Eng Int Syst (2024) 1: 63–75

Engineering
Intelligent Systems

Automated House of Resilience with AI-based Measures

Aloïs Goeury[1], and Elizabeth Chang[2]

[1] *IMT Mines Albi and Griffith University*
[2] *Griffith University*

This paper presents the automated analysis of resilience capability and situation awareness of the capability readiness at one glance through the House of Resilience. We extend the concept of House of Quality into House of Resilience. The resilience attributes and measures were inspired by the work of Jnitova et al. (2022). However, such analysis is complex, usually manual and would take 3 weeks to 6 months to generate, and the interpretation of the result is another level of challenges for both junior or senior officers or employees or executives. because Resilience is an open-end concept, can be interpreted in many ways and resulting in different measures and leading to different understanding. To have a system that can unify the concept, allow common and shared understanding of resilient capability and how it can be measured is the motivation of this study. There is no such tool available in the world. We research into open-source technology aimed at help represent the complex concept of Resilience in a simple and straight way, and we believe such system and approach will be very useful for any organization for resilience capability measure, whether it is related to workforce performance or professionalisation training systems and operation processes. This automated analysis and results visualisation represented by the House of Resilience allow quick understanding of strength and weakness of the capability readiness in near real tine or within few minutes of data collection.

Keywords: House of Resilience, Automation, Resilience Attributes, AI based Measures

1. INTRODUCTION

The House of Resilience offers the automated measures of individuals of all ages the opportunity to acquire resilience capability in the VOCA world (Volatile, Uncertain, Complex and Ambiguous). This paper use the case study from the Australian Defence, and their "Defence Training System," developed based on Jnitova et al. (2021)[1] research and his PhD thesis in 2022. It is centred on the resilience, focusing on six key attributes: "Adaptive Capacity," "Adaptability," "Agility," "Efficiency," "Robustness," "Recovery and Redundancy." After each training session within the Australian Defence system, a comprehensive survey is dispatched to all relevant stakeholders, encompassing individuals ranging from the trainees themselves to the workplace supervisors, instructors, training specialists, and capability managers. This inclusive approach to soliciting feedback is crucial for several compelling reasons. Firstly, each stakeholder's perspective is indispensable, as they bring unique insights and experiences, shaping a holistic understanding of the training's impact. Secondly, considering the diverse locations where training occurs, it becomes evident that what may apply effectively in one training venue may not necessarily yield the same outcomes in another. The distinct geographical, logistical, and operational aspects of various training sites demand a thorough assessment from all perspectives, allowing for tailored evaluations and targeted improvements. By engaging all parties involved, the Australian Defence ensures that the data collected is comprehensive and representative, laying the foundation for informed decision-making and continuous enhancement of their training programs to meet the dynamic needs of different training environments.

E-mails: alois.goeury@mines-albi.fr; e.chang@griffith.edu.au

[1] Jnitova et al., « Improving Enterprise Resilience by Evaluating Training System Architecture: Method Selection for Australian Defence».

2. MOTIVATION OF THE STUDY

The time-consuming and labour-intensive nature of the data analysis process within the Australian Defence's Training System poses significant challenges. The requirement for human intervention in data cleaning, plotting, and report writing, coupled with the extended period for collecting responses, results in a four-month timeframe to obtain post-training results. This delay creates the risk of forgetting critical aspects of the training and facing similar issues during the waiting period.

3. THE CASE STUDY

Within the context of our research, we inherit the work from Jnitova et al. (2021). The several key words stand out as crucial elements as part of case study. The foremost among them is the "survey," a powerful tool for gauging the effectiveness of education and training programs. Specifically, we are interested in surveys related to education, which will provide valuable insights into the quality and impact of training efforts. "Defence training" emerges as a central theme. Understanding the intricacies and nuances of training within the Australian Defence system will be pivotal in comprehending the context in which our data analysis and evaluations are situated. Moreover, we seek to explore the concept of "resilience" in-depth, as it serves as a cornerstone of the Defence Training System, embodied in six key attributes: "Adaptive Capacity," "Adaptability," "Agility," "Efficiency," "Robustness," and "Recovery and Redundancy." Another important aspect is "data analysis," a fundamental process in our way to understand the outcomes and potential improvements of the Defence Training System. We are particularly intrigued by data analysis programs, which encompass both programming languages supporting data analysis and Application Programming Interfaces (APIs) that enable efficient data manipulation and visualization.

As we endeavour to analyse and interpret the free-form responses in the satisfaction surveys, "Natural Language Processing" (NLP) emerges as a critical tool. NLP will enable us to gain valuable insights from unstructured text, allowing us to identify trends, sentiments, and areas of improvement.

4. THE LITERATURE ON THE STUDY OF RESILIENCE AND ITS MEASURES

Jnitova et al. (2022)[2] developed a comprehensive survey aimed at evaluating every aspect of defence training, involving all stakeholders through an adaptive survey approach. Each analysis is meticulously conducted manually after thorough verification. The authors utilized SurveyMonkey to collect a variety of metadata, including survey initiation timestamps. Moreover, they also distributed paper surveys to individuals without internet access. Building upon the groundwork laid by Jnitova et al. (2021), the concept of resilience is distilled into six attributes: Adaptive Capacity, Adaptability, Agility, Efficiency, Robustness, and Recovery and Redundancy. The survey commences with demographic inquiries, followed by questions targeting these resilience attributes, and concludes with open-ended queries. To enhance the survey over time, Anova and T-test procedures were employed to eliminate biases and refine question clarity (fuzzy questions). The Training System Resilience Survey (TSRS) has experienced significant refinement, transitioning from an initial pilot consisting of 107 questions to a more concise version containing 54 questions, all aimed at minimizing potential biases (Jnitova et al., 2022). Given the diverse settings in which training sessions occur, the survey's adaptability to different populations ensures the generation of meaningful and efficient metrics. Within the survey framework, a spectrum of logic scales is employed for questions to solicit quantitative responses, encompassing both positive and negative perspectives. This means that, while the scale always spans from 1 to 5, its interpretation can vary: in some instances, 1 represents the least favourable outcome and 5 the most favourable; conversely, in other scenarios, 5 might be the least favourable score and 1 the most favourable. This quantitative assessment is grounded in a comparative analysis between obtained results and predefined desired values.

Qualitative responses are meticulously organized through a process of clustering, grouping them "according to their depth, breadth (number of themes covered), and strength of themes' affiliation" (Jnitova et al., 2022), a methodology influenced by the approach outlined by Braun and Clarke (2013)[3]. The quantitative responses are analysed at two levels: an initial comprehensive overview, followed by a more intricate scrutiny of each demographic aspect.

It's essential to acknowledge certain limitations associated with the TSRS. One limitation pertains to the "complexity of attribute relationships" (Jnitova et al., 2022), which underscores the intricate interplay between various resilience attributes. Additionally, three other limitations are identified, with the final one concerning information management and sharing. Due to the sensitive nature of Defence information, sharing such data poses challenges. However, the authors offer a potential solution, indicating that a dedicated platform is currently being developed to facilitate information sharing across multiple units. This meticulous development and deployment of the TSRS underscores its commitment to comprehensively assess training efficacy, considering both quantitative and qualitative perspectives. The refinement process, accompanied by strategic analysis, showcases the authors' dedication to ensuring the survey's validity and usefulness in enhancing the Defence Training System's resilience attributes.

