management*, *64*(5), 50-57.

Appendix A: Crowdsourced ideas (name, target audience and flavor description)

1. **Matcha Taro Swirl - adults**
High quality matcha ice cream with ube/taro (paarse yam) swirl. The green ice cream with the purple syrup swirl through it will complement each other perfectly

2. **Cinnabuddy - everyone**
Cinnamon, white chocolate, raspberry.

3. **Crunchy Cocolepine - everyone**
Coconut, Lenin, pineapple with crunchy nuts.

4. **Fun on the beach - adults**
Passionfruit, vodka, cranberry, orange.

5. **Saviour of periods - adults**
HEMA chocolate cake, chocolate ganache, chocolate spread on top. Very rich chocolate on the edge of dark and milk.

6. **Floral gold - adults**
Soft sweetness like saffron and orange blossom water. A yogurt base instead of cream for freshness.

7. **Vaniccella – adults**
Luxurious, creamy vanilla ice cream with crisp chocolate flakes

8. **Harissa honey ice – adults**
Harissa, honey and lemon

9. **Limoncello semifreddo - adults**
Lemon, limoncello semifreddo

10. **Intens genieten - adults**
Pistacchio, vanilla, hazelnut and chocolate

11. **Delicious gem - adults**
Apple Ginger

12. **Summer blessings - everyone**
Pineapple, pomegranate, orange and colored sprinkles

13. **Sweet grape sorbet – everyone**
Grape rolled in sugar, crushed into a sorbet ice cream

14. **White choco strawberry - kids**
Strawberry ice cream with white chocolate chips

15. **Snickers supreme - everyone**
A creamy swirl with the taste of different nuts, small chocolate chunks and caramel sauce

16. **Strawberry matcha tiramisu – adults**
Matcha ice cream, swirled with sweet mascarpone with chunks of lady finger biscuits and strawberries

17. **Fruit choco white – everyone**
Strawberry, white chocolate, banana, raspberry

18. **Deep relief - adults**
A relaxing blend of mint (perhaps THC), melatonin, and lavender. Space blue color with vanilla streaks

19. **Toffee fun - adults**
Banana and toffee

20. **Prickly dream - adults**
Vanilla cream base with cactus fruit juice and white chocolate chips

21. **Dream away - everyone**
Fresh, fruity, with a little bite. Mango, white chocolate and ginger

22. **Sweet n sour ice cream - adults**
Lemon/lime cheesecake ice cream with fresh zest, a creamy based and cookies

23. **Horchata helada – everyone**
Frozen horchata (rice, cinnamon drink)

24. **Chocolate and lemon - everyone**
Chocolate and lemon

25. **Chardonnice - adults**
Wine sorbet ice cream

26. **Cereal killer - adults**
Cornflakes, milk, and cruchy bits with a hint of Baileys

27. **Grandma's lemon cake – adults**
Lemon cake flavor with lemon zest

28. **Chunky cookie - everyone**
Vanilla base. Cookie dough with lots of chocolate chunks, cooked and raw dough pieces, nuts and fudge swirled in.

29. **Virginia fields – everyone**
Double churned French vanilla with real peanut butter on top

30. **Spicy chai - adults**
Chai masala spices like cardamom, cinnamon, black pepper, ginger and vanilla

31. **Spicy rizz** - **adults**
Ginger and raspberry

32. **Caramel s'mores - everyone**
Chocolate with salted caramel and marshmellow

33. **Tough tingle - adults**
Pineapple, chocolate with spicy flakes

34. **Kinder bueno - everyone**
Bueno ice cream with hazelnuts and chocolate

35. **Pumpkin chai pie - adults**
Pumpkin pie and chai latte ice cream

36. **Licopear - adults**
Liquorice and pear

37. Cashew deluxe - everyone
Cashew nut ice cream with little chocolate pieces

38. Sweety peanuts - everyone
Peanut butter and strawberry jam

39. Koffie verkeerd - everyone
Vanilla, mocha and coffee.

40. Pineapples and cream - everyone
Pineapple chunks and vanilla cream

41. Strawberry brownie – everyone
Chocolate, strawberries, and brownie chunks

42. Apenkoppen - everyone
Banana ice cream with a little bit of liquorice

43. Nordic forest - adults
Pine tree flavor with blueberry sauce swirls

44. Strawberry cheesecake with a twist - everyone
Strawberry cheesecake combined with white chocolate

45. Chocolate delight - everyone
Chocolate and pistachio ice cream

46. Grass coffee - adults
Mint, chocolate, matcha, coffee and vanilla ice cream

47. Citrus - adults
Creamy and citrus flavored

48. Mocha dream - adults
Chocolate ice cream with pieces of chocolate cake and thick mocha sauce

49. Zour - everyone
Sour green apple, sorbet ice cream with jelly pieces

50. Dark choco pretzel - adults
Dark chocolate flavored ice cream with salted pretzel chunks

51. Berry blizz - everyone
Vanilla base, tangy strawberry sauce with fried strawberries for crunch and white chocolate drizzle

52. Sea buckthorn (duindoorn bes) with chocolate chip - adults
Sorbet, fresh, light, sea buckthorn, sour but sweet, with small dark chocolate chip

53. Salty pistaramel - everyone
Pistachio and salty caramel

54. Apple crumbles - everyone
Vanilla ice cream base, oven-baked red apple and oat flakes, topped with melted butter and cinnamon sauce

55. Rhubarb cookie swirls - everyone
Vanilla ice cream with rhubarb jam and crumbled white chocolate chip cookies swirled through

56. **Granny's apple pie - adults**
Apple pie, with apples, cinnamon, raising and perhaps a few crust chunks

57. **Strawberry sorbet - everyone**
Strawberry sorbet ice cream with strawberry pieces

58. **Peach and basil - adults**
Roasted peach and basil

59. **Lemon fizz - everyone**
Sour and sweet lemon taste with a carbonated/sparkling effect when you eat it

60. **Strawberry basil – adults**
Strawberry and basil

61. **Pistachio dream - adults**
Pistachio ice cream with salted pistachio nuts mixed in

62. **Neopolitan wishes – everyone**
Vanilla ice cream, strawberry ice cream, chocolate ice cream, white chocolate chips

63. **Vanilla - everyone**
Vanilla, salted caramel, raisins, chocolate chunks and nuts

64. **Caramel nutlicious - everyone**
Rich caramel ice cream, salted caramel with blondie and some nuts

65. **Berry biscuit - everyone**
Vanilla base with oreo crumbs throughout, raspberry pieces and syrup alongside a chocolate syrup drizzled throughout

66. **All of the best - kids**
Box of celebrations chocolate all mixed together

67. **Bokkepootjes delight - adults**
Bokkepootjes, witte chocolate, vanilla ice cream

68. **Lava cake ice cream - everyone**
Vanilla ice cream base, lava cake chunks, chocolate ganache

69. **Twisted choc mint - everyone**
Mint ice cream mixed with creamy milk chocolate ice cream with popping candy and gummy bears mixed in

70. **Sweetella - everyone**
Caramelized hazelnut

71. **Lemon soda - everyone**
Lemon, lime and mint

72. **Pavloved – everyone**
Vanilla/strawberry ice cream, with chewy pavlova and strawberry pieces mixed through

73. **Beloved blueberry – everyone**
Creamy vanilla ice cream, with blueberry jam swirls and white chocolate chunks and cookie crumble

