Business Analytics Project Report

Academic Year 2024/2025

D'Ambrosio Luca, Filippi Edoardo, Liani Tancredi, Luque Ochoa Santiago Andres, Negri Giacomo

Students' ID: 3182051, 3136295, 3319819, 3297703, 3155287
Group number: 5

Contents

Index of figures:

Index of tables:

Introduction

Navigating uncertainty is at the core of effective decision-making, particularly in today's complex and fast-evolving business landscape. This report explores how structured theory-building, data analysis, and experimentation can guide strategic decisions, drawing on real-world examples and experimental findings. By combining practical cases and theoretical insights, it demonstrates the value of understanding the nuanced relationships between time, complexity, and decision outcomes.

The report begins addressing two key cases: Google's strategic approach to entering the online video market in the mid-2000s and Booking.com's data-driven decision-making processes to completely change the hospitality industry. These cases highlight how organizations can leverage a theory-based approach to evaluate potential strategies, weight risks, and identify opportunities for long-term growth and competitive advantage.

Building on these examples, the report further investigates two core directions in decision-making. The first explores the non-linear relationship between time, complexity, and success probabilities, revealing a "sweet spot" where performance peaks before diminishing returns set in. The second examines how time investment and environmental factors, signals, influence the willingness to adapt decisions based on new information, emphasising the importance of flexibility in uncertain environments.

Through a carefully designed experiment and survey, the report tests these theories, focusing on how the amount of information available impacts problem-solving success. The findings confirm the dual challenges of insufficient information and "information overload", underscoring the importance of striking the right balance to optimize outcomes.

This report combines theoretical exploration, experimental insights, and practical applications to illustrate how businesses can make more informed, adaptable, and effective decisions. It provides insights for understanding and applying theory-driven frameworks to achieve success in an unpredictable world.

Theory Building

Our thought process to address the theory-building task adhered to the following framework:

1) Situate ourselves in the historic context in which the decision problem took place

Given both cases are historic, during this phase we highlighted the fact that many of the widespread tools of today were not present at the time the decision problem took place, or at least not in their current form. This included mobile applications and the internet, among others. Additionally, historical trends were brought to the table, such as the tech boom of 2007 (Gurdus, 2016).

2) List the attributes depending on how significant they were for the accomplishment of the end state

By means of discussion and questioning, we argued which were the most significant attributes for the achievement of the end state. For each argument proposed by a member of the group, the others had the task of questioning his reasoning and the soundness of his arguments. This approach mimics the more notorious Devil's Advocate technique. In the end, the attributes who "survived" were taken into consideration for building the theory and the links between them and the end state.

3) Identify causal relationships between attributes

Among all the attributes at disposal, we identified plausible causal relationships. The purpose of this task was to work on the complexity of our theory, considering that we believed a certain level of complexity is beneficial to the robustness of a theory. Contemporarily, we wanted to avoid overfitting our model and overcomplicating our theory, because these could make it harder to trace the cause-and-effect relationship once experiments are done or the theory is implemented. In the following sections we will explain how the framework presented was applied to each case.

Google Case

In 2006, Eric Schmidt, CEO of Google, faced a crucial strategic decision in an environment particularly fertile for technological innovation. The rapid exponential growth of technology and the increasing use of the internet created an ideal context for expanding Google's offerings. Schmidt had to decide how to position Google in the emerging field of online video, a rapidly growing sector thanks to platforms like YouTube, which allowed anyone to upload and share video content.

In this promising context, Schmidt considered two possible directions for Google: to internally develop an "all-media ecosystem" by integrating video across Google's various services, collaborating with media partners without producing original content, or to acquire YouTube to leverage its already established user base and expertise in the sector. The "Google All-Media Ecosystem Theory" suggested that, given the growing spread of online video, Google could offer a comprehensive service where users could watch, share, and search for videos, integrating them with the other services provided by the company.

Initially, we situated ourselves in the historical context of the decision problem. During the process, we debated about the idea of quality and how it changed from 2006 until today: today, videos are fast to load, high quality (up to 4K), which often include subtitles, while in 2006 a high-quality video may have been conceived as a video that included some sort of basic effect and generic music in the background. As a result, we argued that, given the historical period, high initial probabilities were meant to be assigned to each attribute thanks to the vibrant environment of 2006. We were convinced that the momentum of the tech boom was a strong proof that the attributed were occurring. For the "All-Media Ecosystem Theory", the lack of many competitors in the market and rapidly evolving technological potentials bolstered confidence that Google, with the right mix of technology and strategy, could emerge as a dominant multimedia platform This approach aimed at taking advantage of a still-expanding market allowed Google to create long-term value by capitalising on favourable conditions of the time.

Subsequently, the acquisition of YouTube was evaluated as an opportunity to accelerate growth in the online video sector. YouTube, with its vast user base and strong engagement capacity, represented a springboard for entering this expanding market. Schmidt viewed the acquisition as a strategic investment that, despite the high cost, could bring significant long-term advantages, since implementing Google's search engine into YouTube could lead to a solid return.

Google intended to leverage the synergies between the two platforms: by integrating YouTube with its search and advertising technologies, it could improve targeting for advertisers and offer a complete multimedia experience for users. However, Schmidt was aware of the cultural and technological integration challenges that such a large acquisition would entail. Therefore, we identified causal relationships from the success of cultural integration and user data synergies: if teams work better together, they are more prone to produce successful advancements such as data synergies.

In a growing competitive context, with players like Facebook and Microsoft, Schmidt believed that acquiring YouTube could provide Google with a strategic advantage, transforming it into a dominant multimedia platform and positioning it for sustainable long-term growth. We decided that the most important attributes were Competitive Edge and User Data Synergies, given the prior knowledge we had about the potential of algorithms in providing tailored advertisement to users. These attributes were, in fact, the ones with the most evident and strong causality regarding the end state.

Moreover, the possible effectiveness of the acquisition was evaluated through another theory, where the final probability of accelerated growth appeared high, as the only potential obstacle could be the merging of the two corporate cultures, given that, according to many sources (e.g., Financier Worldwide, 2022) 80% of the success of M&As depends on this aspect. On the other hand, the acquisition could give Google a significant competitive edge over Facebook and Microsoft.

Finally, experiments were conducted on both theories. Although the "All-Media Ecosystem" theory had a higher degree of uncertainty, we decided that conducting an experiment on both was beneficial as the difference between the expected probability of the theories was not large. In the end, the experiment provided a larger signal for the "Accelerated Growth" theory, and we chose it as our final one.

Booking Case

In the late 1990s, Geert-Jan Bruinsma, a university student, faced a critical challenge while planning a holiday: booking accommodations was cumbersome, with fragmented offerings, inconsistent standards, and limited transparency. This problem was particularly pronounced in Europe, where most hotels were independent and lacked uniformity, unlike the standardized chains in the United States. Bruinsma saw an opportunity in the emerging Internet to address these issues and simplify the process for both travellers and hotel owners.

At the time, the hospitality industry was dominated by offline travel agencies that acted as intermediaries, leveraging the information asymmetry between hotels and travellers. Bruinsma envisioned an online platform, "Booking.nl," where users could search for and book accommodations seamlessly, while hotels could advertise their rooms efficiently. Competitors such as Expedia also began offering online booking, but their platforms lacked key features, such as detailed descriptions and photos of hotels, which Bruinsma sought to provide.

The timing was critical, as the Internet was beginning to disrupt traditional industries, yet both travellers and hotel owners were hesitant to move away from offline methods like brochures and phone bookings. Given this background, Bruinsma had two choices: develop an “Online Marketing Space” or a “Travel Platform”.

For the first choice, and as part of the first phase of our framework, it was important to remark that the internet was not a widespread technology. It was important to remember that, as it was a novelty, hotels were probably not willing to spend enormous amounts of money on advertising in an empty space such as the internet. Additionally, as part of the second phase, we all agreed that the adoption of the internet was going to occur with a high probability, given the vast benefits and minimal backlash it promised; this fact guided our assignment of a high probability to this attribute (and the link between it and the end state) and lower probabilities for the rest.