Finally, Jnitova et al. (2022) also emphasize the significance of qualitative analysis in their study. In their pursuit of qualitative insights, the authors employed mind mapping techniques facilitated by the MAXGDA package. However, there is a lack of automation in this operation.

Go back to the survey, in terms of response rates, research in the medicinal domain indicated an approximately 40%

[2] Jnitova, V., M., Joiner, K.F., Xavier, A., Chang, E., Ferris, T., and Camelia, F. (2024). Is Your Training Service Resilient and Postured to Support Organisational Sustainment? Australian Journal of Multi-Disciplinary Engineering, currently in production to be published.

[3] Clarke et Braun, « Successful qualitative research: A practical guide for beginners. »

response rate for online surveys after three waves (Aerny-Perreten et al., 2015)[4]. Similar rates were observed in a psychopathological profiling survey examining individuals' weight (Varela et al., 2016)[5]. Varela et al. (2016) employed SurveyMonkey and achieved a response rate of 42%. Non-response items represent a significant challenge, although this issue is not notably distinct from other survey modes (Èehovin et al., 2023)[6], and it can be addressed through reminders or prerequisite questions, particularly in written surveys.

Comparative analysis has revealed that online surveys possess advantages in terms of being cost-free, expeditious, and requiring relatively minimal labour for creation (Vasantha Raju N. and N.S.Harinarayana, 2016)[7]. Google Forms, for instance, is entirely free but has certain limitations related to data privacy. The free version of SurveyMonkey, while providing a basic level of functionality, offers very limited capabilities (Vasantha Raju N. and N.S.Harinarayana, 2016). In contrast, the proprietary version of SurveyMonkey comes at a monthly cost of $26, offering more comprehensive features and customization options (Vasantha Raju N. and N.S.Harinarayana, 2016). However, it's important to note that the implementation of online surveys requires a certain level of expertise (Vasantha Raju N. and N.S.Harinarayana, 2016).

5. CHALLENGES IN THE ANALYTICS AND MEASURE OF RESILIENCE

In the realm of big data analysis, Agrawal et al. (2015)[8] highlight three significant challenges that emphasize the complexities of working with vast datasets: perceptual scalability, real-time scalability, and interactive scalability. In their exploration of effective data visualization tools, the authors introduce several software options, including D3.js, a free and open-source web framework. D3.js empowers users to create dynamic and engaging visualizations for comprehensive data representation. We trailed Python and R (Siddiqui and Alkadri, 2017)[9] and tools like Jupyter Notebook, along with libraries like pandas and matplotlib, which streamline the creation of essential charts (Sahoo et al., 2019)[10] we developed user-friendly software using the tkinter framework (Beniz and Espindola, 2017)[11].

In terms of backend development for web applications, we trailed Flask and Django are prominent Python libraries for APIs and smaller projects building and larger web applications, including HTML pages (Idris, Mohd Foozy, and Shamala, 2021)[12]. Flask support streamlined design approach and Django help handle more intricate web applications (Ghimire, 2010[13]; Gore et al., 2021[14]; Smyth, 2010[15]). As the digital landscape continues to evolve, trial and decision making on applying the right technology stack becomes a critical and that impacts the efficiency and effectiveness of data analysis and application development.

To analyse the qualitative responses in surveys, Python, as a versatile tool, provides solutions for text summarization using tools like SpaCy and Natural Language Processing (NLP) (JUGRAN et al., 2021)[16].

To enable AI and ChatGPT capabilities, we trailed Cambria and White in 2014[17] which delved into the field of NLP and concluded that while NLP had room for improvement in interpreting emotions and cultural nuances, GPT has effectively shattered these barriers (Bhaskar, Fabbri, and Durrett, 2023)[18]. GPT, with its impressive metrics-driven understanding of human opinions, has revolutionized the landscape. However, it's crucial to acknowledge that GPT still faces limitations, such as challenges in translation (Jiao et al., 2023)[19] and aspects of creativity (Sawicki et al., 2023)[20].

Interestingly, when it comes to text summarization, discerning between human-generated summaries and those produced by ChatGPT becomes exceedingly challenging (Soni and Wade, 2023)[21]. This observation underscores the remarkable capabilities of AI language models like GPT in effectively distilling information while maintaining the essence of the original content.

Further, the important need for an automated and user-friendly data analysis system for satisfaction surveys becomes apparent. Such a system could empower the Defence to take full control of their data analysis, eliminating the reliance on external analysts and expediting the evaluation process. Thus, the central research question becomes: Can the implementation of an automated data analysis system in the Australian Defence's Training System lead to more efficient and timely evaluation of education quality, ultimately enhancing the resilience attributes in trainees?

An additional aspect to consider is whether this automated system can provide comprehensive data visualization that synthesizes all relevant information. The ability to interpret results briefly on a single page could be beneficial in facilitating quick and informed decision-making. Therefore, the supplementary research question is: Can the implementation of an automated data analysis system offer data visualization

[4] Aerny-Perreten et al., « Participation and Factors Associated with Late or Non-Response to an Online Survey in Primary Care ».

[5] Varela et al., « Advantages and Disadvantages of Using the Website SurveyMonkey in a Real Study: Psychopathological Profile in People with Normal-Weight, Overweight and Obesity in a Community Sample ».

[6] Èehovin, Bosnjak, et Lozar Manfreda, « Item Nonresponse in Web Versus Other Survey Modes ».

[7] Vasantha Raju N. et N.S.Harinarayana, « Online survey tools: A case of study of Google Forms ».

[8] Agrawal et al., «Challenges and Opportunities with Big Data Visualization».

[9] Siddiqui et Alkadri, « Review of Programming Languages and Tools for Big Data Analytics ».

[10] Sahoo et al., « Exploratory Data Analysis Using Python ».

[11] Beniz et Espindola, « Using Tkinter of Python to Create Graphical User Interfae (GUI) for Scripts in LNLS ».

[12] Idris, Mohd Foozy, et Shamala, « A Generic Review of Web Technology ».

[13] Ghimire, « Comparative Study on Python Web Frameworks: Flask and Django ».

[14] Gore et al., « Django: Web Development Simple & Fast ».

[15] Smyth, « Creating Web APIs with Python and Flask ».

[16] JUGRAN et al., « Extractive Automatic Text Summarization using SpaCy in Python & NLP ».

[17] Cambria et White, « Jumping NLP Curves ».

[18] Bhaskar, Fabbri, et Durrett, « Prompted Opinion Summarization with GPT-3.5 ».

[19] Jiao et al., « Is ChatGPT A Good Translator? »

[20] Sawicki et al., « Bits of Grass ».

[21] Soni et Wade, « Evaluating and Detecting ChatGPT's Responses on Abstractive Summarization ».

Figure 1 Tkinter Logo from "iot4beginners.com".