74. Vanilla currant – adults
Vanilla based ice cream with blackcurrant sauce and tiny chocolate flakes

75. Berry-crumblicious - everyone
Raspberry and rhubarb pie crumble, vanilla ice cream

76. Twirly girly - everyone
A pink strawberry and raspberry and cherry ice cream coated with white chocolate

77. Cookie crumbs - everyone
A cookie crumble ice cream base with toppings such as cinnamon apples, cherry or chocolate

78. Nutty & Nice (or Hazel Daze) - everyone
Hazelnut base with brownie pieces and chocolate chunks

79. CaraMAZING - everyone
Extremely sweet, salted, caramel ice cream with a bunch of caramel swirls

80. Carameliosa - adults
Vanilla ice cream, with pecans, caramel and fudge with a hint of salt

81. Cool caramel - adults
Mint and soft caramel with a swirl of caramel sauce

82. Pearlique - adults
A swirl with pear and dark chocolate ice cream topped with brownie pieces and salty liquorice

83. Macadamia Mocca - adults
Chocolate and coffe ice cream swirled togethers with bits of macadamia nuts and marshmelows

84. Blueberry honey - adults
Blueberries and honey, with a small hint of coconut

85. Hot - adults
Lemon, black pepper, chili and dates dipped in dark chocolate

86. Fudge dream - everyone
High quality vanilla base with some soft fudge chunks and salty caramel sauce

87. Vanilla choco-pistach-ioooo - everyone
Chocolate, pistachios and vanilla

88. Chocolate fudge brownie - everyone
Fudgy browny chunks, dark chocolate ice cream

89. Crunchy black raspberry madness - everyone
Vegan, liquorice and raspberry ice cream with cookie bits (digestive style)

90. Melon and chocolate - everyone
Melon sorbet ice cream with chocolate

91. Tiramisu twist - adults
Tiramisu ice cream with caramel swirled in and a layer of chocolate on top

92. Choco vanilla cookie dough bonanza - everyone
Strong vanilla mixed with chocolate chip cookie dough

93. Blueberry vanilla - everyone
Blueberry and vanilla soft-serve ice cream, no chunks

94. Pechpricot - everyone
Peach and apricot yoghurt ice cream

95. Chocolate chip banana bread - everyone
Chocolate banana bread, vanilla base with banana flavor, cinnamon and burnt butter. Swirled in with tempered dark chocolate and fresh banana bread chunks

96. PB&J ice cream - kids
Peanut flavored ice cream with pieces of strawberry or other red fruits

97. Stracciatella - adults
Vanilla and chocolate

98. Coco mango - everyone
Coconut and mango ice cream swirled together

99. Strawberryvan - everyone
Vanilla and strawberry

100. Exotic explosion - everyone
Passionfruit and lemon

Appendix B: Survey Questions Crowd

Introduction: This survey is part of a master's thesis in Innovation and Entrepreneurship at Radboud University. The purpose is to explore how people generate creative new product ideas. In this short survey, you will be asked to invent ONE new ice cream flavor and describe it in detail. Your idea should be original, creative, and realistic. You are asked to come up with a flavor that customers would find both exciting and enjoyable. The survey will take approximately 5-7 minutes to complete.

Important: This is a test of your idea generation skills. To ensure valid results, you are NOT allowed to use any form of AI assistance (e.g., ChatGPT, Gemini, Copilot etc.) to complete this task. Submissions should reflect your own thinking only.

Question 1: What is your age?

- Younger than 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- Older than 65

Question 2: Describe your ice cream flavor. What are the main ingredients and flavor combinations? Please write approximately 20 words.

Question 3: What is the name of your ice cream flavor?

Question 4: Who do you think this ice cream flavor would appeal to most?

- Kids
- Adults
- Everyone

Question 5: Even though this is a hypothetical contest, please indicate to what extent the following statements apply to you if this were a real competition

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I am more likely to contribute ideas if there is a (financial) reward					
I would be more interested if the winning flavor were actually produced					
I would feel more pressure to do well if it were a real contest					
I enjoy coming up with ideas regardless of rewards					

Question 6: Please answer honestly: While completing this survey, did you use any form of AI assistance (e.g., ChatGPT, Gemini, Copilot etc.) to help generate your ice cream flavor idea?

- No, I generated the idea entirely by myself
- Yes, I used AI to help me generate or refine the idea

Appendix C: ChatGPT generated ideas (name, target audience and flavor description)

1. **Midnight Campfire Crunch – everyone**
Smoky dark chocolate ice cream with gooey marshmallow swirls, honeycomb shards, and a hint of cinnamon.

2. **Mango Tango Surprise – adults**
Creamy mango ice cream swirled with spicy chili-lime ribbons and bursting pomegranate seeds for a sweet, tangy, spicy twist.

3. **Lemon Lavender Dream - adults**
Light lemon ice cream with lavender honey swirls and white chocolate chips.

4. **Caramel Cloudberry Crave - everyone**
Nordic cloudberry (kruising tussen frambozen, bramen en rode bessen) swirled into caramel ice cream with oat crumble

5. **Chai'd Up Delight - everyone**
Spiced chai tea ice cream with ginger cookie chunks and a vanilla glaze swirl

6. **Sakura Snowfall – everyone**
Cherry blossom-flavored cream with mochi pieces and strawberry pearls

7. **Banana Burnt Sugar Crunch – adults**
Roasted banana base with burnt sugar ripple and candied walnuts.

8. **Gummy Galaxy Swirl – kids**
Mixed berry ice cream with sour gummy clusters and cosmic popping candy.

9. **Espresso Biscotti Bash – adults**
Bold espresso ice cream loaded with almond biscotti and chocolate shavings.

10. **Golden Milk & Honey - adults**
Turmeric-vanilla base with golden honeycombs and a cinnamon finish.

11. **Dragonfruit Lemon Pop – everyone**
Vibrant dragonfruit sorbet with lemon fizz crystals and chia seeds. Electric, refreshing, and playful.

12. **Sakura White Chocolate – adults**
Cherry blossom–infused base with white chocolate petals and matcha mochi bits. Floral, delicate, and creamy.

13. **Churro Dulce Dream – adults**
Cinnamon churro base with dulce de leche core and sugar dust. Crunchy, creamy, and decadent.

14. **Peach Basil Breeze – adults**
Juicy peach cream with fresh basil ripple and a buttery biscuit crumble.

15. **Toasty Taro Crisp – everyone**
Taro-flavored (tropische wortelknol met milde, nootachtige smaak) base with toasted coconut and crunchy rice bits.

16. **Matcha Mochi Madness – everyone**
Creamy matcha green tea ice cream with chewy mochi bites

17. Coconut Lime Lightning – everyone
Coconut ice cream swirled with zesty lime curd and mint flakes.

18. Midnight Earl Grey Crunch – adults
Earl Grey-infused base with chocolate-covered sunflower seeds and orange zest.

19. S'moresplosion – everyone
Toasted marshmallow ice cream with graham cracker chunks and melted chocolate core.

20. Pumpkin Pie Palooza – everyone
Spiced pumpkin cream with pie crust bites and whipped cream ribbons.

21. Raspberry Rose Ritual – adults
Delicate rose ice cream swirled with tart raspberry puree and pistachio dust.

22. Bubblegum Stardust – kids
Bright bubblegum base with rainbow sprinkles and glittery popping pearls.

23. Macadamia Maple Melt – adults
Rich maple cream with roasted macadamia clusters and a brown sugar swirl.

24. Cherry Cola Freeze – everyone
Cherry cola ice cream with fizzy candy crunch and maraschino cherry bits.

25. Turkish Delight– adults
Rose and lemon ice cream with chewy candy chunks and powdered sugar swirl.

26. Churro Choco Crunch – everyone
Cinnamon sugar cream with churro bites and chocolate ripple

27. Ginger Snap Snap! – adults
Ginger ice cream with molasses cookie crumbles and candied ginger bits.

28. Lychee Lime Pop – everyone
Lychee cream swirled with zesty lime and tapioca pearls.

29. Yuzu Explosion – everyone
Bright yuzu citrus base with white chocolate shavings and popping candy.

30. Black Sesame Swirl – adults
Nutty black sesame base with caramel ribbon and matcha crunch.

31. Pineapple Jalapeño Twist – adults
Sweet pineapple sorbet with a touch of jalapeño heat and lime zest. Spicy-sweet and bold.