For the second choice, a similar approach followed. We recognized that at the time, online payments were not common: neither for the everyday person nor for the hotels’ owner, and we identified that this was probably the most important attribute of all, followed by online purchasing habits, and engagement of users and hotels. The assignment of probabilities for the links between these attributes and the end state followed this descending order of importance; higher probabilities were assigned to the most important attributes.

Causal relationships were stablished subsequently. In a unanimous way we agreed that online purchasing habits impacted the probability of users engaging with the platform. Additionally, we noted that there might exist a bidirectional causal relationship between user and hotel engagement, but through discussion we concluded that in this case hotel engagement was more probable to impact user engagement and not vice versa.

Finally, we experimented on the first theory, “Online Marketing Space”, as it was the most uncertain given the probabilities and links we assigned. The experiment did not provide a signal strong enough for us to drastically change our minds about the structure, probabilities, and links assigned. For this reason, we chose the “Travel Platform” as our final one.

Data Analysis

The analysis presented in this section will concern the users’ behaviour during different sessions in which they play with the platform Plato of the Bocconi IMSL. Two datasets were shared: game_headers.csv and game_records.csv.

Data Preparation and Cleaning

Approach

- **Initial phase:** we first cleaned the dataset and extracted the necessary information. Then, we began by exploring how variations in time and complexity influenced outcomes, categorising results as either success (winner) or failure (loser). This phase involved analysing discrete statistics, creating correlation matrices, and visualising data through graphs to uncover trends and direct our theoretical evaluation.

- **Removing noise**: a critical cut-off point for the time required to complete the task was established based on data and hands-on experience. Below this threshold, the results were deemed unreliable or misleading, making their exclusion necessary to avoid introducing noise into the analysis.

Implementation

- The dataset was refined by extracting the necessary variables from the experiment JSON file, cleaning the data, and addressing inconsistencies. Approximately 80 observations were removed because participants left the first theory blank, suggesting a lack of active participation in theory development or hypothesis testing. These participants had an average game completion time of 8 minutes and 7 seconds (486.64 seconds), which is considered unusually low for this task. Moreover, their results could have impaired the ability to evaluate with sounding reason the result, since they would have contaminated the real effect that time, attributes, and complex thinking have on the outcome. Consequently, their theories and results were excluded from the analysis. After these adjustments, the final dataset comprised 452 observations.

First Analysis

Theory

The relationship between time, complexity, and the likelihood of success when addressing a decision problem is non-linear. Initially, the probability of success improves as time and complexity increase, but this trend reverses beyond a critical threshold were excessive time and complexity lead to diminishing returns and decreased chances of success. The ability of an individual to handle complexity is influenced by three factors: the number of attributes they consider when constructing their theory, the connections they establish among these attributes, and the time they dedicate to analysing and reasoning through the problem.

Approach

- **Discrete and graphical analysis**: this phase involved analysing discrete statistics, creating correlation matrices, and visualising data through graphs to uncover trends and direct our theoretical evaluation.
- **Tree building:** we developed a classification tree to take some variables as inputs and to classify these features dependently on the result. To enhance robustness, each terminal node (leaf) of the tree was required to include some minimum number of observations. This minimum size helped mitigate overfitting and ensured meaningful classifications.
- **Analysis**: the final analysis of this theory was performed using a randomised classification tree model, which was designed to validate the insights and refine our understanding of the key factors influencing success.

Results

- **Discrete statistic**: discrete statistics were computed to evaluate the sample. Notably, one important observation is that the distribution of winners is relatively balanced, with 67% of the sample being winners after data cleaning.
- **Correlation matrix**: a correlation matrix was constructed to examine the relationships between various variables.

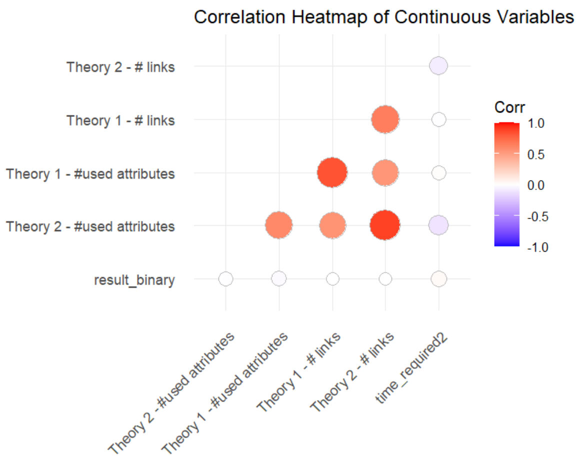

Figure 1: Correlation Heatmap of Continuous Variables

The matrix in Figure 1 highlights several possible patterns in how participants develop and revise their theories. Notably, Theory 1 and Theory 2 exhibit strong positive correlations in both the number of links and attributes used. This suggests that participants tend to maintain similar levels of complexity across the two theories, considering also that an increase in attributes inherently leads to a higher number of links. While there is a positive correlation between the number of attributes and links within each theory, the time required to build the theories shows a slight negative correlation with the number of attributes used in Theory 2. This could imply that as participants gain familiarity with the situation and tools, they become more efficient in their theory development. Interestingly, the binary outcome variable (result_binary) shows weak correlations with all other variables, reinforcing the idea that success is more likely when participants aim at an optimal balance among the various factors, rather than maximising the number of elements considered for their theory-building process.

- **Graphs:** three key graphs were created to assess the relationship between variables and the result. The x-axis of each graph represents one of the following variables: the number of attributes (total attributes used), the number of links (total links created), and play time (time spent completing the game).
    - **Total attributes used:** in the first graph we can observe that:

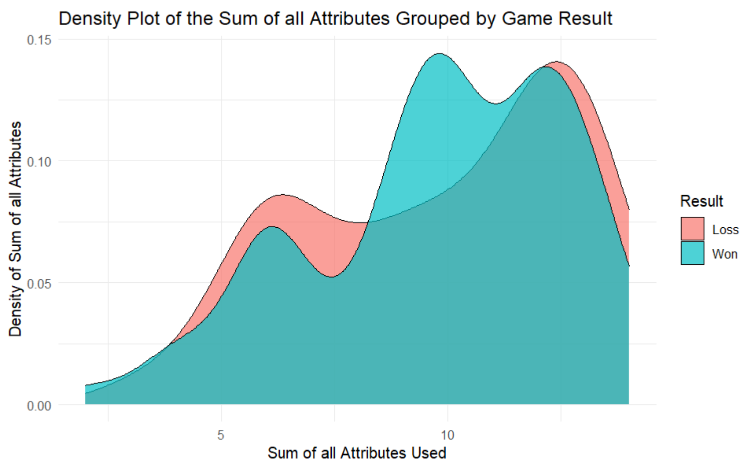

Figure 2: Density plot of the sum of all attributes grouped by game result

the distribution of attributes used shows optimal success rates when players incorporate 9 to 11 attributes into their analysis. Interestingly, those who lost tend to choose either a very small number of attributes or a very large one, while successful player tends to concentrate around the "sweet spot", which falls right in the middle of the groove left by the losers' attributes. This seems to suggest that selecting 6 to 8 attributes is an awkward middle ground where the model is neither simple enough to be manageable nor complex enough to be comprehensive. The data indicates that successful problem solving requires considering a substantial number of relevant attributes, but exceeding about 13 attributes begins to negatively impact success rates, likely due to information and variables overload.
As proof that excluding the previously mentioned empty theory elements was necessary, is that without it analyses would have had result as follows:

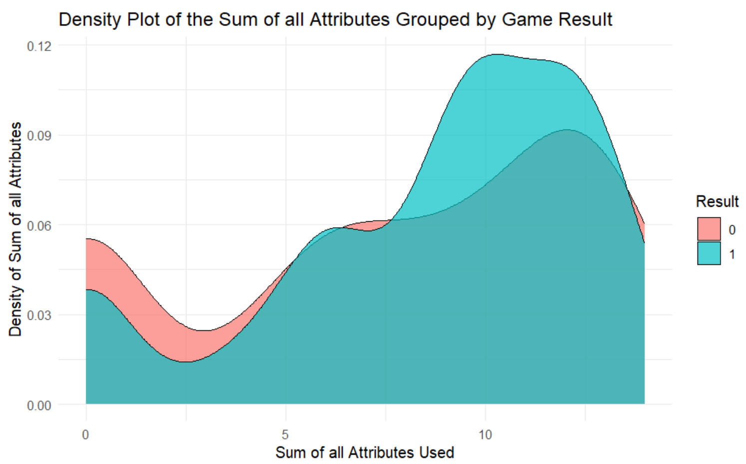

Figure 3: Density plot of the sum of all attributes grouped by game result

As Figure 3 shows, the actual sweet spot would have been impossible to find, and how attributes and user's ability really influence the outcome would've been an unanswered question. Since we're dealing with a binary outcome, without excluding those who left blank theory one, the best guess is, on average, 50% correct.

- **Total links created:** the density plot of links created follows a similar pattern, highlighting an optimal range between 8 and 10 links, where players achieved the highest success rates. This suggests that effective problem-solving requires a sufficient level of interconnection between concepts, without becoming overly complex. Success rates drop significantly when players create more than 15 links, likely due to cognitive overload and unnecessary complexity. Conversely, fewer than 3 links are associated with higher failure rates, indicating that overly simplified models fail to capture the complexity of the situation.

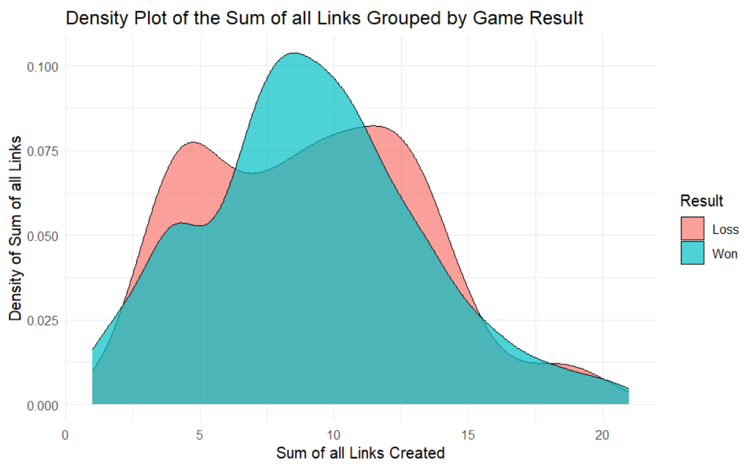

Figure 4: Density plot of the sum of all links grouped by game result

- **Play time:** the time distribution shows that most successful outcomes occur between 20 and 40 minutes, with a peak around 25 minutes. This "golden window" represents the optimal balance between thorough analysis and efficient problem-solving. Success rates sharply decline after 45 minutes, suggesting that extended time may reflect difficulties with the task or counterproductive overthinking, as indicated by the unusual red heavy tail. Conversely, very short completion times (under 10 minutes) are associated with higher failure rates, implying inadequate engagement with the problem's complexity.

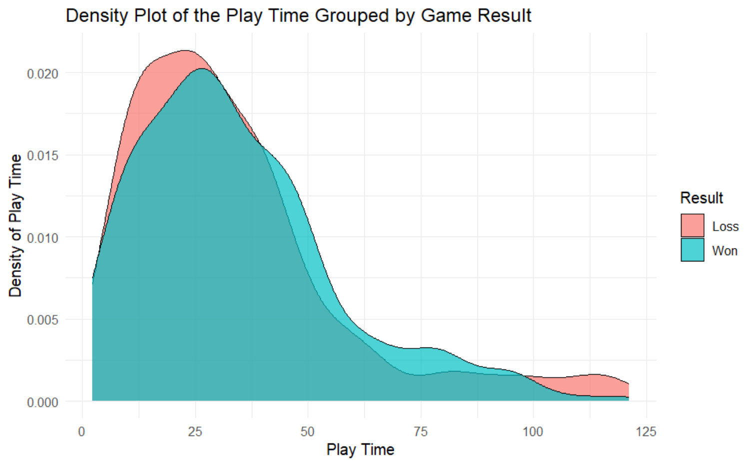

Figure 5: Density plot of the play time grouped by game result

- **Tree analysis:** the classification tree highlights key thresholds in time and complexity that impact success rates. Players who spent less than 25 minutes have a lower win rate (63%), suggesting insufficient time for thorough problem-solving. Success improves significantly (73%) when time exceeds this threshold, but further success depends on the number of connections. With fewer than 16 connections, players face high failure rates (44%), while success peaks (76%) when connections exceed 26, provided enough time is spent. Intermediate levels of complexity (16–26 connections) show diminished returns, emphasising that excessive or insufficient complexity hampers outcomes. This demonstrates that effective problem-solving requires both adequate time and an optimal level of complexity.

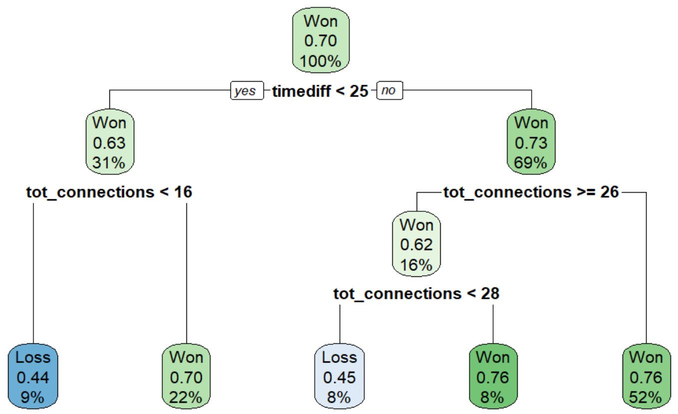

Figure 6: Classification tree

- **Generalization**: the findings underline the importance of balance in problem-solving. Success is not about maximising time or complexity but about finding the right combination of both. Players who rush or oversimplify fail to engage deeply, while those who overcomplicate without clear focus face inefficiencies. Practical thresholds, such as spending at least 25 minutes and maintaining connections within a meaningful range, suggest that balancing thoughtful problem design and manageability is critical for creating a successful theory.

Second Analysis

Theory

Based on intuition, it is possible to argue that the more time is spent on building something, the less prone one is to change it. Also, one could argue that the environment can influence one's decisions. This reasoning guided the development of the first theory: the time spent on developing a theory and the difference between the experimental result (signal) and the initial expected value influence the magnitude of the "adjustment" made to the theory after receiving the signal. A smaller difference between the signal experimental result and the initial expected value leads to smaller adjustments, while a larger discrepancy causes larger adjustments. With respect to time, as if time spent building a theory increases is large, the adjustment decreases is smaller, independently of the difference between the experimental result and the initial expected value, and vice versa. This could mean that the less time spent on building the theory, the less confident one is about it.

For convention, in the following sections "time spent developing a theory", "difference between the signal and the initial expected value of a theory", and "adjustment made to the theory after receiving the signal" will be referred to as "time", "signal difference", and "adjustment", respectively.