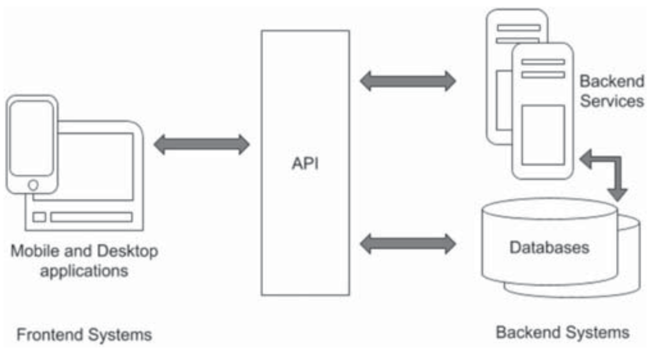


Figure 2 The role of API from "A Seven-Step Guide to API-First Integration".

that presents a comprehensive overview of results, enabling immediate interpretation on the same page?

It's important to emphasize that while this research is conducted within the specific context of the Australian Defence, its implications extend far beyond this domain. The challenges associated with handling satisfaction surveys are not unique to the defence sector; they resonate across a wide array of organizations and industries, including healthcare, education, customer service, and more. Therefore, this research serves as a valuable model for improving satisfaction survey analysis, with the potential to be adapted and applied effectively in diverse contexts.

6. AUTOMATED MEASURE OF RESILIENCE AND THE PROTOTYPE

6.1 Centralizing and Automating Data Analysis

The objective of the research was to streamline the data analysis process to improve the efficiency. In pursuit of this goal, the versatility of Python was opted to be leveraged due to its rich library ecosystem that supports data analytics, API interaction, and application development. The application comprises two distinct components: the frontend and the backend. The frontend, developed using Tkinter, serves a specific purpose in our prototype, to facilitate data input for report creation.

In the prototype, aesthetics was not a primary concern; rather, the frontend was designed to efficiently gather requisite information for report generation. Shifting to the backend, the Flask framework has been employed due to its user-friendly learning curve and strong suitability for creating APIs. Notably, the application exclusively interfaces with APIs, thereby allowing for future iterations of the frontend to be revamped and enhanced in subsequent research.

The resultant output of the system is an Excel file. This format was opted, as it has proven effective in similar contexts, such as in the work of Jnitova et al. (2022), for generating comprehensive reports. Excel seamlessly integrates with Python through libraries like OpenPyxl[22] and XlsxWriter[23], making it an ideal choice. Moreover, Excel's multi-sheet capabilities enable us to provide in-depth explanations in a structured manner, extending to a second level of detail. The user-friendly nature of Excel makes it accessible for users

[22] Hunt, *Advanced Guide to Python 3 Programming*.

[23] Tetali, « A PYTHON TOOL FOR EVALUATION OF SUBJECTIVE ANSWERS (APTESA) ».

across various devices, as it does not necessitate licensing constraints, allowing for seamless sharing and collaboration.

6.2 Create Surveys

Achieving the goal of centralization encompasses the creation of surveys within our application. To facilitate this process, a dedicated tab was implemented within the application interface.

Several key components are integral to this survey creation tab. Firstly, the survey's name, serving as its unique identifier for response collection, is specified. Additionally, the unit for which the survey is intended, the questionnaire version, the repetition count of the survey with the same questionnaire and unit, are all essential parameters. The questionnaires themselves are sourced from an Excel file, a compilation resulting from the efforts of Jnitova et al. (2022), encapsulating input from all relevant stakeholders.

The Google Forms API was selected to be used as the chosen online survey platform. This decision was informed by our literature review, which indicated that Google Forms offered distinct advantages for our prototype. Notably, Google Forms imposes no restrictions on the number of questions, surveys, or responses, a crucial consideration for the comprehensive research approach. Conversely, alternatives such as SurveyMonkey were evaluated but found to have limitations, including a cap of 10 different surveys using their API and a limit of 100 responses for the free version (Vasantha Raju N. and N.S.Harinarayana, 2016). Our selection of Google Forms aligns with our aim for scalability and unrestricted data collection, enabling us to effectively gather insights from all stakeholders. Additionally, it's imperative to note that the integration of Google Forms necessitates the selection of a credential file for API access, further enhancing security and streamlining the survey management process.

6.3 Collect Responses

To facilitate the retrieval of survey responses, a user interface with a dropdown list was developed which dynamically populates with the names of the created surveys, along with corresponding details such as version and repeatability. Given the reliance on Google Forms API, the inclusion of credentials remains pivotal to enable seamless access. Figure 8 present the initial prototype interface.

Within this interface, a button facilitates the transition to a dedicated page for both response collection and report generation. While the upper section mirrors the component for downloading results, the lower segment diverges, introducing distinct functionalities. In the lower section, users are presented with several options.

6.4 Create the Report From the Results File

An automated interface has been designed to address the necessity of analysing results obtained prior to the creation of the application. The survey data can be in a singular Excel file encompassing online surveys. As shown in figure 6, the pertinent survey information is also essential to compare surveys from different period.

6.5 Import data for Pre-Processing

The final page serves as a platform for opening the Excel reports that have been generated. Through the selection of the reference name, a dropdown list containing all the reports associated with that particular reference becomes accessible. By choosing a specific report from the list, users can initiate the process of opening and viewing the selected report.

6.6 Resilience Measure and Implementation

The backend mirrors the structure of the frontend, consisting of five APIs developed to facilitate automated functionalities, including report creation, survey generation from an Excel file, and response collection. The API is responsible for generating surveys from an Excel file containing questionnaires and stakeholder's feedbacks and each sheet is organized into three sections. The first section covers demographic questions, featuring columns for the question code, the question itself (separated by a semicolon), response options, and corresponding numerical values for analysis. The second section pertains to resilience questions, with similar code and question columns, as well as columns for response options such as 'always,' 'often,' 'sometimes,' 'rarely,' 'never,' and 'IET' (for trainees). The final section deals with free-response questions, containing code and question columns. Upon completion, the API returns the necessary URLs, which can then be distributed to the intended recipients.

We also developed an API to gather survey responses. Leveraging the JSON file generated by the preceding API, which encompasses the survey IDs required for response collection, this API is capable of interfacing with the Google Forms API (version 1) to retrieve and transform responses that are in JSON file. Subsequently, our API converts this JSON file into an Excel file. The Excel file is structured with each question code as a column and every respondent's answers as rows, organized into sheets corresponding to each stakeholder (Jnitova et al., 2022).

There exist two distinct APIs within the backend framework to create reports. Each catering to the instantaneous generation of reports and the creation of reports via file inputs. Despite their differences, these APIs share common elements that contribute to their functionality. One prominent variation between them revolves around the way responses are integrated during the report generation process, synchronously downloading responses as part of the creation process itself, a concept that logically aligns with the purpose of data visualization and analysis, which remains at the core of this research.

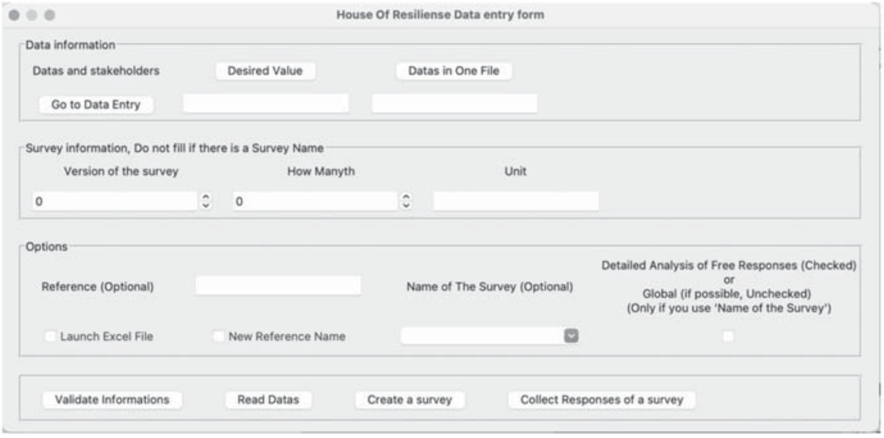


Figure 3 Sample Interface use to Generate House of Resilience.