32. Key Lime Thunder – everyone
Tangy key lime pie base with cookie swirl and candied lime peel.

33. Butterscotch Bliss – kids
Butterscotch ice cream swirled with toffee chunks and magic dust.

34. Date Night Delight – adults
Medjool date ice cream with vanilla bean and caramelized pecans.

35. Blueberry Pancake – everyone
Maple base with blueberry compote and pancake chunks.

36. Rosemary Grapefruit Chill – adults
Herbal grapefruit base with rosemary shortbread crumble.

37. Almond Cherry Cobbler – everyone
Almond ice cream with cherry compote and buttery crumble swirls.

38. Molten Chocolate Lava – everyone
Deep chocolate base with molten fudge core and brownie pieces.

39. Cactus Pear Pop – everyone
Prickly pear sorbet with coconut cream swirls and lemon zest

40. Saffron Pistachio Silk – adults
Rich saffron ice cream with crunchy pistachio and rose syrup swirls.

41. Chili Chocolate Tango – adults
Dark chocolate base with smoky chili swirls and chocolate nibs.

42. Apple Crumble Sundae – everyone
Spiced apple base with oat crumble and caramel core.

43. Vanilla Bean Basil Bomb – adults
Vanilla bean base with basil swirl and honey drizzle.

44. Kiwi Coconut Fizz – everyone
Tangy kiwi sorbet with coconut cream and sparkling popping powder.

45. Cardamom Cookie Craze – adults
Spiced cardamom ice cream with sugar cookie chunks and white chocolate.

46. Pineapple Sage Sorcery – adults
Pineapple base with sage leaf essence and candied ginger.

47. Hibiscus Heaven –adults
Floral hibiscus ice cream with mixed berry ripple and honey crystals

48. Cookies & Matcha – everyone
Matcha base with cookie chunks and chocolate drizzle.

49. Root Beer Ripple – everyone
Root beer ice cream with vanilla swirl and toffee bits.

50. Pistachio Baklava Crunch – adults
Pistachio ice cream with honey swirls, phyllo (krokant filodeeg) shards, and toasted nuts

51. Peach Bourbon Glaze – adults
Juicy peach cream infused with bourbon and glazed pecan crunch.

52. Charcoal Vanilla Marble – adults
Dramatic black vanilla ice cream marbled with activated charcoal and dark chocolate flecks.

53. Toasted Marshmallow Mocha – adults
Coffee-chocolate base with toasted marshmallow swirl and graham chunks.

54. Mulled Wine Sorbet – adults
Wine sorbet steeped with orange, cinnamon, clove, and star anise. Fruity and spiced.

55. Pomegranate Pistachio Velvet – adults
Rich pomegranate cream with pistachio crumble and white chocolate ripple.

56. Apple Sage Smash – adults
Stewed (gestoofd) apple ice cream with sage-infused caramel and pie crust pieces.

57. Oat Milk Espresso Fudge – adults
Vegan oat base with espresso and thick chocolate fudge ribbons.

58. Plum & Star Anise Cream – adults
Plum sorbet rippled through star anise-infused vanilla base. Sweet, spiced, and layered.

59. Fig & Walnut Delight – adults
Honey-roasted fig base with candied walnuts and cinnamon dust.

60. Iced Tea Twist – everyone
Creamy tea base with sweet, condensed milk ripple and chewy tapioca balls.

61. Ginger Peach Harmony –everyone
Peach ice cream with ginger syrup and crunchy almond praline.

62. Orange Cardamom Fusion – adults
Orange zest ice cream with cardamom swirl and pistachio nibs.

63. Rum Raisin Reboot – adults
Classic vanilla rum base with juicy dark rum-soaked raisins and caramel core.

64. Basil Strawberry Ripple – everyone
Strawberry base with fresh basil swirl and lemon zest.

65. Cantaloupe Mint Freeze – everyone
Melon sorbet blended with mint leaves and coconut water ribbons.

66. Spicy Mango Mezcal – adults
Mango sorbet with smoky mezcal swirl and chili salt.

67. Creme Brûlée Crunch – everyone
Custard base with caramelized sugar shards and buttery biscuit crumble.

68. Boba Milk Tea Bonanza – everyone
Creamy milk tea base with chewy tapioca pearls and caramel drizzle.

69. Honey Hibiscus Rose – adults
Floral hibiscus base with rose syrup swirl and honeycomb crunch.

70. Cherry Almond Rhapsody – everyone
Almond cream with tart cherry ribbons and toasted almond crunch.

71. Smoked Vanilla Bean – adults
Smoky vanilla ice cream with burnt sugar and charred wood essence.

72. Lime Jalapeño Blast – adults
Lime sorbet infused with jalapeño syrup and candied lime peel.

73. Blueberry Lemon Thyme – everyone
Blueberry cream with lemon curd swirl and thyme sprinkles.

74. Gingerbread Espresso Swirl – adults
Espresso base with gingerbread chunks, molasses ribbons, and nutmeg dust.

75. Carrot Cake– everyone
Spiced carrot cake base with cream cheese frosting ripple and walnut crunch.

76. Sea Salt Rose Caramel – adults
Salted caramel base with rose syrup swirl and candied pistachios. Floral and buttery.

77. Pineapple Rum Swirl – adults
Grilled pineapple base with spiced rum ripple and cinnamon dust.

78. Black Tea Raspberry Ripple – adults
Black tea base with raspberry swirl and almond flakes.

79. Toasted Almond Joy – everyone
Coconut ice cream with toasted almonds and chocolate fudge core.

80. Sour Cherry Cheesecake – everyone
Cream cheese base with sour cherry swirl and crust crumble

81. Papaya Lime Cream – everyone
Creamy papaya ice cream with lime zest and coconut sugar. Tropically sweet.

82. Hazelnut Amaretto Crunch – adults
Hazelnut base with amaretto swirl and praline shards.

83. Cotton Candy Potion – kids
Cotton candy-flavored cream with blue raspberry syrup, star-shaped sprinkles, and a glittery core that fizzes on your tongue.

84. Jellybean Jungle – kids
Tropical fruit ice cream with chewy jellybeans, swirled with kiwi-lime syrup.

85. Tropical Coconut Swirl – everyone
Coconut base with pineapple swirl and toasted coconut flakes.

86. Caramel Apple Pie – everyone
Apple-flavored ice cream with caramel ribbons and pie crust chunks.

87. Strawberry Waffle Swirl – everyone
Strawberry ice cream with waffle cone chunks and strawberry syrup ribbons.

88. Lemon Sugar Cookie – everyone
Tangy lemon ice cream with soft sugar cookie dough bits.

89. Dark Cherry Espresso – adults
Espresso base with black cherry ribbons and chocolate-covered espresso beans.

90. Vanilla Bean Pretzel Pop – everyone
Classic vanilla bean base with crushed pretzel pieces and salted caramel drizzle.

91. Watermelon Mint Sorbet – everyone
Juicy watermelon sorbet with cool mint leaves and lemon zest.

92. Choco Banana Swirl – everyone
Creamy banana ice cream with fudge ribbons and chocolate chip chunks.

93. Rainbow Sorbet – everyone
A fruity trio of mango, raspberry, and blueberry sorbets

94. Blueberry Muffin Scoop – everyone
Vanilla base with blueberry swirls and muffin crumble chunks.

95. Salted popcorn Chocolate Swirl – everyone
A creamy chocolate base with cookie dough and a hint of salted popcorn

96. Cinnamon Caramel Swirl
A creamy caramel base with coconut flakes and a hint of cinnamon

97. Bourbon Hazelnut Swirl – adults
A creamy hazelnut base with pistachio crunch and a hint of bourbon

98. Layered Vanilla Dream – everyone
Vanilla ice cream with pretzel chunks, topped with gold dust, and layered with cheesecake chunks.