Approach

- **Integration of new variables:** for the purpose of this analysis, it was necessary to generate new variables. Therefore, in addition to the variables included by default in the *game_records.csv* dataset, the following variables were created to serve our analysis:
    - experiments_done: represents the experiments conducted by the group, including the theory on which they were conducted and the number of experiments.
    - th_experiment1: binary variable which takes the value of 1 if the first experiment was conducted on theory 1 and 2 if it was conducted on theory 2. If no experiments were conducted, then it takes de value "NA".
    - th_experiment2: binary variable which takes the value of 1 if the second experiment was conducted on theory 1 and 2 if it was conducted on theory 2. If no experiments were conducted, then it takes de value "NA".
    - adjustment ("adjustment"): difference between the expected probability of theory 1 after the experiment and the expected probability of theory 1 before the experiment. If the group lastly experimented on theory 2, then the variable takes the value of the difference between the expected probability of theory 2 after the last experiment and the expected probability of theory 2 before all experiments. This variable quantifies how much a theory's expected value was modified after experimenting on it.
    - signal ("signal difference"): difference between the drawn probability of the experiment and the expected probability of the theory before the experiment, the omega and alternative value were collected. This variable quantifies how

distant the initial expected probability of a theory was from the result of the experiment.
    - time_difference (“time”): difference in time between start_time and end_time. This variable indicates how much time took the group to perform the activity.
    - last_exp_value: theory’s last updated expected probability (prior to the omega and the expected value. This variable represents the same as the variable “Expected” in the JSON mapping of the Experiment.
    - fin_exp_draw: signal provided by the experiment (only final experiment’s signal is considered). This variable represents the same as the variable “Final exp draw” provided in the original dataset.
- **Exploratory data analysis (EDA):** with the goal of identifying trends of relationships between variables, EDA was performed on the dataset. This included plots for single, specific variables as well as plots that could illustrate the relationship between two or more of them.
- **Regression analysis:** incorporating the new variables mentioned above, a linear regression analysis was performed. Adjustment was regressed on time_difference and signal.

Results

- **Exploratory data analysis:** initially, the plotting of individual variables was performed to identify their distribution.

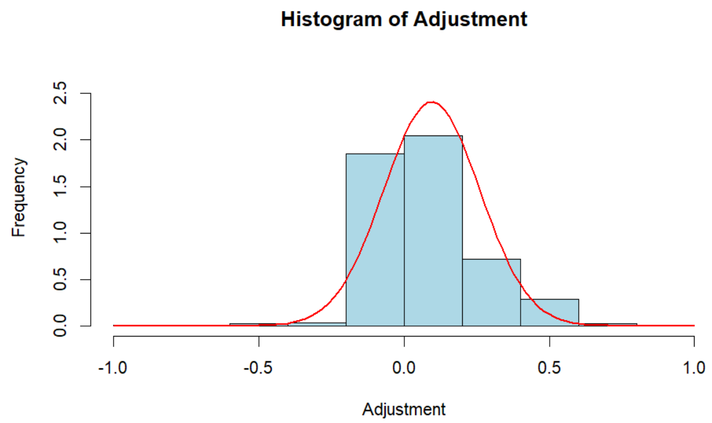

Figure 7: Histogram of the Adjustment

From Figure 7 it is observable that there are some values of the adjustment that are more common than others. At the same time, this variable is distributed around a slightly negative mean, which could indicate that groups, on average, are overconfident when constructing their theories and end up adjusting them to obtain a smaller expected probability.

With respect to signal difference, the initial hypothesis is again supported. Figure 8 seems to suggest that groups tend to develop theories with lower expected probability than that obtained through the experiments. This fact explains the positive Adjustment shown in Figure 7.

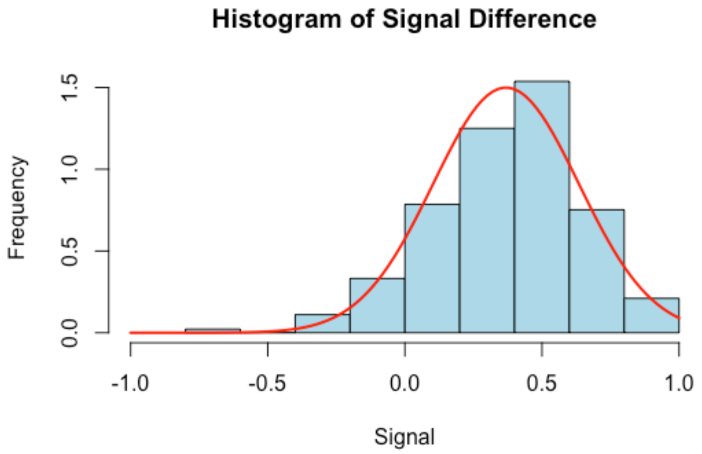

Figure 8: Histogram of the Signal

To support further analysis, a correlation matrix was constructed to examine the relationships between our variables of interest.

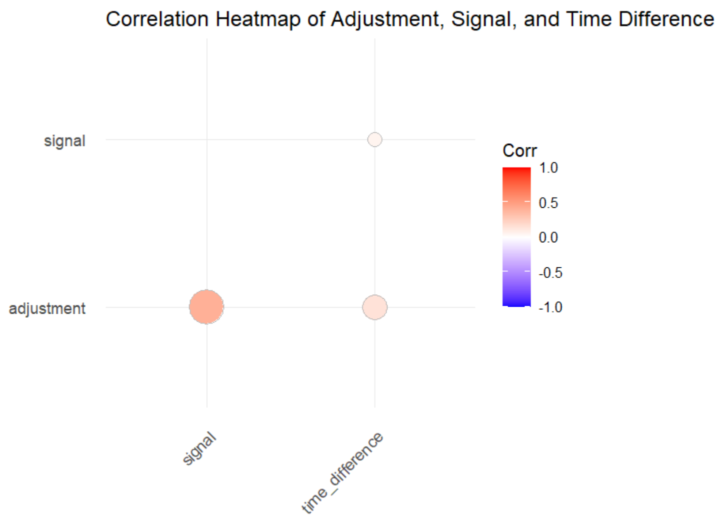

Figure 9: Correlation Heatmap of Continuous Variables

Figure 9 shows that there is a strong positive correlation between the signal difference and the adjustment, as well as a mild positive correlation between time and the adjustment.

The outcome of the regression analysis for the "adjustment" is presented in Table 1.

	Coefficient estimate	Std. Error	t value	P value	Significance[1]
Intercept	-2.868*10-1	1.921*10-2	-14.925	<2*10-16	***
Time	1.689*10-6	8.408*10-7	2.009	0.0451	*
Signal difference	5.629*10-1	4.234*10-2	13.296	<2*10-16	***

Table 1: Linear Regression of "Adjustment" on Time and Signal difference

As suggested by the correlation matrix, the coefficients of time and signal difference are positive. Both are significant at the 0.05 and 0.001 levels, respectively, which indicate they are good regressors for the adjustment made to the theory. This way, for every 1% variation in the signal difference, for example, between 44% and 45%, it is expected that the adjustment made to the theory will cause a 0.5629% variation to its expected probability. It is important to remark that a 0.5629% variation is intended to represent the variation observed, for example, between 60.0% and 60.5629%.

Experiment

Survey

Considering the first theory, an experiment was carried out. The proposed research question is as follows:

When solving a problem, is there a non-linear relationship between time spent solving it, amount of information available and the chance of success at solving the specified problem? If so, does this imply there is a "sweet spot" of information and time?

As emerged from the antecedent analysis, it is evident that a lack of information can negatively impact performance, but a similar effect has also been observed for the case of "information overload", where an excessive amount of information is detrimental to the chance of success. We decided to focus on the following question: how does the amount of information available influences performance?

To accomplish this task, a survey was designed in which participants had to solve an inferential problem. In this case, each participant had the task of answering the following question to the best of their abilities:

Imagine an Italian province 'X' with a population of about 350 000 people.
How many pharmacies are there in X?[2]

[1] Significance legend: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[2] In the English version there was also an additional phrase: "*As a reminder, Italy's administrative division is as follows: Country > Region > Province > Municipalities*".

Having defined the problem to solve an experimental design was developed as explained in the following section. It is important to note that in this context, “success” is defined as how close the respondent’s answer is to the actual number of pharmacies in province X: 140.