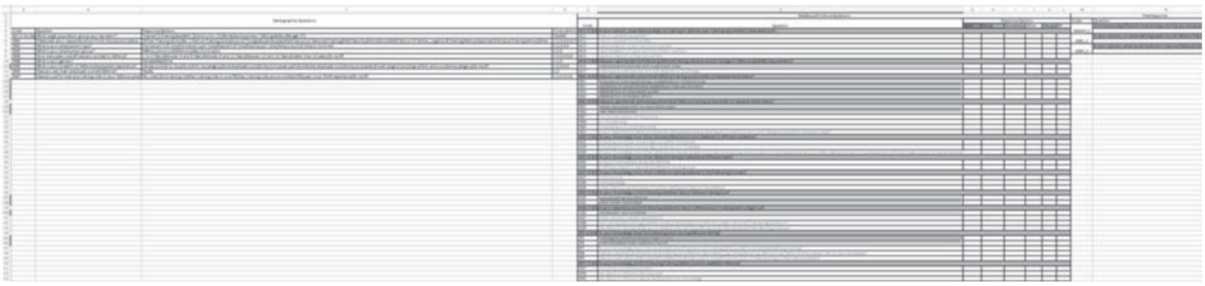
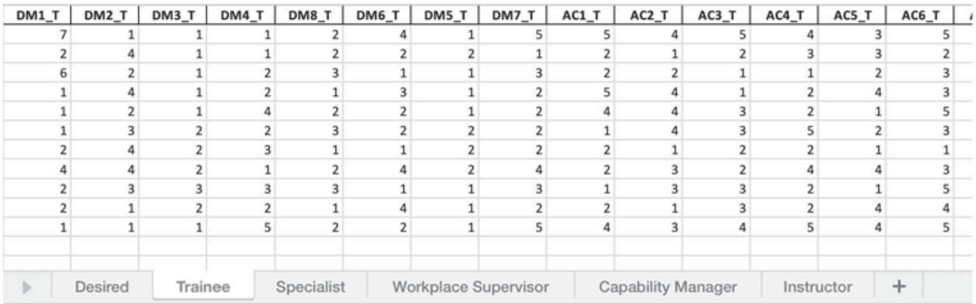


Figure 4 The Questionaries.



DM1_T	DM2_T	DM3_T	DM4_T	DM8_T	DM6_T	DM5_T	DM7_T	AC1_T	AC2_T	AC3_T	AC4_T	AC5_T	AC6_T
7	1	1	1	2	4	1	5	5	4	5	4	3	5
2	4	1	1	2	2	2	1	2	1	2	3	3	2
6	2	1	2	3	1	1	3	2	2	1	1	2	3
1	4	1	2	1	3	1	2	5	4	1	2	4	3
1	2	1	4	2	2	1	2	4	4	3	2	1	5
1	3	2	2	3	2	2	2	1	4	3	5	2	3
2	4	2	3	1	1	2	2	2	1	2	2	1	1
4	4	2	1	2	4	2	4	2	3	2	4	4	3
2	3	3	3	3	1	1	3	1	3	3	2	1	5
2	1	2	2	1	4	1	2	2	1	3	2	4	4
1	1	1	5	2	2	1	5	4	3	4	5	4	5

Desired | Trainee | Specialist | Workplace Supervisor | Capability Manager | Instructor | +

Figure 5 Questionnaires.

6.7 Automated Generation of the House of Resilience and the Sequence Diagram

Figure 10 illustrates the sequence diagram detailing the creation process from Excel files. The concept of the "House of Quality" originates from the realm of product engineering and serves as a cornerstone of the Quality Function Deployment (QFD) methodology. Initially introduced in 1972 within the Japanese automobile industry, this concept employs a matrix-based structure resembling a house (Hauser and Clausing, 1988)[24]. It facilitates the alignment of product attributes with customer needs and expectations. Customers' attributes constitute the left side of the house, whereas engineers' attributes are positioned at the top. The interior of the house features a relationship matrix that correlates these attributes. The roof, on the other hand, functions as a correlation matrix among various attributes of engineers. The bottom section comprises measurements used by engineers to evaluate products, while the left side encompasses customer perceptions (Hauser, 1993)[25].

The adaptability of the House of Quality extends beyond the automotive industry and has found significant use in software engineering (Liu, 2001)[26], treating technical components as software features. An empty column signifies that a particular feature doesn't address any customer need, rendering it redundant. Furthermore, the concept has evolved to enable complex decision-making through multi-layered structures (Isaac et al., 2015)[27]. Moreover, the House of Quality can be enriched through the application of fuzzy logic, a method that incorporates uncertainty and imprecision into decision-making processes (Temponi, Yen, and Amos Tiao, 1999)[28]. This integration aids in handling

[24] Hauser et Clausing, « The House of Quality ».

[25] Hauser, « How PURITAN-BENNETT used the HOUSE OF QUALITY ».

[26] Liu, « Software quality function deployment ».

[27] Isaac, Olumide, et Rasaki, « Application of House of Quality Matrix to Material Selection for Engineering Designs ».

[28] Temponi, Yen, et Amos Tiao, « House of Quality: A Fuzzy Logic-Based Requirements Analysis ».

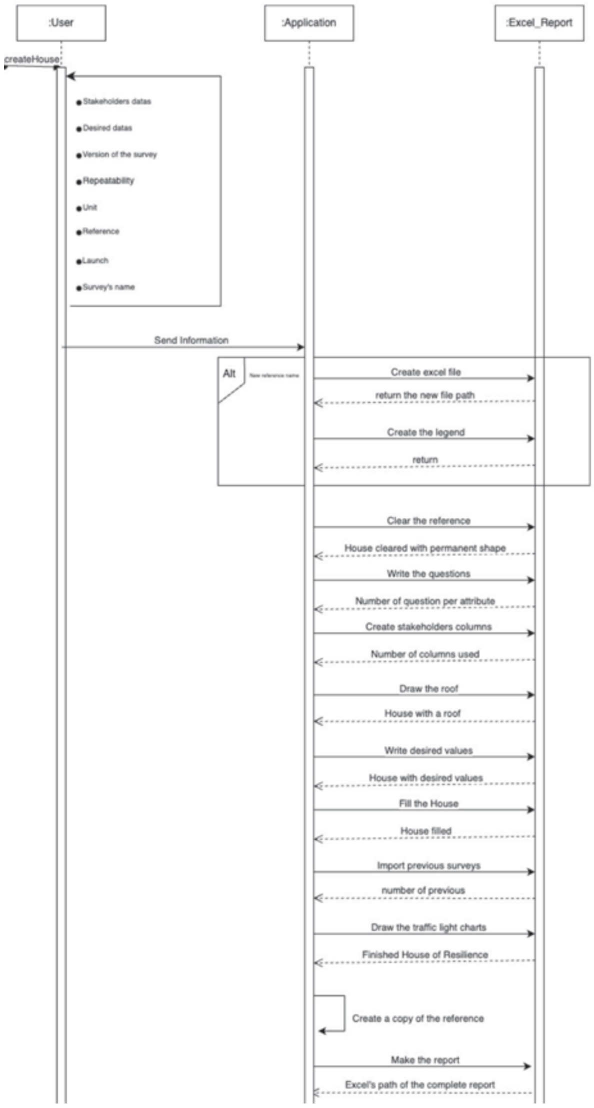


Figure 6 Sequence diagram.

situations where precise, binary distinctions might not be feasible.