99. Gummy orange – kids
Creamy orange ice cream with gummy bears

100. Espresso Biscotti - adults
Creamy espresso ice cream mixed with crunchy almond biscotti pieces and a swirl of chocolate ganache

Appendix D: ChatGPT Prompt

(asked once)

"You will be asked to invent ONE new ice cream flavor and describe it in detail. Your idea should be original, creative, and realistic. You are asked to come up with a flavor that customers would find both exciting and enjoyable. Describe your ice cream flavor. What are the main ingredients and flavor combinations? Please write approximately 20 words. What is the name of your ice cream flavor? Who do you think this ice cream flavor would appeal to most (kids, adults, or everyone)?"

Follow-up prompt (asked 99 times)

"*Give me another idea that is different from the previous ones"*.

Appendix E: List of Evaluation Items for Raters

Idea Quality: Novelty
Self-developed, based on Kudrowitz & Wallace (2013), and Poetz and Schreier (2012).
5-point Likert scale.

Compared to products currently on the market, this new product idea…

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is original and uncommon					

Idea Quality: Feasibility
Self-developed, based on Kijkuit and Van Den Ende (2007), and Kudrowitz and Wallace (2013).
5-point Likert scale.

Compared to products currently on the market, this new product idea…

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is both feasible and marketable					

Idea Quality: Customer Benefit
Self-developed, based on Im and Workman (2004).
5-point Likert scale.

Compared to competitors, the new product idea…

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is considered suitable for customers' desires					

External Motivation to Participate: (Non-) Monetary Rewards
Self-developed, based on Amabile (1994) and Bretschneider et al. (2012).
5-point Likert scale.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I am more likely to contribute ideas if there is a (financial) reward					
I enjoy coming up with ideas regardless of rewards					
I would feel more pressure to do well if it were a real contest					
I would be more interested if the winning flavor were actually produced					

Appendix F: SPSS Output Inter-rater Reliability Test

Appendix F1: ICC Novelty

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	Sig
Single Measures	-.015[a]	-.153	.124	.971	199	199	.583
Average Measures	-.030[c]	-.361	.220	.971	199	199	.583

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Novelty_IJswinckel	200	4	1	5	3.51	1.156	1.337
Novelty_Kees	200	3	2	5	3.68	.774	.599
Valid N (listwise)	200						

Appendix F2: ICC Customer Benefit

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	Sig
Single Measures	.205[a]	.069	.334	1.516	199	199	.002
Average Measures	.340[c]	.128	.501	1.516	199	199	.002

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
CustomerBenefit_IJswinckel	200	4	1	5	3.39	1.197	1.434
CustomerBenefit_Kees	200	3	2	5	4.52	.770	.593
Valid N (listwise)	200						

Appendix F3: ICC Feasibility

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.043[a]	-.096	.180	1.090	199	199	.272
Average Measures	.083[c]	-.212	.306	1.090	199	199	.272

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Feasibility_IJswinckel	200	4	1	5	3.50	1.147	1.317
Feasibility_Kees	200	2	3	5	4.19	.562	.315
Valid N (listwise)	200						