This question is quantitative but does not require any specific knowledge, allowing it to be answered objectively by anyone. It provides quantitative feedback on individual guesses and, more importantly, on how participants adjust their responses in relation to the information available. Moreover, the question has many accurate and verified details that refer to it, making it suitable for the survey. Additionally, using a real case like the province of Novara avoids creating misleading or biased values, while its status as a lesser-known Italian city means most respondents are unlikely to have prior knowledge about it.

Experiment Design

Section 1: Introduction

The initial section of the survey collects demographic information about the respondent, which include:

- Age Group
- Gender
- If the participant grew up in Italy or not
- Highest level of education completed
- Area of most recent area of study

The role of these variables is to work as controls for a balanced and representative treatment assignment. Interest is also destined to evaluate if the individual’s background and familiarity with the context of the problem might raise a significant effect on performance when solving a generic problem.

Section 2: First question

After answering demographic questions, all participants were faced again with the following question:

Imagine an Italian province 'X' with a population of about 350 000 people.
How many pharmacies are there in X?

This question collects participant’s initial guess, without any additional information other than the population of the province of interest. This question served as the “before” state of the groups. It is important to clarify that, since the problem participants faced was related to Italy, in the demographic-questions stage the question “Did you grow up in Italy?” was included to serve as control.

Section 3: Randomization

The three treatments were administered with equal probability, ensuring that approximately one-third of the participants in the survey received each treatment.

The information provided to the participants was randomly sampled from a set of twelve possible data points:

1. The population density is 271 people per km².
2. The region is in the north of Italy.
3. The number of (small and large) hospitals in the region is 92.
4. The number of municipalities in province X is 88.
5. The number of pharmacies in Italy is 20,079.
6. The area is 1340 km².
7. The number of public doctors in the region is 8,443.
8. The region has 4,256,350 residents.
9. Percentage of elderly population in the Region is 26,6%.
10. The annual public health expenditure in the region is €1,880 per resident.
11. 43% of the region's pharmacies are in rural areas (municipalities with fewer than 5,000 residents).
12. There are 99,000 pharmacists in Italy.

The three treatment conditions were structured as follows:

- **Treatment 1**: the participant received one randomly selected piece of information from the set of twelve.
- **Treatment 2**: the participant received five randomly selected pieces of information from the set of twelve.
- **Treatment 3**: the participant received the complete set of twelve pieces of information.

After being presented with the relevant information, participants were then asked to answer the second question.

Section 4: Second question

After responding to the initial question, all participants must address the following query:

Based on the new information provided, how many pharmacies do you estimate are in Province X?

To evaluate how the availability of information influences participants' reasoned estimates, we employ a randomized design. Participants are exposed to one of three different information treatments, administered to ensure variation in the number of information and the specific information provided.

Approach

The analysis employs a difference-in-differences regression model to evaluate how varying the amount of information (1, 5, or 12 pieces) affect participants' ability to estimate the number of pharmacies in an Italian province. The outcome variable is the absolute percentage difference from the correct answer (140 pharmacies), normalized by dividing by 140.

The variables considered are defined as follows:

- Treatment_5info: dummy variable for receiving 5 pieces of information
- Treatment_12info: dummy variable for receiving 12 pieces of information
- period_dummy: indicator for post-treatment period
- Base case: receiving 1 piece of information (captured in intercept).

Results

Sample balance

The demographic analysis of the sample does not report any unexpected result. Is highlight that most respondents fall within the 18–24 age range, representing the majority cohort, as illustrated in the histogram in the appendix 1. Smaller but notable groups are present in the 25–34 and 45–54 age ranges, while the youngest (<18) and oldest (65 and over) groups are minimally represented. Regarding nationality, a larger portion of the participants did not grow up in Italy ("No"), compared to those who did ("Yes"). As for education levels, most participants possess a bachelor's degree, followed by those with a master's degree in much smaller percentage. Participants with technical/professional education and high school education are less common, and those with a doctoral degree are rare. This demographic breakdown ensures a diverse yet predominantly young and academically educated participant pool, suitable for exploring the study's hypothesis on problem-solving efficiency.

Regression results

After performing a regression, the results obtained are presented in Tables 2 and 3.

	Coefficient	**Std. Error**	**t value**	**P**	**Significance[3]**
Intercept	15.0556	8.0255	1.876	0.0622	.
Treatment_5info	-14.3243	12.3155	-1.163	0.2463	
Treatment_12info	-12.8273	13.1383	-0.976	0.3301	
period_dummy	0.1616	11.3497	0.014	0.9887	
Treatment_5info*period_dummy	4.1569	17.4167	0.239	0.8116	
Treatment_12info*period_dummy	-1.3140	18.7013	-0.070	0.9441	

Table 2: Experiment regression's results

Minimum	1Q	Median	3Q	Maximum
-15.19	-14.00	-4.62	-0.63	483.94

Table 3: Experiment regression's residuals

[3] Significance legend: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The regression analysis reveals interesting patterns in how varying levels of information affect problem-solving accuracy, though with notable statistical limitations. The model's residuals range from -15.19 to 483.94, with a median of -4.62, indicating a right-skewed distribution of errors. This skewness suggests that while most participants' estimates clustered relatively close to the true value, some participants made substantially larger overestimates.

The intercept coefficient of 15.0556 ($p = 0.0622$) represents the baseline error rate for participants who received only one piece of information. This marginally significant result suggests that with minimal information, participants tend to overestimate the number of pharmacies by approximately 15% relative to the true value. The borderline significance level ($p < 0.1$) provides weak evidence that this baseline error is systematically different from zero.

When examining the treatment effects, participants who received five pieces of information (Treatment_5info) showed an initial reduction in error of -14.3243 percentage points compared to the baseline group, though this effect is not statistically significant ($p = 0.2463$). Similarly, those who received twelve pieces of information (Treatment_12info) demonstrated a reduction in error of -12.8273 percentage points, also not statistically significant ($p = 0.3301$). These coefficients suggest that additional information might help reduce initial estimation errors, though we cannot rule out that these apparent improvements occurred by chance.

The period_dummy coefficient (0.1616, $p = 0.9887$) indicates virtually no change in accuracy over time for the baseline group.

The values 4.1569 and -1.3140 can be interpreted respectively as the increase in percentage points of error after the treatment for those who had 5 and 12 information with respect to the baseline group. Since the p-values are not statistically significant (0.81 and 0.94) we can affirm that, as expected, the estimates are not statistically different from zero.

These are standardized factors but this that are reported are directional change, which is not clear whether close on or divert from the actual right answer (140).

The high maximum residual (483.94) compared to the relatively modest first and third quartiles (-14.05 and -0.69 respectively) indicates that while most participants made reasonable estimates, there were some extreme outliers that might warrant further investigation. These outliers could represent either genuine difficulty in processing larger amounts of information or might indicate issues with the experimental design or data collection process.

The overall pattern of results suggests that this effect is likely more nuanced. Future research might benefit from larger sample sizes, more controlled information presentation, or additional variables capturing individual differences in information processing capabilities.

Plato – Data:

Theory

This analysis investigates the factors influencing participants' theoretical updates after conducting experiments, focusing on their exposure to external shocks and the AI tool Aristotle. Specifically, it examines how changes in probability estimates, measured as V_diff (absolute magnitude of change) and V_diff_nonabs (directional change) are shaped by participants' cognitive stimulus and environmental factors.

Cognitive stimuli include variables such as exposure to Aristotle, familiarity with GPT, frequency of GPT usage, and participants' liking of algorithmic tools. Environmental factors: like the administered shock treatments, are also expected to play a significant role in shaping users' beliefs.

The central hypothesis is that external shocks will significantly affect directional updates (V_diff_nonabs), while cognitive stimulus will impact the absolute magnitude of change (V_diff), potentially reducing the overall impact of shocks, because of the user's possibility to interact with tools enabling the explanation of such shocks.