The House of Quality's adaptability and flexibility make it well-suited to navigate complex and nuanced situations.

In our context, the automated House of Quality has been tailored to visualize survey results. The left side represents survey questions, the top represents stakeholders, and traffic light charts are positioned on the right. While the conventional relationship matrix is absent, it has been reimagined. The interior of the matrix now incorporates the most frequently provided responses by stakeholders for each question, along with their respective percentages. The cell colours denote the answer types. The process commences with the creation of the "House of Resilience," followed by the remaining report components generated through the chart templates devised by Jnitova et al. (2022), fully automated via our Python program.

Expressing a desire to utilize ChatGPT for analysis, a strategic approach was developed to effectively leverage the capabilities of the OpenAI GPT-3.5 turbo model. The objective was to summarize the responses to free-form questions and extract the key subjects for each question. The selection of this model was guided by insights from the literature review. A comprehensive and generalized prompt was meticulously formulated to ensure optimal outcomes. The process involves concatenating the responses from each question and assessing the token count to determine the suitable model version, either 8k or 16k tokens. If the token limit is exceeded, a similar procedure is executed on a per-question basis for each stakeholder. In the event of persistent token limit breaches, responses are randomly chosen until adherence to the token limit is achieved. The inputs for this procedure are essential to validate the operational costs and ensure efficient analysis. Subsequently, the responses generated by the GPT model are seamlessly integrated into the Excel file, occupying the designated sheet reserved for this purpose. Upon the report's completion, it is deposited within the app's designated folder structure, specifically under its corresponding reference name folder. In cases where the launch button is activated, the Excel file automatically opens, providing an efficient and user-friendly experience.

We developed the API for the backend serves the purpose of enabling users to access reports that have been generated. By inputting the reference name, the API facilitates the presentation of a list containing all the reports created under that

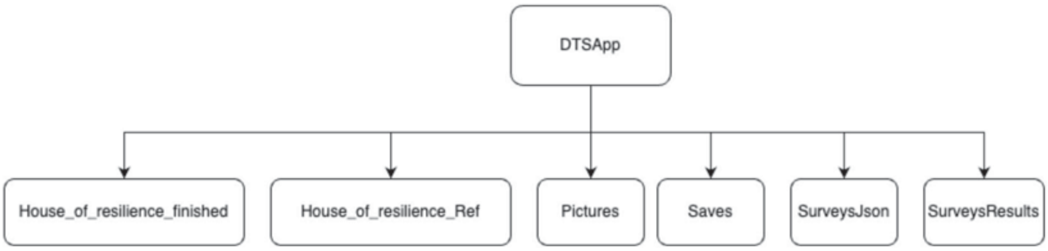


Figure 7 Folder structure.

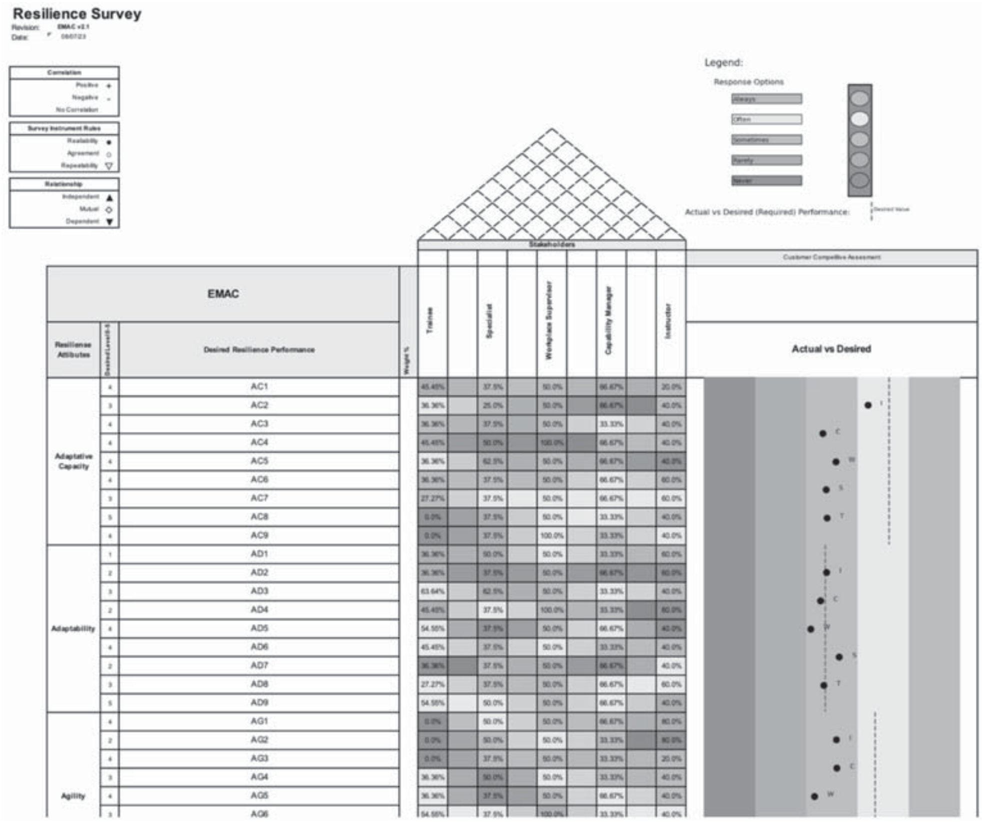


Figure 8 House of Resilience.

specific reference. This functionality streamlines the process of accessing and reviewing previously generated reports, contributing to a more efficient and organized workflow.

6.8 Automated House of Resilience and Visualization

For the results, the House of Resilience brings forth its merits based on reviews, albeit without the opportunity for extensive testing within the limited timeframe. Practical testing under real conditions has not yet been conducted. The House of Resilience offers a unique vantage point to observe both levels of analysis, employing visual representations to provide insights into training efficacy.

The Traffic Light Chart, a crucial component of the House of Resilience, serves to visualize the average responses of each stakeholder (represented by dots with the initial letter of their names) for each resilience attribute. This visualization is then compared to the desired value, indicated by a purple line. The colour coding employed in the chart signifies the opinion of the respondents, with colours ranging from green to red. However, this colour scheme isn't always straightforward, as "never" might represent a positive outcome in specific cases. For instance, if a question pertains to the punctuality of training starting at least 20 minutes late, a "never" response could actually be desirable, and thus, it would be depicted as green.