Appendix G: SPSS Output Assumption Tests

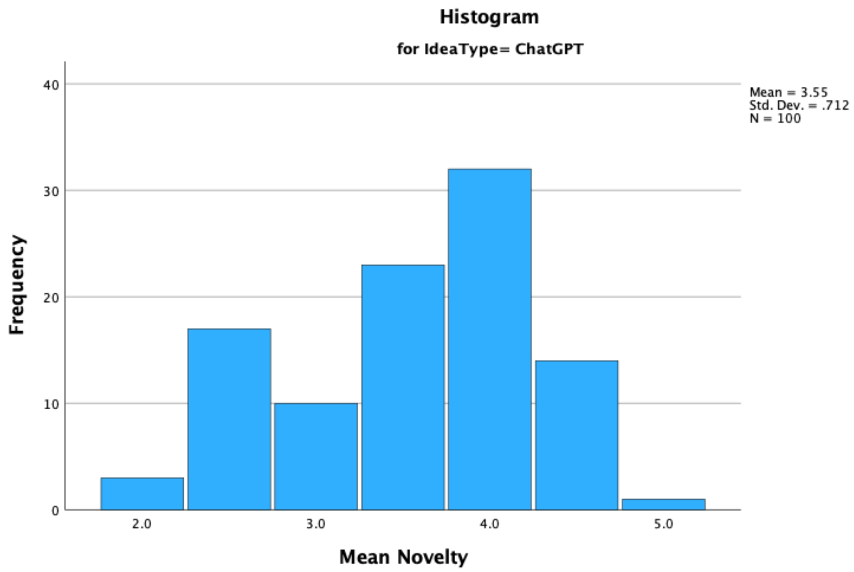
Appendix G1: Normality test

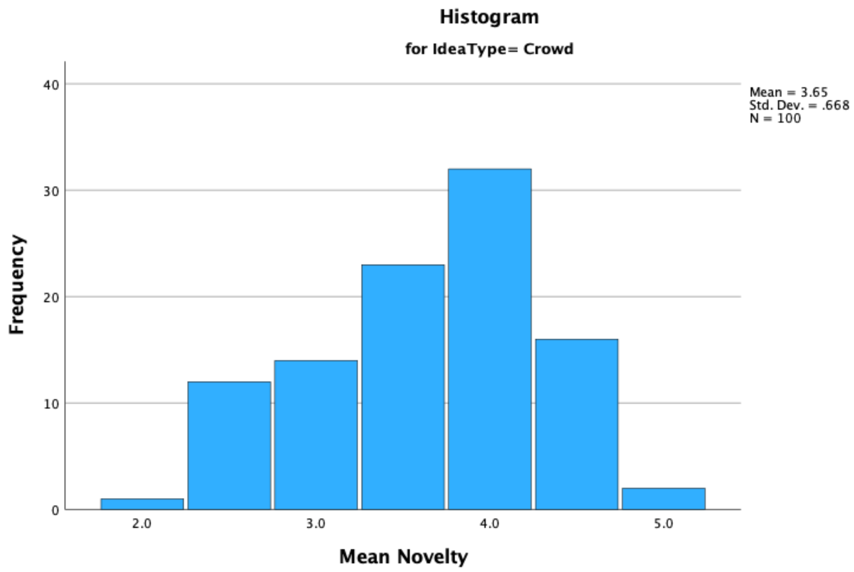
Tests of Normality

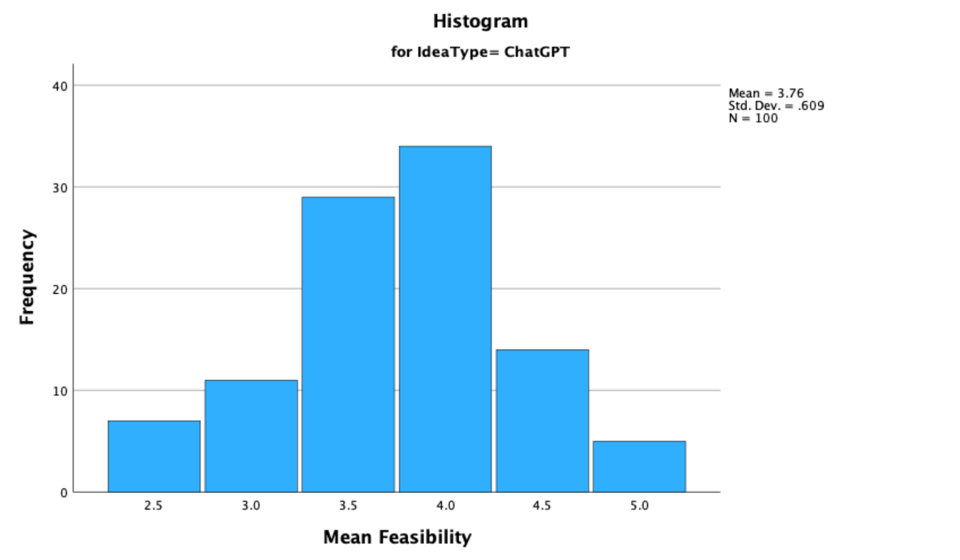
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Idea Type	Statistic	df	Sig.	Statistic	df	Sig.
Mean Novelty	ChatGPT	.206	100	<.001	.912	100	<.001
	Crowd	.203	100	<.001	.927	100	<.001
Mean Feasibility	ChatGPT	.183	100	<.001	.930	100	<.001
	Crowd	.201	100	<.001	.916	100	<.001
Mean Customer Benefit	ChatGPT	.194	100	<.001	.923	100	<.001
	Crowd	.186	100	<.001	.886	100	<.001

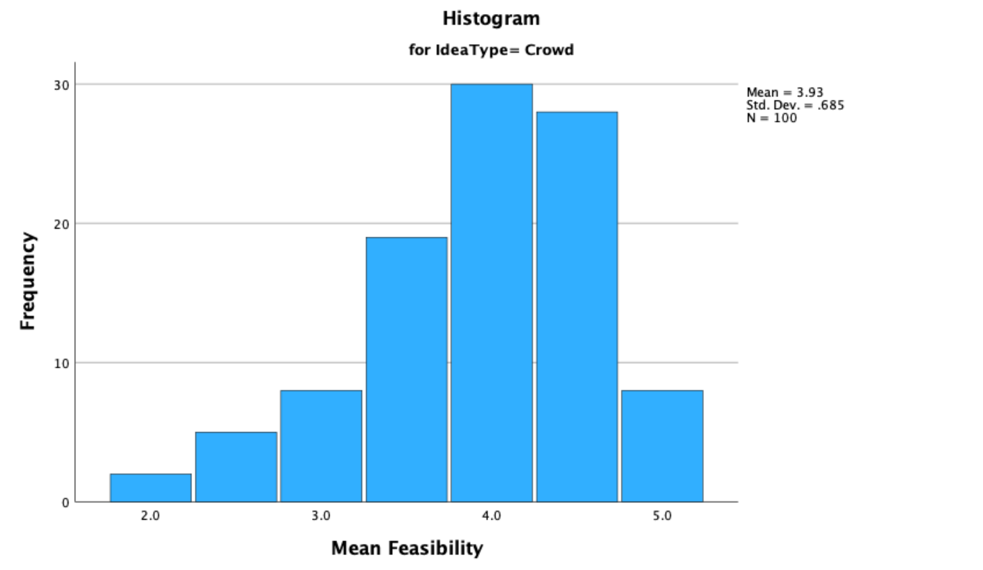
a. Lilliefors Significance Correction

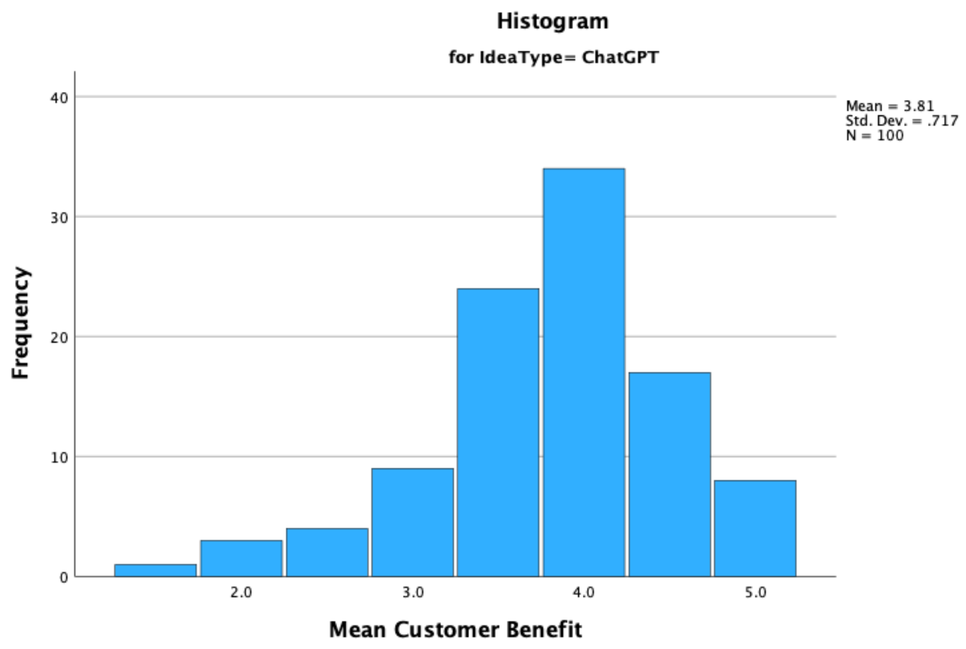
Appendix G2: Histograms

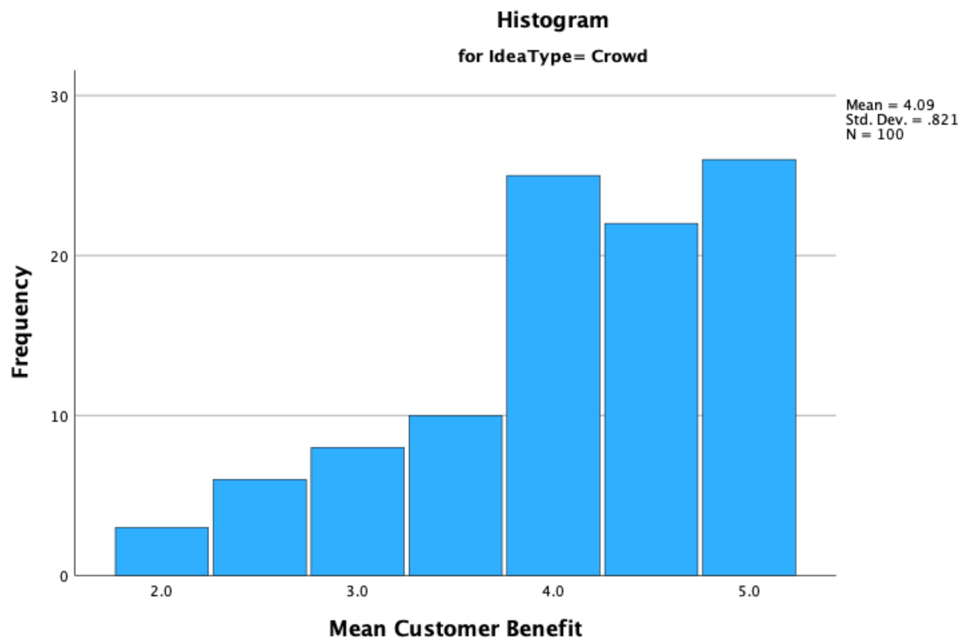





Histogram
for IdeaType= ChatGPT
Mean = 3.76
Std. Dev. = .609
N = 100
Frequency
0
10
20
30
40
2.5
3.0
3.5
4.0
4.5
5.0
Mean Feasibility