Approach

The analysis followed three primary steps:

- **Exploratory Data Analysis (EDA)**: the dataset's structure and summary statistics were examined, with significant distributions of numeric variables visualized through histograms and boxplots to identify potential group-level differences. Additionally, the presence of missing values was assessed.
- **Correlation analysis**: a correlation matrix was generated for numeric attributes, excluding identifiers and dependent variables such as group or Algorithmic_aversion paired with Algorithmic_liking. Relationships between variables like Knowledge_Depth and V_diff were also analysed to inform the design of the regression models.
- **Linear regression**: two regression models were developed. Model 1 predicted absolute changes in theoretical values (V_diff), while Model 2 focused on directional changes (V_diff_nonabs). Both models incorporated Shock and Aristotle as key predictors, alongside other control variables, and employed a linear modelling approach.

Result

- **EDA:** during the initial phase of exploration, it can be verified the distribution of various variables' values, being balance in the treatment assignment the most relevant one. The four groups have small and balanced in the number of individuals assigned to it, with values ranging from 15 to 17.

  The following table illustrates some discrete statistics.

Group	**mean_V1**	**mean_V2**	**mean_Vdiff**	**mean_SD1**	**mean_SD2**	**mean_SDdiff**
1	60.40000	67.06667	7.600000	5.133333	5.400000	0.2666667
2	69.05882	74.29412	8.058824	5.000000	5.176471	0.1764706
3	63.33333	74.40000	11.600000	4.266667	4.800000	0.5333333
4	56.25000	56.56250	7.312500	4.437500	4.750000	0.3125000

Table 4: Plato's dataset discrete statistics

Table 4 highlights notable differences in how treatment and shock influence participants' evaluation of theory success. The combination of treatment without shock (Group 3) resulted in the largest increase in the probability of success for participants' theories (mean V_diff = 11.6) and the highest change in variability (SD_diff = 0.53). This suggests that participants exposed to treatment without experiencing a shock may

have leveraged the provided tools to confirm their initial theories, potentially exhibiting confirmation bias. The increased variability indicates a significant gain in information that prompted participants to revise their models more extensively.
By contrast, the no treatment with shock (Group 2) group showed a moderate increase in theory change (V_diff = 8.06) but minimal variability change, implying that shock alone had a less pronounced and more uniform effect on theory updates. Participants might have adjusted their theories conservatively in response to the shock.
The no treatment/no shock (Group 1) group exhibited the smallest change in theory values (V_diff = 7.6) and the highest variability in both initial and updated theories, reflecting a lack of strong intervention to guide theory adjustments, leaving participants with diverse interpretations. Finally, treatment with shock (Group 4) produced a minor theory change (V_diff = 7.31) and a variability profile like Group 1, suggesting that participants in this group recalibrated their models effectively after the shock, leading to relatively stable evaluations.
These findings suggest that treatment, particularly in the absence of shock, played a significant role in shaping participants' assessments. However, when combined with a shock, treatment's impact appears tempered, likely due to the participants' ability to integrate the shock's influence into their evaluations. These dynamic underscores the nuanced interplay between cognitive bias and the external factors of treatment and shock.
The previous effects are visible also in the subsequent boxplot about V_diff_nonabs.

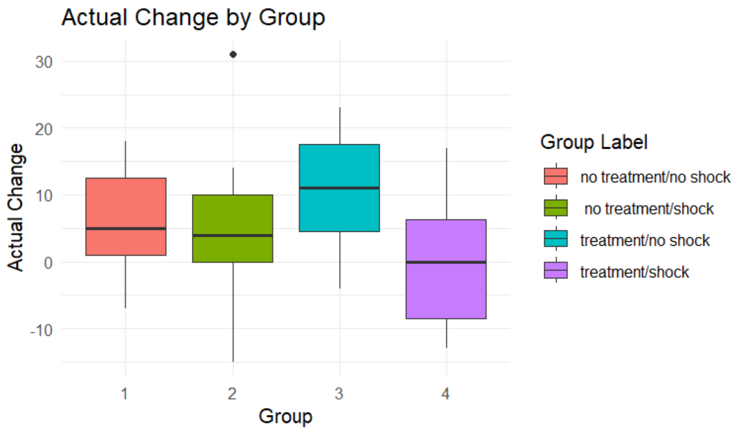

Figure 10: boxplot of actual change by group

- **Correlation analysis**: the correlation matrix reveals key relationships between cognitive and environmental factors influencing participants' theory updates. Strong positive correlations among Knowledge_Breadth, Knowledge_Depth, and Confidence suggest that participants with broader and deeper initial theories tend to be more confident. Cognitive stimuli, such as GPT_Usage, GPT_Familiarity, and Algorithmic_liking, are closely linked, indicating that participants familiar with AI tools are more likely to trust them. Conversely, Algorithmic_aversion shows a strong negative relationship with Algorithmic_liking but weak correlations with other variables, suggesting it operates independently. While Shock has limited direct correlation with V_diff, its weak negative association with Knowledge_Breadth and variability measures (SD_1, SD_2) indicates that it might indirectly influences theory updates.

These patterns highlight the interactions between cognitive stimuli, environmental factors, and theory adjustments.

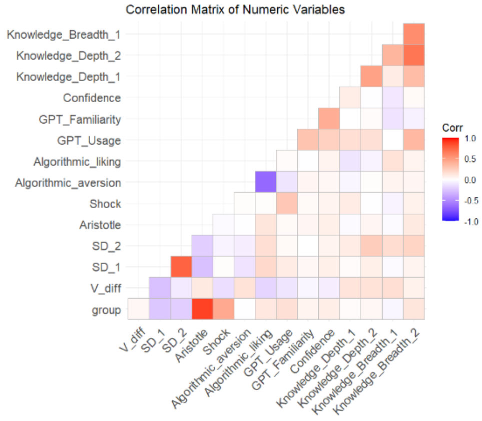

Figure 11: correlation matrix of Aristotle's dataset

- **Regression analysis**: we can find the results of the regression analysis in the following tables:

Coefficient	Estimate	Std. Error	t value	P value	Significance
Intercept	2.8525	9.427	0.303	0.7634	
Aristotle	4.5038	2.585	1.742	0.0875	*
Shock	0.8529	2.5914	0.329	0.7434	
Algorithmic_liking	-0.9177	0.6969	-1.317	0.1938	
GPT_Usage	-0.5529	1.0877	-0.508	0.6134	
GPT_Familiarity	0.3384	0.9665	0.35	0.7277	
Confidence	-0.2311	0.9564	-0.242	0.81	
Knowledge_Depth_1	0.762	0.8872	0.859	0.3945	
Knowledge_Depth_2	0.5373	1.5867	0.339	0.7363	
Knowledge_Breadth_1	1.332	0.9212	1.446	0.1543	
Knowledge_Breadth_2	-0.818	1.3893	-0.589	0.5586	
Aristotle:Shock	-4.5698	3.6686	-1.246	0.2186	

Table 5: V_diff regression's results

Minimum	1Q	Median	3Q	Maximum
-10.7146	-4.7486	-0.6687	5.2948	20.0021

Table 6: V_diff regression's residuals

Coefficient	Estimate	Std. Error	t value	P value	Significance
Intercept	3.102	12.1748	0.255	0.7999	
Aristotle	5.2112	3.3385	1.561	0.1247	
Shock	-0.9291	3.3468	-0.278	0.7824	
Algorithmic_liking	-0.5888	0.9	-0.654	0.5159	
GPT_Usage	-0.8943	1.4047	-0.637	0.5272	

GPT_Familiarity	0.6172	1.2483	0.494	0.6231	
Confidence	-1.0849	1.2352	-0.878	0.3839	
Knowledge_Depth_1	2.0549	1.1459	1.793	0.0789	.
Knowledge_Depth_2	-0.0615	2.0491	-0.03	0.9762	
Knowledge_Breadth_1	1.5022	1.1897	1.263	0.2125	
Knowledge_Breadth_2	-1.1815	1.7943	-0.658	0.5132	
Aristotle:Shock	-9.2823	4.7379	-1.959	0.0556	.