The Roof of the House of Resilience symbolizes the correlation between different stakeholders. A plus or minus sign is used to indicate positive or negative correlations, respectively, between pairs of stakeholders. This element offers insights into relationships and connections among various stakeholders' responses.

The most significant innovation lies in the Interior of the House, which takes the form of a two-dimensional table. Within this matrix, each cell represents the predominant response provided by a particular stakeholder (column) to a specific question (row). The colour of each cell corresponds to the response grade, while the percentage represents the

questions	Trainee_%	Trainee_TXT	Specialist_%	Specialist_TXT	Workplace Supervisor_%	Workplace Supervisor_TXT
AC1	45,45	Rarely	37,5	Sometimes	50	Sometimes
AC2	36,36	Often	25	Always	50	Rarely
AC3	36,36	Sometimes	37,5	Rarely	50	Always
AC4	45,45	Rarely	50	Never	100	Never
AC5	36,36	Often	62,5	Rarely	50	Always
AC6	36,36	Never	37,5	Never	50	Never
AC7	27,27	Always	37,5	Often	50	Often
AC8	0	No Data	37,5	Rarely	50	Often
AC9	0	No Data	37,5	Rarely	100	Often
AD1	36,36	Rarely	50	Rarely	50	Often
AD2	36,36	Rarely	37,5	Never	50	Rarely
AD3	63,64	Often	62,5	Rarely	50	Always
AD4	45,45	Rarely	37,5	Often	100	Sometimes
AD5	54,55	Often	37,5	Never	50	Sometimes
AD6	45,45	Often	37,5	Always	50	Often
AD7	36,36	Never	37,5	Always	50	Rarely
AD8	27,27	Often	37,5	Always	50	Always
AD9	54,55	Often	50	Often	50	Always
AG1	0	No Data	50	Rarely	50	Never
AG2	0	No Data	50	Never	50	Rarely
AG3	0	No Data	37,5	Never	50	Never
AG4	36,36	Rarely	50	Always	50	Rarely
AG5	36,36	Rarely	37,5	Always	50	Sometimes
AG6	54,55	Rarely	37,5	Rarely	100	Often
AG7	54,55	Sometimes	62,5	Often	50	Always
AG8	36,36	Rarely	50	Rarely	50	Often
AG9	45,45	Rarely	25	Always	50	Sometimes
EF1	0	No Data	50	Often	50	Rarely
EF2	0	No Data	62,5	Never	50	Rarely
EF3	0	No Data	37,5	Rarely	100	Rarely
EF4	0	No Data	25	Never	50	Rarely

Figure 9 Visualization of answers.

proportion of respondents who aligned with that majority grade. Here, red signifies the lowest grade, while green indicates the highest. This method of representation allows for swift identification of problematic areas. When less than 8 stakeholders participate, columns between stakeholders showcase a blend of colours, denoting consensus or divergence in opinions. Additionally, grey cells indicate questions for which no response was provided by the stakeholder.

A cursory glance at the House of Resilience reveals potential issues, indicated by red spots. These red and green markers act as indicators for decision-makers, enabling them to tailor future training sessions based on the observed discrepancies. The first level of analysis is achieved by examining the overall house to identify problematic areas and comparing attribute averages against desired values. Subsequently, the second level of analysis involves a more detailed investigation of each question and stakeholder response. This preparatory analysis offers a preliminary understanding of stakeholder perspectives and facilitates comparisons between them. Adjacent to the Traffic Light Charts, another essential feature of the House of Resilience framework comes into play: the presentation of previous survey results on a per-question basis. This inclusion serves a crucial purpose, to enable a comprehensive comparison and tracking of the training's evolution over time. These previous surveys consist of responses collected using the same questionnaire and involving the same unit. The sole variable that changes between these surveys is the repeatability factor. This feature holds significant potential as it offers a longitudinal view of how training attributes and stakeholder perceptions have evolved across multiple iterations of the same training sessions. By juxtaposing the latest survey responses with those from earlier instances, decision-makers can discern patterns, trends, and shifts in stakeholder feedback. This historical perspective provides insights into areas of improvement, helps to identify persistent challenges, and highlights positive changes that have been accomplished over time. By incorporating this comparative element, the House of Resilience fosters an in-depth understanding of the training's trajectory, which is essential for making informed decisions and refining future training initiatives. The ability to visualize the evolution of training effectiveness further enriches the framework's utility as a comprehensive assessment tool.

6.9 Automate the Advanced Analytics for Resilience Measures

To delve deeper into the analysis, a range of additional charts has been integrated into separate sheets within the House of Resilience framework. These charts provide a more nuanced exploration of the survey responses, offering insights into various dimensions of the training's effectiveness.

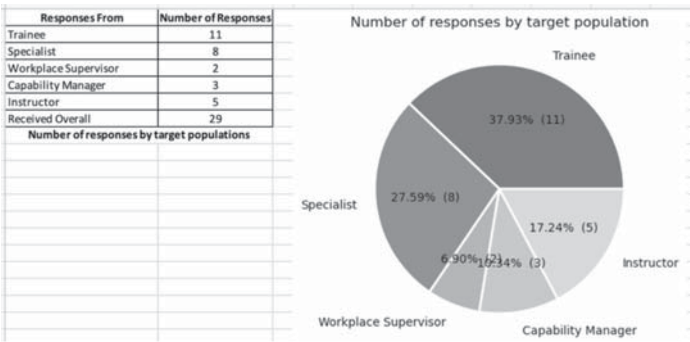
The "Visualization of Answers" sheet presents a detailed view of the individual responses provided by stakeholders. This sheet serves as an expanded version of the interior of the house, displaying the precise text of each answer alongside the stakeholder who provided it.

The "Nb_Responses" chart takes the form of a pie chart, illustrating the distribution and proportion of responses among the different stakeholders. This visualization provides a visual snapshot of how various stakeholder groups have contributed to the survey data.

The "Overall Resilience Performance" sheet contains the calculated averages of resilience performance, consolidating the opinions of all stakeholders. This metric provides a high-level overview of the training's overall effectiveness, capturing a consensus view that encompasses all stakeholder perspectives.

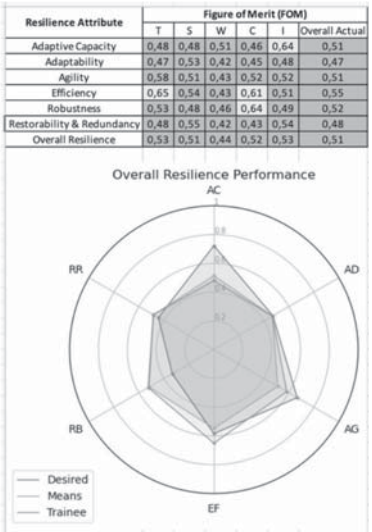
The "Radar Chart" is a visualization, featuring three spider webs. Webs represent the desired value for every resilience attribute, the trainee's assessment, and the average evaluation from other stakeholders. This visualization format offers a clear comparison between these three perspectives, highlighting areas of alignment or disparity. This radar chart is complemented by an accompanying table that facilitates the numerical comparison of these values, providing a quantitative representation of the disparities among the attributes.