Histogram
for IdeaType= Crowd
Mean = 3.93
Std. Dev. = .685
N = 100
Frequency
0
10
20
30
2.0
3.0
4.0
5.0
Mean Feasibility





Appendix G3: Levene's Test

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Mean Novelty	.268	1	198	.605
Mean Feasibility	.414	1	198	.521
Mean Customer Benefit	2.065	1	198	.152

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + WordCount + IdeaType

Appendix G4: Box’s M test

Box's Test of Equality of Covariance Matrices[a]

Box's M	5.662
F	.928
df1	6
df2	284044.075
Sig.	.473

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + WordCount + IdeaType

Appendix G5: VIF and Multicollinearity

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	1.948	.308		6.335	<.001		
	Mean Feasibility	.234	.106	.221	2.198	.029	.421	2.376
	Mean Customer Benefit	.192	.088	.217	2.177	.031	.428	2.338
	Word Count	.000	.009	-.003	-.051	.959	.973	1.028

a. Dependent Variable: Mean Novelty

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	1.269	.205		6.192	<.001		
	Mean Customer Benefit	.589	.042	.706	14.192	<.001	.847	1.181
	Word Count	-.011	.006	-.084	-1.833	.068	.990	1.010
	Mean Novelty	.103	.047	.109	2.198	.029	.852	1.174

a. Dependent Variable: Mean Feasibility

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	.160	.271		.590	.556		
	Word Count	.004	.007	.023	.500	.618	.974	1.026
	Mean Novelty	.123	.057	.109	2.177	.031	.852	1.174
	Mean Feasibility	.860	.061	.718	14.192	<.001	.833	1.201

a. Dependent Variable: Mean Customer Benefit

Appendix H: SPSS Output Descriptive Statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Novelty	400	1	5	3.60	.986	.973
Feasibility	400	1	5	3.84	.966	.933
Customer Benefit	400	1	5	3.95	1.154	1.331
Total idea quality (sum of 3 items)	400	3	15	11.39	2.577	6.640
Average Idea Quality score between raters	400	7.0	14.0	11.392	1.7428	3.037
IdeaType_code	400	1.00	2.00	1.5000	.50063	.251
ID	400	1	200	100.50	57.807	3341.604
Valid N (listwise)	400					

Descriptives

	IdeaType_code			Statistic	Std. Error
Novelty	ChatGPT	Mean		3.55	.071
		95% Confidence Interval for Mean	Lower Bound	3.41	
			Upper Bound	3.69	
		5% Trimmed Mean		3.60	
		Median		4.00	
		Variance		1.003	
		Std. Deviation		1.001	
		Minimum		1	
		Maximum		5	
		Range		4	
		Interquartile Range		1	
		Skewness		-.791	.172
		Kurtosis		.086	.342
	Crowd	Mean		3.65	.069
		95% Confidence Interval for Mean	Lower Bound	3.51	
			Upper Bound	3.78	
		5% Trimmed Mean		3.69	
		Median		4.00	
		Variance		.944	
		Std. Deviation		.971	
		Minimum		1	
		Maximum		5	
		Range		4	
		Interquartile Range		1	
		Skewness		-.864	.172
		Kurtosis		.317	.342
Feasibility	ChatGPT	Mean		3.76	.069
		95% Confidence Interval for Mean	Lower Bound	3.62	
			Upper Bound	3.90	
		5% Trimmed Mean		3.82	
		Median		4.00	
		Variance		.947	
		Std. Deviation		.973	
		Minimum		1	
		Maximum		5	
		Range		4	

		Interquartile Range		1	
		Skewness		-.823	.172
		Kurtosis		.503	.342
	Crowd	Mean		3.93	.067
		95% Confidence Interval for Mean	Lower Bound	3.80	
			Upper Bound	4.06	
		5% Trimmed Mean		4.01	
		Median		4.00	
		Variance		.910	
		Std. Deviation		.954	
		Minimum		1	
		Maximum		5	
		Range		4	
		Interquartile Range		1	
		Skewness		-1.088	.172
		Kurtosis		1.264	.342
Customer Benefit	ChatGPT	Mean		3.81	.082
		95% Confidence Interval for Mean	Lower Bound	3.65	
			Upper Bound	3.97	
		5% Trimmed Mean		3.88	
		Median		4.00	
		Variance		1.361	
		Std. Deviation		1.166	
		Minimum		1	
		Maximum		5	
		Range		4	
		Interquartile Range		2	
		Skewness		-.640	.172
		Kurtosis		-.610	.342
	Crowd	Mean		4.09	.080
		95% Confidence Interval for Mean	Lower Bound	3.93	
			Upper Bound	4.25	
		5% Trimmed Mean		4.21	
		Median		4.00	
		Variance		1.268	
		Std. Deviation		1.126	
		Minimum		1	
		Maximum		5	
		Range		4	
		Interquartile Range		2	
		Skewness		-1.181	.172
		Kurtosis		.629	.342

Appendix I: SPSS Output MANOVA and ANOVA analyses

Between-Subjects Factors

		Value Label	N
Idea Type	1	ChatGPT	100
	2	Crowd	100

Descriptive Statistics

	Idea Type	Mean	Std. Deviation	N
Mean Novelty	ChatGPT	3.550	.7124	100
	Crowd	3.645	.6678	100
	Total	3.597	.6904	200
Mean Feasibility	ChatGPT	3.760	.6092	100
	Crowd	3.930	.6854	100
	Total	3.845	.6524	200
Mean Customer Benefit	ChatGPT	3.810	.7170	100
	Crowd	4.090	.8208	100
	Total	3.950	.7814	200

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.979	3013.164[b]	3.000	196.000	<.001	.979
	Wilks' Lambda	.021	3013.164[b]	3.000	196.000	<.001	.979
	Hotelling's Trace	46.120	3013.164[b]	3.000	196.000	<.001	.979
	Roy's Largest Root	46.120	3013.164[b]	3.000	196.000	<.001	.979
IdeaType	Pillai's Trace	.032	2.183[b]	3.000	196.000	.091	.032
	Wilks' Lambda	.968	2.183[b]	3.000	196.000	.091	.032
	Hotelling's Trace	.033	2.183[b]	3.000	196.000	.091	.032
	Roy's Largest Root	.033	2.183[b]	3.000	196.000	.091	.032

a. Design: Intercept + IdeaType

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Mean Novelty	.451[a]	1	.451	.947	.332	.005
	Mean Feasibility	1.445[b]	1	1.445	3.437	.065	.017
	Mean Customer Benefit	3.920[c]	1	3.920	6.601	.011	.032
Intercept	Mean Novelty	2588.401	1	2588.401	5429.206	<.001	.965
	Mean Feasibility	2956.805	1	2956.805	7032.401	<.001	.973
	Mean Customer Benefit	3120.500	1	3120.500	5254.797	<.001	.964
IdeaType	Mean Novelty	.451	1	.451	.947	.332	.005
	Mean Feasibility	1.445	1	1.445	3.437	.065	.017
	Mean Customer Benefit	3.920	1	3.920	6.601	.011	.032
Error	Mean Novelty	94.398	198	.477			
	Mean Feasibility	83.250	198	.420			
	Mean Customer Benefit	117.580	198	.594			
Total	Mean Novelty	2683.250	200				
	Mean Feasibility	3041.500	200				
	Mean Customer Benefit	3242.000	200				
Corrected Total	Mean Novelty	94.849	199				
	Mean Feasibility	84.695	199				
	Mean Customer Benefit	121.500	199				

a. R Squared = ,005 (Adjusted R Squared = ,000)

b. R Squared = ,017 (Adjusted R Squared = ,012)

c. R Squared = ,032 (Adjusted R Squared = ,027)