Table 7: V_diff_nonabs regression's results

Min	1Q	Median	3Q	Max
-17.1722	-5.7528	0.1708	4.9062	23.4901

Table 8: V_diff_nonabs regression's residuals

The regression results, integrated with earlier analyses, provide insight into how cognitive tools, shocks, and individual factors shape participants' theory updates. In the first model (V_diff), Aristotle significantly increases the magnitude of adjustments (coefficient = 4.50, p = 0.0875), suggesting it promotes greater reassessments. However, its interaction with shocks (-4.57, p = 0.2186) indicates a stabilising effect during disruptions, aligning with previous findings that treatment alone (Group 3) led to the largest adjustments, while treatment combined with shock (Group 4) produced smaller, more calibrated changes. Shocks alone have limited impact on the absolute magnitude of change in this model, as indicated by their small coefficient (0.85).

In the second model (V_diff_nonabs), directional theory adjustments are better predicted ($R^2 = 0.2723$) than in the first model, with Aristotle again demonstrating a positive but more pronounced effect (5.21, p = 0.1247). The interaction between Aristotle and shocks (-9.28, p = 0.0556) approaches significance, underscoring Aristotle's role in helping participants rationalize and stabilize their updates in the presence of shocks, ensuring that directional adjustments remain measured. This dynamic is evident in Group 4's smaller directional shifts, suggesting that participants in this group make use of Aristotle to integrate shocks into their evaluations without overreacting. Conversely, shocks primarily influence directional changes, pushing theories downward (coefficient = -0.93).

Initial knowledge depth (Knowledge_Depth_1) emerges as a near-significant factor (2.05, p = 0.0789), suggesting that participants with deeper theoretical understanding make more informed directional adjustments, as they are better prepared to process new information. Interestingly, neither model reveals significant effects for variables related to AI familiarity, such as GPT_Usage and Algorithmic_liking, indicating that prior exposure to AI tools does not strongly influence theory adjustments. This highlights the important role of cognitive and environmental factors, such as Aristotle's availability and the presence of shocks, over participants' familiarity with AI tools in shaping theory updates.

Aristotle Analysis

Theory

AI-generated attributes may fail to adequately account for the specific context defined by a company, its decision maker, and the nature of the decision problem. This issue arises when the AI prioritizes attributes that are prevalent in its training data, favouring widely discussed or generic trends without sufficiently considering their relevance or practicality for the company's size, sector, or strategic goals. Consequently, attributes that are valuable in a broad sense might be poorly suited to the unique circumstances of a particular industry or organization. This lack of contextual sensitivity can result in inappropriate recommendations, such as applying strategies suited for large corporations to start-ups or indiscriminately proposing unrelated trends across diverse sectors.

Implementation

Classification

To evaluate the contextual relevance of AI-generated attributes, a structured taxonomy is used. This taxonomy organizes attributes into four macro categories: Economic, Technology, Strategic, and Regulatory. Macro categories are further divided into three micro categories:

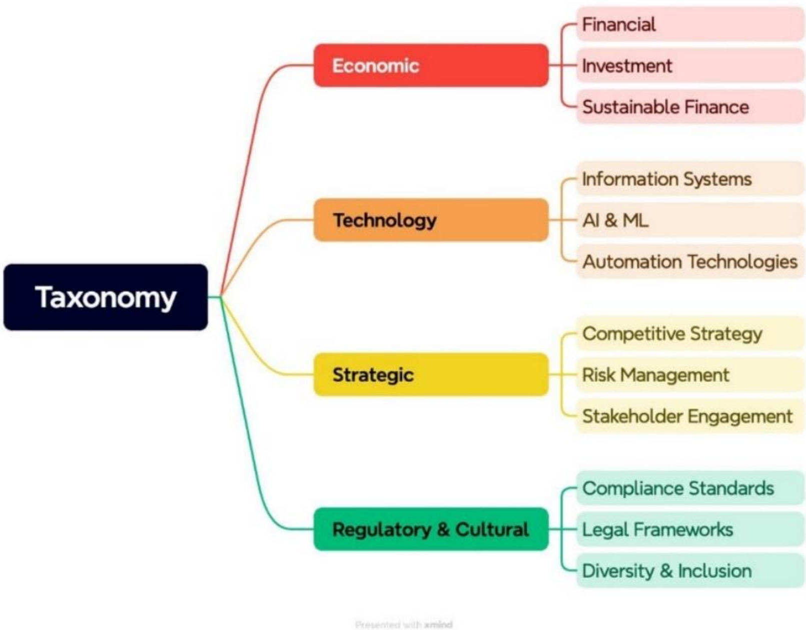

Figure 12: Attributes' taxonomy

This classification is applied to three variables: the attributes, the decision problem, and the company itself. To evaluate the alignment effectively, it is crucial to classify and categorize companies based on their characteristics, including sector, product type, revenue, headcount, and years of operation. Additionally, companies should be classified by their status, such as start-up, small-cap, mid-cap, large-cap, corporate, or multinational.

Fit Score & Generalization Bias

The next step is to compute a Fit Score, which quantifies how well AI-generated attributes align with a company's specific context and decision problem. This score combines two key metrics: relevance (R) and applicability (A). Relevance measures how well an attribute addresses the decision problem at hand, while applicability evaluates whether the attribute is feasible within the company's situation. The fit score is calculated as:

$$Fit\ Score = w_R R + w_A A$$

where both R and A scores range from 0 to 1, with 0 indicating no relevance or applicability, and 1 representing perfect relevance or applicability. The sum of w_R and w_A equals 1, representing the weights assigned to relevance and applicability, which can be determined through expert judgment or pre-validated business case data. To calculate the generalization bias (GB), we use the following formula:

$$GB = 1 - Fit\ Score$$

This value reflects how much the attribute is overly generic or misaligned with the company's context or the decision problem it seeks to address.
Initially, R and A should be evaluated by human experts in theory value creation, while considering the specific industry standards, business constraints, and insights from past successful and unsuccessful strategies. Additionally, methods such as text mining of attribute definitions can complement the evaluation of relevance and applicability.
In conclusion, generalization bias can be analysed and summarized considering also internal or external factors, once these classifications are defined. Together, this scoring system and its supporting analyses provide a comprehensive framework to evaluate the contextual validity of AI-generated recommendations, ensuring they are both relevant and actionable for the company and its decision problem.

Conclusion

This framework provides a structured approach to assess the contextual fit of AI-generated attributes by integrating Relevance, Applicability, and Generalization Bias. By classifying both attributes and companies according to industry-specific characteristics and decision contexts, the system ensures that AI recommendations are a good fit to the unique problem. This extensive evaluation, supported by expert insights and analytical methods like text mining, helps ensure that AI-generated suggestions are not only contextually relevant but also practical and actionable, minimising the risk of generic (ChatGPT, n.d.) or inappropriate recommendations.

Bibliography

Gurdus, L. (2016, November 22). *CNBC*. Retrieved from https://www.cnbc.com/2016/11/22/your-world-changed-forever-in-2007-with-iphone-and-tech-boom-thomas-friedman.html#:~:text=Also%20in%202007%2C%20Facebook%20and,interview%20on%20CNBC's%20Squawk%20Box."

Teppo Felin, T. R. (2017). The Theory-Based View: Economic Actors as Theorists. *STRATEGY SCIENCE*, 258-271.

Gambardella, A. (2024, August). A Theory-Driven Perspective on Strategic Decisions Under Uncertainty.

Arnaldo Camuffo, A. G. (2024). THEORY-DRIVEN STRATEGIC MANAGEMENT DECISIONS.