The final two sheets, "Weakest and Strongest," reveal which stakeholders hold the most favourable and least favourable opinions for each resilience attribute. These insights can



Responses From	Number of Responses
Trainee	11
Specialist	8
Workplace Supervisor	2
Capability Manager	3
Instructor	5
Received Overall	29
Number of responses by target populations	

Figure 10 Number of Responses.



Resilience Attribute	Figure of Merit (FOM)					
	T	S	W	C	I	Overall Actual
Adaptive Capacity	0,48	0,48	0,51	0,46	0,64	0,51
Adaptability	0,47	0,53	0,42	0,45	0,48	0,47
Agility	0,58	0,51	0,43	0,52	0,52	0,51
Efficiency	0,65	0,54	0,43	0,61	0,51	0,55
Robustness	0,53	0,48	0,46	0,64	0,49	0,52
Restorability & Redundancy	0,48	0,55	0,42	0,43	0,54	0,48
Overall Resilience	0,53	0,51	0,44	0,52	0,53	0,51

Figure 11 Radar Chart.

Resilience Attribute	Weakest Performance		Strongest Performance		Desired Performance
	Target population	Score	Target population	Score	
Adaptive Capacity	C	0,46	I	0,64	0,72
Adaptibility	W	0,42	S	0,53	0,47
Agility	W	0,43	T	0,58	0,67
Efficiency	W	0,43	T	0,65	0,58
Robustness	W	0,46	C	0,64	0,33
Restorability & Redundancy	W	0,42	S	0,55	0,44
Weakest and strongest resilience attributes' performance as stated by target populations					

Figure 12 Weakest and Strongest.

be valuable for identifying key influencers or outliers in stakeholder perceptions.

The "Gaps and Surplus" sheet quantifies the gaps or surplus between the overall performance of the training and the desired performance levels. This analysis helps to pinpoint areas where the training is exceeding expectations or falling short, offering guidance for targeted improvements.

These additional sheets and charts, collectively woven into the House of Resilience framework, offer a comprehensive and multidimensional view of the training's effectiveness, thereby empowering decision-makers to make well-informed adjustments and enhancements.

A comprehensive assessment of survey creation time was conducted through a series of 15 tests. These tests were instrumental in determining the average duration required for survey generation. The standard survey size used for these tests emulated the structure employed by Jnitova et al. (2022), comprising 9 demographic questions, 3 free-form questions, and 54 questions related to resilience attributes. In total, 6 surveys were created, with some tailored for trainees and thus featuring fewer attribute-related questions. These surveys collectively resulted in the generation of 477 individual lines sent to the Google Form API.

The analysis revealed that, on average, it takes approximately 45.4250 seconds to complete the creation process for a typical survey of this scale. In two additional tests, the scope of the questionnaires was expanded. The first variation included the addition of 2 and 4 stakeholders, resulting in a higher number of elements to be processed by the Google Form API, 642 elements for the former and 809 for the latter. Notably,

Attribute	Desired Target	Actual Performance	Gap/Surplus
Adaptive Capacity	0,72	0,51	0,21
Adaptibility	0,47	0,47	0
Agility	0,67	0,51	0,16
Efficiency	0,58	0,55	0,03
Robustness	0,33	0,52	0,19
Restorability & Redundancy	0,44	0,48	0,04
Overall Resilience	0,54	0,51	0,03

Figure 13 Gaps and Surplus.

Table 1 Table 1Time to create surveys.

X	Test 1	Test 2	Test 3
Normal size	46.1083s	44.5017s	45.6649s
+2 stakeholders	57.9761s	58.7257s	60.7602s
+4 stakeholders	69.82s	73.3267s	73.0954s
+50% questions	47.7093s	47.5060s	49.2858s
+100% questions	50.7977	50.6861s	50.5543s

Table 2 GPT prompts.

X	coherence	length	Resume
"Can you resume these responses :"	3	4	1
"Resume me theses {number of responses} reviews answering '{question}' with the 3 main points within 200 tokens:"	1	1	3
"Resume me theses {number of responses} reviews answering '{question}' with 3 key points within 200 words:"	2	3	2
"There is a training where people answer this question: {question} with theses answers: {responses} Resume in 3 key points theses answers in maximum 200Tokens."	4	2	4
"Find the keywords of these sentences in 200 tokens:"	5	5	5

the average creation time increased to 59.1540 seconds for the former and 72.0807 seconds for the latter, reflecting a discernible rise in processing time.

Conversely, tests involving the augmentation of questions demonstrated interesting results. Specifically, when 50% and 100% more questions were included in the initial questionnaire, expanding the number of rows to 715 and 954 respectively, the impact on creation time was relatively modest. The average times recorded were 48.1670 seconds and 50.6794 seconds respectively. This observation suggests that a significant portion of the time consumption stems from the unique survey creation process using the Google Form API.

These findings collectively shed light on the factors influencing survey creation time, demonstrating that the intricacies of survey composition and structure, particularly when utilizing the Google Form API, significantly impact the efficiency of the process.

In the pursuit of qualitative analysis, a series of diverse prompts were subjected to testing. The results of these tests are illustrated in the provided table. Multiple variations of sentences were tried, with each response evaluated based on key criteria such as coherence, length within the token limit, and the effectiveness of the summarization. Each prompt was iterated five times to enable a thorough assessment, leading to the subsequent ranking. The following table illustrates the comparative rankings of each prompt based on key attributes:

The first and last prompts consistently exceeded the token limit, while the third prompt resulted in three appropriate responses and two that exceeded the length limit. Considering the differing qualities of the responses and the constraints of the token limit, the prompt "Resume me theses {number of responses} reviews answering '{question}' with 3 key points within 200 words:" was selected as the most suitable for effectively summarizing the qualitative question responses.

The presented outcome is a result of the analysis conducted on this prototype, utilizing GPT-3.5 to perform response summarization and extract three key points explicitly mentioned within the surveys.

7. CONCLUSIONS AND FUTURE WORK

In conclusion, inspired by the work of Jnitova et al. (2022), we research into open-source technology to help represent the complex concept of Resilience and that can be used for any enterprise wide resilience capability measure, whether it is related to workforce performance or system and processes. Enable situation awareness of the capability readiness at one glance. This automated analysis and results visualisation represented by the House of Resilience allow quick understanding of strength and weakness of the capability training system, or workforce performance and allow decision making in near real tine. Such work usually took 3 weeks to 3 months to generate. With such tool, it can be generated within few minutes. Provided value for money and save the officers time, so that they can work on the strategic tasks. The adaptation of the House of Quality framework has enabled the identification of both positive and negative aspects, aiding training managers in decision-making. In addition, the House of Resilience serves as a remarkable tool that provides an immediate and comprehensive overview of the overall satisfaction level achieved by the training. With a single glance, the House of Resilience encapsulates the entire spectrum of stakeholder responses, delivering a succinct and impactful portrayal of training satisfaction. The implemented solution encompasses four functionalities within a single application, operating through APIs, which ensures flexibility in the frontend design. Additionally, the utilization of GPT-3.5 for qualitative response summarization enhances the report generation process.