Appendix J: SPSS Output Pearson Correlations

Correlations

		Mean Novelty	Mean Feasibility	Mean Customer Benefit
Mean Novelty	Pearson Correlation	1	.385**	.384**
	Sig. (2-tailed)		<.001	<.001
	N	203	200	200
Mean Feasibility	Pearson Correlation	.385**	1	.756**
	Sig. (2-tailed)	<.001		<.001
	N	200	200	200
Mean Customer Benefit	Pearson Correlation	.384**	.756**	1
	Sig. (2-tailed)	<.001	<.001	
	N	200	200	200

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix K: SPSS Output MANCOVA and ANCOVA analyses

Descriptive Statistics

	Idea Type	Mean	Std. Deviation	N
Mean Novelty	ChatGPT	3.550	.7124	100
	Crowd	3.645	.6678	100
	Total	3.597	.6904	200
Mean Feasibility	ChatGPT	3.760	.6092	100
	Crowd	3.930	.6854	100
	Total	3.845	.6524	200
Mean Customer Benefit	ChatGPT	3.810	.7170	100
	Crowd	4.090	.8208	100
	Total	3.950	.7814	200

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.895	553.532[b]	3.000	195.000	<.001	.895
	Wilks' Lambda	.105	553.532[b]	3.000	195.000	<.001	.895
	Hotelling's Trace	8.516	553.532[b]	3.000	195.000	<.001	.895
	Roy's Largest Root	8.516	553.532[b]	3.000	195.000	<.001	.895
WordCount	Pillai's Trace	.027	1.774[b]	3.000	195.000	.153	.027
	Wilks' Lambda	.973	1.774[b]	3.000	195.000	.153	.027
	Hotelling's Trace	.027	1.774[b]	3.000	195.000	.153	.027
	Roy's Largest Root	.027	1.774[b]	3.000	195.000	.153	.027
IdeaType	Pillai's Trace	.032	2.149[b]	3.000	195.000	.095	.032
	Wilks' Lambda	.968	2.149[b]	3.000	195.000	.095	.032
	Hotelling's Trace	.033	2.149[b]	3.000	195.000	.095	.032
	Roy's Largest Root	.033	2.149[b]	3.000	195.000	.095	.032

a. Design: Intercept + WordCount + IdeaType

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Mean Novelty	.779[a]	2	.389	.815	.444	.008
	Mean Feasibility	3.554[b]	2	1.777	4.314	.015	.042
	Mean Customer Benefit	5.013[c]	2	2.506	4.239	.016	.041
Intercept	Mean Novelty	450.821	1	450.821	944.101	<.001	.827
	Mean Feasibility	550.423	1	550.423	1336.352	<.001	.872
	Mean Customer Benefit	561.408	1	561.408	949.439	<.001	.828
WordCount	Mean Novelty	.327	1	.327	.686	.409	.003
	Mean Feasibility	2.109	1	2.109	5.120	.025	.025
	Mean Customer Benefit	1.093	1	1.093	1.848	.176	.009
IdeaType	Mean Novelty	.437	1	.437	.915	.340	.005
	Mean Feasibility	1.381	1	1.381	3.353	.069	.017
	Mean Customer Benefit	3.843	1	3.843	6.499	.012	.032
Error	Mean Novelty	94.070	197	.478			
	Mean Feasibility	81.141	197	.412			
	Mean Customer Benefit	116.487	197	.591			
Total	Mean Novelty	2683.250	200				
	Mean Feasibility	3041.500	200				
	Mean Customer Benefit	3242.000	200				
Corrected Total	Mean Novelty	94.849	199				
	Mean Feasibility	84.695	199				
	Mean Customer Benefit	121.500	199				

a. R Squared = ,008 (Adjusted R Squared = -,002)

b. R Squared = ,042 (Adjusted R Squared = ,032)

c. R Squared = ,041 (Adjusted R Squared = ,032)

Appendix L: SPSS Output Post-hoc Analysis

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.611	50.253[b]	3.000	96.000	<.001	.611
	Wilks' Lambda	.389	50.253[b]	3.000	96.000	<.001	.611
	Hotelling's Trace	1.570	50.253[b]	3.000	96.000	<.001	.611
	Roy's Largest Root	1.570	50.253[b]	3.000	96.000	<.001	.611
ExternalMotivation	Pillai's Trace	.035	1.164[b]	3.000	96.000	.327	.035
	Wilks' Lambda	.965	1.164[b]	3.000	96.000	.327	.035
	Hotelling's Trace	.036	1.164[b]	3.000	96.000	.327	.035
	Roy's Largest Root	.036	1.164[b]	3.000	96.000	.327	.035

a. Design: Intercept + ExternalMotivation

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Mean Feasibility	.152[a]	1	.152	.321	.572	.003
	Mean Novelty	1.174[b]	1	1.174	2.677	.105	.027
	Mean Customer Benefit	.917[c]	1	.917	1.367	.245	.014
Intercept	Mean Feasibility	47.505	1	47.505	100.424	<.001	.506
	Mean Novelty	50.475	1	50.475	115.107	<.001	.540
	Mean Customer Benefit	59.562	1	59.562	88.746	<.001	.475
ExternalMotivation	Mean Feasibility	.152	1	.152	.321	.572	.003
	Mean Novelty	1.174	1	1.174	2.677	.105	.027
	Mean Customer Benefit	.917	1	.917	1.367	.245	.014
Error	Mean Feasibility	46.358	98	.473			
	Mean Novelty	42.974	98	.439			
	Mean Customer Benefit	65.773	98	.671			
Total	Mean Feasibility	1591.000	100				
	Mean Novelty	1372.750	100				
	Mean Customer Benefit	1739.500	100				
Corrected Total	Mean Feasibility	46.510	99				
	Mean Novelty	44.148	99				
	Mean Customer Benefit	66.690	99				

a. R Squared = ,003 (Adjusted R Squared = -,007)

b. R Squared = ,027 (Adjusted R Squared = ,017)

c. R Squared = ,014 (Adjusted R Squared = ,004)

Appendix M: Lists of Top-rated Ideas

The following lists are unordered, as all ice cream flavors within each list were given the same top score.

Appendix M1: Top Rated Ideas of Both Raters Combined

	Flavor code	Novelty	Feasibility	Customer Benefit	Idea Quality	Idea Generation Type
1.	Flavor 70	4.5	4.5	5	14	ChatGPT
2.	Flavor 73	5	4	5	14	Crowd
3.	Flavor 89	4.5	4.5	5	14	ChatGPT
4.	Flavor 104	4.5	4.5	5	14	Crowd
5.	Flavor 136	4.5	4.5	5	14	Crowd
6.	Flavor 137	4.5	4.5	5	14	Crowd
7.	Flavor 145	4	5	5	14	Crowd
8.	Flavor 151	4.5	4.5	5	14	Crowd
9.	Flavor 155	4.5	4.5	5	14	Crowd
10.	Flavor 162	4	5	5	14	Crowd
11.	Flavor 169	4.5	4.5	5	14	Crowd
12.	Flavor 170	4.5	4.5	5	14	ChatGPT
13.	Flavor 177	4.5	4.5	5	14	ChatGPT
14.	Flavor 185	4	5	5	14	ChatGPT