Tennant, F. (2022, July). *Financier Worldwide*. Retrieved from https://www.financierworldwide.com/culture-clashes-in-ma-new-perspectives

Federazione Nazionale Unitaria Titolari di Farmacia. (n.d.). *Fedefarma.it*. Retrieved from https://federfarma.it/Farmaci-e-farmacie/Farmacie-e-farmacisti-in-Italia/La-presenza-delle-farmacie-sul-territorio/Piemonte.aspx

Provincia di Novara. (n.d.). Retrieved from https://www.regioinsubrica.org/il-territorio/provincia-di-novara#:~:text=E'%20articolata%20in%2088%20comuni,Orta%20(quasi%20per%20intero).

Federfarma. (2024). *La Farmacia Italiana 2024.* Federfarma. Retrieved from https://farmaciavirtuale.it/wp-content/uploads/2024/05/4515-La-farmacia-italiana-2024-Report-Federfarma.pdf

IRES Piemonte. (2019). *10 NUMERI SUL PERSONALE NELLA SANITÀ PUBBLICA PIEMONTESE*. IRES Piemonte.

IRES Piemonte. (2024). *COME SI INVECCHIA IN UN PIEMONTE SEMPRE PIÙ ANZIANO? UNO SGUARDO SULLE OPPORTUNITÀ DELL'INVECCHIAMENTO ATTIVO*. IRES Piemonte.

Provincia di Novara. (n.d.). *La Provincia di Novara*. Retrieved from https://www.provincia.novara.it/Provincia/territorio.php#:~:text=La%20Provincia%20di%20Novara%20con,provincia%20del%20Piemonte%20per%20estensione.

Regione Piemonte. (n.d.). *Il sistema sanitario piemontese*. Retrieved from https://www.regione.piemonte.it/web/temi/sanita/sistema-sanitario-piemontese

IRES Piemonte. (2021). *Rete Ospedaliera e Rete Territoriale in Piemonte.* IRES Piemonte.

(n.d.). Retrieved from ChatGPT: chatgpt.com

Appendix 1

In this appendix, we present graphs that provide insights into the balance of the experimental sample. The following graphs are included:

- Histogram of age groups
- Histogram of sex groups
- Histogram of Italian responders
- Histogram of educational levels
- Histogram of educational field

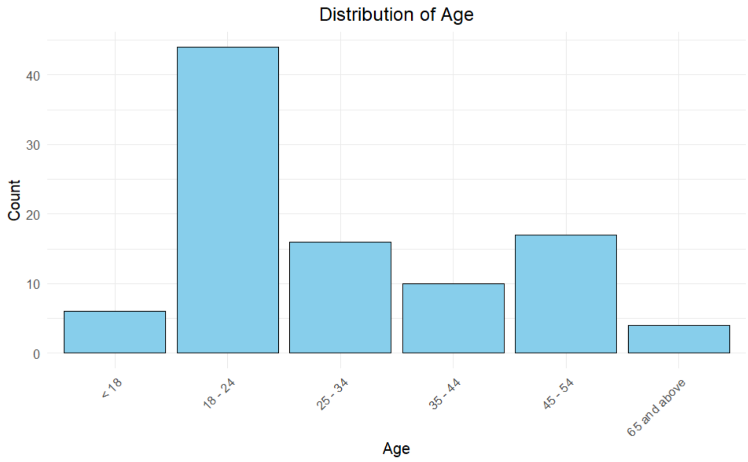

Figure 13: Histogram of age group

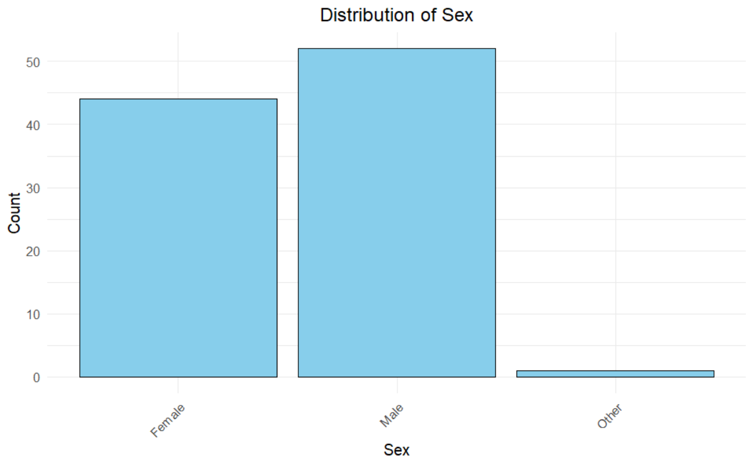

Figure 14:Histogram of sex groups

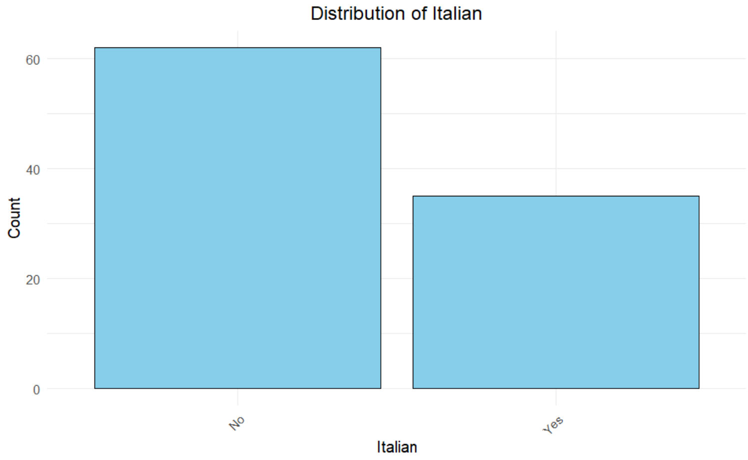

Figure 15: Histogram of Italian responders

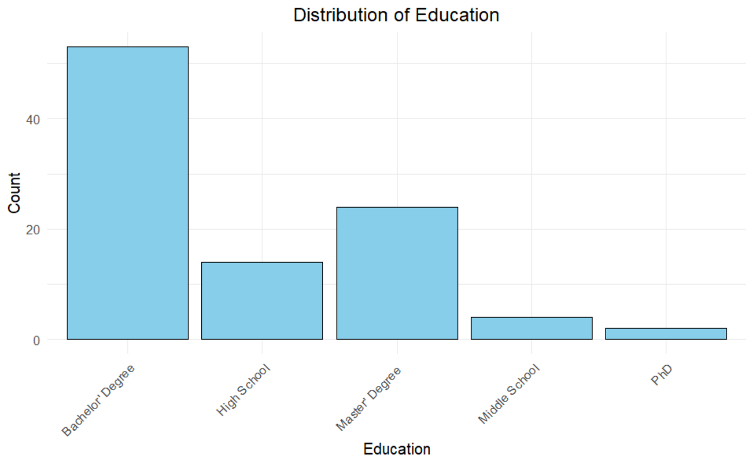

Figure 16: Histogram of educational levels

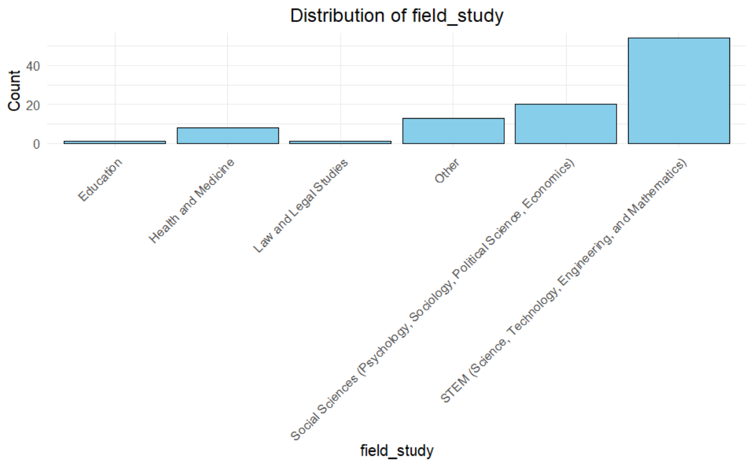

Figure 17: Histogram of educational field