However, certain limitations should be acknowledged. The reliance on GPT-3.5 introduces the potential for errors, particularly when analysing responses generated by ChatGPT itself due to a scarcity of real data. Furthermore, the use of Excel for reporting restricts interactivity, hindering the exploration of second-level analyses as outlined by Jnitova et al. (2022). The Google Form API, being an initial version, exhibits sluggishness. The API's constraints result in one survey per stakeholder, and its inability to adapt surveys based on prior responses may pose challenges. These limitations could potentially be addressed in future research, perhaps by transitioning to a web-based interface. Additionally, attempts to employ machine learning for discovering relationships between subsets of questions and overall training evaluations yielded limited meaningful outcomes, indicating a potential avenue for further investigation.

REFERENCES

1. Aerny-Perreten, Nicole, Ma. Felicitas Domínguez-Berjón, María D. Esteban-Vasallo, et Carmen García-Riolobos. « Participation and Factors Associated with Late or Non-Response to an Online Survey in Primary Care: Late and Non-Response to Online Surveys ». *Journal of Evaluation in Clinical Practice* 21, no 4 (août 2015): 688–93. https://doi.org/10.1111/jep.12367.
2. Agrawal, Rajeev, Anirudh Kadadi, Xiangfeng Dai, et Frederic Andres. « Challenges and Opportunities with Big Data Visualization ». In *Proceedings of the 7th International Conference on Management of Computational and Collective IntElligence in Digital EcoSystems*, 169–73. Caraguatatuba Brazil: ACM, 2015. https://doi.org/10.1145/2857218.2857256.
3. Beniz, D B, et A M Espindola. « Using Tkinter of Python to Create Graphical User Interface (GUI) for Scripts in LNLS », 2017.
4. Bhaskar, Adithya, Alexander R Fabbri, et Greg Durrett. « Prompted Opinion Summarization with GPT-3.5 », juillet 2023.
5. Cambria, Erik, et Bebo White. « Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article] ». *IEEE Computational Intelligence Magazine* 9, no 2 (mai 2014): 48–57. https://doi.org/10.1109/MCI.2014.2307227.
6. Čehovin, Gregor, Michael Bosnjak, et Katja Lozar Manfreda. « Item Nonresponse in Web Versus Other Survey Modes: A Systematic Review and Meta-Analysis ». *Social Science Computer Review* 41, no 3 (juin 2023): 926–45. https://doi.org/10.1177/08944393211056229.
7. Clarke, V., et V. Braun. « Successful qualitative research: A practical guide for beginners. »*Successful Qualitative Research*, 2013.
8. Ghimire, Devndra. « Comparative Study on Python Web Frameworks: Flask and Django », 2010.
9. Gore, Himanshu, Rakesh Kumar Singh, Ashutosh Singh, Arnav Pratap Singh, Mohammad Shabaz, Bhupesh Kumar Singh, et Vishal Jagota. « Django: Web Development Simple & Fast »25, no 6 (2021).
10. Hauser, John. « How PURITAN-BENNETT used the HOUSE OF QUALITY », Sloan management review, janvier 1993.
11. Hauser, John R., et Don Clausing. « The House of Quality ». *Harvard Business Review*, 1 mai 1988. https://hbr.org/1988/05/the-house-of-quality.
12. Hunt, John. *Advanced Guide to Python 3 Programming*. Undergraduate Topics in Computer Science. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-25943-3.
13. Idris, Nuruldelmia, Cik Feresa Mohd Foozy, et Palaniappan Shamala. « A Generic Review of Web Technology: DJango and Flask ». *International Journal of Advanced Science Computing and Engineering* 2, no 1 (7 mai 2021): 34–40. https://doi.org/10.30630/ijasce.2.3.29.
14. Isaac, Ogedengbe, Olugbade Olumide, et Olagunju Rasaki. « Application of House of Quality Matrix to Material Selection for Engineering Designs ». *British Journal of Applied Science & Technology* 10, no 4 (10 janvier 2015): 1–11. https://doi.org/10.9734/BJAST/2015/19105.
15. Jiao, Wenxiang, Wenxuan Wang, Jen-tse Huang, Xing Wang, et Zhaopeng Tu. « Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine ». arXiv, 19 mars 2023. http://arxiv.org/abs/2301.08745.
16. Jnitova, Victoria, M. Efatmaneshnik, DR Keith Joiner, et Elizabeth Chang. « Improving Enterprise Resilience by Evaluating Training System Architecture: Method Selection for Australian Defence », 2021.
17. Jnitova, V., M., Joiner, K.F., Xavier, A., Chang, E., Ferris, T., and Camelia, F. (2024). Is Your Training Service Resilient and Postured to Support Organisational Sustainment? Australian Journal of Multi-Disciplinary Engineering, currently in production to be published.
18. JUGRAN, SWARANJALI, ASHISH KUMAR, BHUPENDRA SINGH TYAGI, et VIVEK ANAND. « Extractive Automatic Text Summarization using SpaCy in Python & NLP ». In *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 582–85, 2021. https://doi.org/10.1109/ICACITE51222.2021.9404712.
19. Liu, X.F. « Software quality function deployment ». *IEEE Potentials* 19, no 5 (décembre 2001): 14–16. https://doi.org/10.1109/45.890072.
20. Sahoo, Kabita, Abhaya Kumar Samal, Professor, Dept. of Comp. Sc. & Engg., Trident Academy of Technology, Bhubaneswar, India., Jitendra Pramanik, Asst. Professor, Centurion University of Technology and Management, Odisha, India., Subhendu Kumar Pani, et Associate Professor, Orissa Engineering College, Bhubaneswar, India. « Exploratory Data Analysis Using Python ». *International Journal of*

Innovative Technology and Exploring Engineering 8, no 12 (30 octobre 2019): 4727–35. https://doi.org/10.35940/ijitee.L3591.1081219.
21. Sawicki, Piotr, Marek Grzes, Fabricio Goes, Dan Brown, Max Peeperkorn, et Aisha Khatun. « Bits of Grass: Does GPT Already Know How to Write like Whitman? »arXiv, 10 mai 2023. http://arxiv.org/abs/2305.11064.
22. Siddiqui, Tamanna, et Mohammad Alkadri. « Review of Programming Languages and Tools for Big Data Analytics ». *International Journal of Advanced Research in Computer Science*, mai 2017.
23. Smyth, Patrick. « Creating Web APIs with Python and Flask », 2018.
24. Soni, Mayank, et Vincent Wade. « Evaluating and Detecting ChatGPT's Responses on Abstractive Summarization ». arXiv, 29 mai 2023. http://arxiv.org/abs/2303.17650.
25. Temponi, Cecilia, John Yen, et W. Amos Tiao. « House of Quality: A Fuzzy Logic-Based Requirements Analysis ». *European Journal of Operational Research* 117, no 2 (septembre 1999): 340–54. https://doi.org/10.1016/S0377-2217(98)00275–6.
26. Tetali, Dharma Reddy. « A PYTHON TOOL FOR EVALUATION OF SUBJECTIVE ANSWERS (APTESA) », 2017.
27. Varela, Carmen, José Ruiz, Ana Andrés, Rubén Roy, Adela Fusté, et Carmina Saldaña. « Advantages and Disadvantages of Using the Website SurveyMonkey in a Real Study: Psychopathological Profile in People with Normal-Weight, Overweight and Obesity in a Community Sample », 2016, 77–89.
28. Vasantha Raju N., et N.S.Harinarayana. « Online survey tools: A case of study of Google Forms », janvier 2016.