Appendix M2: Top Rated Ideas by De IJswinckel

	Flavor code	Novelty	Feasibility	Customer Benefit	Idea Quality	Idea Generation Type
1.	Flavor 25	5	5	5	15	Crowd
2.	Flavor 65	5	5	5	15	Crowd
3.	Flavor 71	5	5	5	15	ChatGPT
4.	Flavor 70	5	5	5	15	ChatGPT
5.	Flavor 73	5	5	5	15	Crowd
6.	Flavor 74	5	5	5	15	Crowd
7.	Flavor 89	5	5	5	15	ChatGPT
8.	Flavor 90	5	5	5	15	Crowd
9.	Flavor 92	5	5	5	15	Crowd
10.	Flavor 103	5	5	5	15	Crowd
11.	Flavor 104	5	5	5	15	Crowd
12.	Flavor 129	5	5	5	15	ChatGPT
13.	Flavor 136	5	5	5	15	Crowd
14.	Flavor 137	5	5	5	15	Crowd
15.	Flavor 145	5	5	5	15	Crowd
16.	Flavor 155	5	5	5	15	Crowd
17.	Flavor 162	5	5	5	15	Crowd
18.	Flavor 168	5	5	5	15	Crowd
19.	Flavor 169	5	5	5	15	Crowd
20.	Flavor 170	5	5	5	15	ChatGPT
21.	Flavor 177	5	5	5	15	ChatGPT
22.	Flavor 179	5	5	5	15	Crowd
23.	Flavor 180	5	5	5	15	ChatGPT
24.	Flavor 185	5	5	5	15	ChatGPT
25.	Flavor 193	5	5	5	15	ChatGPT

Appendix M3: Top Rated Ideas by IJssalon Kees

	Flavor code	Novelty	Feasibility	Customer Benefit	Idea Quality	Idea Generation Type
1.	Flavor 4	4	4	5	13	ChatGPT
2.	Flavor 5	3	5	5	13	ChatGPT
3.	Flavor 6	5	3	5	13	ChatGPT
4.	Flavor 8	4	4	5	13	Crowd
5.	Flavor 9	4	4	5	13	Crowd
6.	Flavor 10	4	4	5	13	ChatGPT
7.	Flavor 12	4	4	5	13	Crowd
8.	Flavor 13	4	4	5	13	Crowd
9.	Flavor 16	4	4	5	13	Crowd
10.	Flavor 18	4	4	5	13	ChatGPT
11.	Flavor 21	4	4	5	13	Crowd
12.	Flavor 28	4	4	5	13	ChatGPT
13.	Flavor 30	4	4	5	13	Crowd
14.	Flavor 33	4	4	5	13	ChatGPT
15.	Flavor 34	4	4	5	13	ChatGPT
16.	Flavor 35	4	4	5	13	Crowd
17.	Flavor 37	4	4	5	13	Crowd
18.	Flavor 38	4	4	5	13	ChatGPT
19.	Flavor 39	4	4	5	13	Crowd
20.	Flavor 42	4	4	5	13	ChatGPT
21.	Flavor 44	3	5	5	13	Crowd
22.	Flavor 51	3	5	5	13	Crowd
23.	Flavor 53	4	4	5	13	ChatGPT
24.	Flavor 54	4	4	5	13	ChatGPT
25.	Flavor 55	4	4	5	13	ChatGPT
26.	Flavor 56	4	4	5	13	Crowd
27.	Flavor 58	4	4	5	13	ChatGPT
28.	Flavor 60	3	5	5	13	Crowd
29.	Flavor 61	4	4	5	13	Crowd
30.	Flavor 62	4	4	5	13	Crowd
31.	Flavor 64	4	4	5	13	ChatGPT
32.	Flavor 67	4	4	5	13	Crowd
33.	Flavor 68	4	4	5	13	Crowd
34.	Flavor 70	4	4	5	13	ChatGPT
35.	Flavor 72	4	4	5	13	Crowd
36.	Flavor 73	5	3	5	13	Crowd
37.	Flavor 75	3	5	5	13	ChatGPT
38.	Flavor 80	3	5	5	13	Crowd
39.	Flavor 81	4	4	5	13	ChatGPT
40.	Flavor 82	3	5	5	13	ChatGPT
41.	Flavor 83	3	5	5	13	Crowd
42.	Flavor 84	4	4	5	13	Crowd
43.	Flavor 87	3	5	5	13	ChatGPT
44.	Flavor 88	4	4	5	13	ChatGPT

Flavor Code	Novelty	Feasibility	Customer Benefit	Idea Quality	Idea Generation Type
45. Flavor 89	4	4	5	13	ChatGPT
46. Flavor 91	4	4	5	13	Crowd
47. Flavor 93	3	5	5	13	ChatGPT
48. Flavor 94	3	5	5	13	ChatGPT
49. Flavor 95	4	4	5	13	ChatGPT
50. Flavor 97	4	4	5	13	Crowd
51. Flavor 98	4	4	5	13	Crowd
52. Flavor 104	4	4	5	13	Crowd
53. Flavor 105	4	4	5	13	ChatGPT
54. Flavor 110	3	5	5	13	Crowd
55. Flavor 111	4	4	5	13	ChatGPT
56. Flavor 113	3	5	5	13	ChatGPT
57. Flavor 114	4	4	5	13	Crowd
58. Flavor 115	4	4	5	13	Crowd
59. Flavor 117	4	4	5	13	Crowd
60. Flavor 118	4	4	5	13	ChatGPT
61. Flavor 120	4	4	5	13	Crowd
62. Flavor 121	4	4	5	13	Crowd
63. Flavor 122	4	4	5	13	Crowd
64. Flavor 123	4	4	5	13	ChatGPT
65. Flavor 124	4	4	5	13	ChatGPT
66. Flavor 125	4	4	5	13	Crowd
67. Flavor 126	4	4	5	13	Crowd
68. Flavor 127	3	5	5	13	Crowd
69. Flavor 130	4	4	5	13	Crowd
70. Flavor 131	4	4	5	13	Crowd
71. Flavor 134	4	4	5	13	ChatGPT
72. Flavor 136	4	4	5	13	Crowd
73. Flavor 137	4	4	5	13	Crowd
74. Flavor 139	4	4	5	13	Crowd
75. Flavor 140	3	5	5	13	Crowd
76. Flavor 142	4	4	5	13	ChatGPT
77. Flavor 143	4	4	5	13	ChatGPT
78. Flavor 144	4	4	5	13	Crowd
79. Flavor 145	4	4	5	13	Crowd
80. Flavor 146	4	4	5	13	ChatGPT
81. Flavor 148	4	4	5	13	ChatGPT
82. Flavor 149	4	4	5	13	Crowd
83. Flavor 151	4	4	5	13	Crowd
84. Flavor 152	4	4	5	13	ChatGPT
85. Flavor 154	3	5	5	13	ChatGPT
86. Flavor 155	4	4	5	13	Crowd
87. Flavor 159	3	5	5	13	ChatGPT
88. Flavor 162	3	5	5	13	Crowd
89. Flavor 163	3	5	5	13	Crowd
90. Flavor 166	4	4	5	13	ChatGPT

Flavor code	Novelty	Feasibility	Customer Benefit	Idea Quality	Idea Generation Type
91. Flavor 167	3	5	5	13	Crowd
92. Flavor 169	4	4	5	13	Crowd
93. Flavor 170	4	4	5	13	ChatGPT
94. Flavor 171	3	5	5	13	ChatGPT
95. Flavor 174	4	4	5	13	ChatGPT
96. Flavor 176	4	4	5	13	ChatGPT
97. Flavor 177	4	4	5	13	ChatGPT
98. Flavor 178	4	4	5	13	Crowd
99. Flavor 182	3	5	5	13	ChatGPT
100. Flavor 184	3	5	5	13	ChatGPT
101. Flavor 185	3	5	5	13	ChatGPT
102. Flavor 186	4	4	5	13	ChatGPT
103. Flavor 187	4	4	5	13	ChatGPT
104. Flavor 188	3	5	5	13	ChatGPT
105. Flavor 189	3	5	5	13	ChatGPT
106. Flavor 191	3	5	5	13	ChatGPT
107. Flavor 194	4	4	5	13	ChatGPT