



Criteria definition for digital requirements using hesitant fuzzy linguistic terms sets: an application to the automotive industry

Pietro Fronte^{1,2} · Núria Agell¹ · Marc Torrens¹ · Diana Mesa^{2,3}

Received: 2 January 2024 / Accepted: 10 December 2024
© The Author(s) 2024

Abstract

Managing a portfolio of digital products is challenging, particularly in a context of limited economic resources and workforce. Therefore, prioritization of activities and new developments is crucial. In Software Development environment, almost all well-known prioritization techniques are based on experts' knowledge and opinion, leaving little room for a data-driven, objective approach. In this paper, we propose a methodology that adopts the Delphi framework and Hesitant Fuzzy Linguistic Term Sets for collecting experts' opinions, evaluating perceived importance, and computing group consensus. The objective is to provide a framework to define a group-consensual set of relevant criteria that would represent the basis for a data-driven prioritization process for digital requirements. Implementation and results from a real case application in a European automotive company are presented to understand the relevance of criteria and suggest their inclusion or exclusion for prioritization purposes.

Keywords Business value · Prioritization · Hesitant fuzzy linguistic term sets · Automotive · Consensus measure

1 Introduction

Digital transformation encompasses various changes occurring across different levels in both society and enterprises (Vial, 2019; Majchrzak et al., 2016). Organizations undergo internal transformations by adapting strategies, reorganizing, adjusting processes, shaping

✉ Núria Agell
nuria.agell@esade.edu

Pietro Fronte
pietro.fronte@alumni.esade.edu

Marc Torrens
marc.torrens@esade.edu

Diana Mesa
diana.mesa@code.seat

¹ Department of Operations, Innovation and Data Science, ESADE Business School - Ramon Llull University, Avinguda de la Torre Blanca, 59, 08172 Sant Cugat del Vallès, Barcelona, Spain

² CUPRA Digital Business, CUPRA, Autovía A-2, KM-585, 08760 Martorell, Barcelona, Spain

³ SEAT CODE, Carrer de Badajoz, 97, Sant Martí 08018, Barcelona, Spain

culture, and evolving business models (Riasanow et al., 2017; Matt et al., 2015; Chantias et al., 2019; Li et al., 2023). A careful focus on these actions can empower companies to achieve desired performance outcomes, fostering customer-centricity, agility, collaboration, and profitability (Imran et al., 2021). Notably, the manufacturing sectors struggle with the adoption of new digital technologies, compared to more agile sectors like entertainment or Information Technology (IT) (Dremel, 2017). This is mainly attributed to the rigid process and structure of manufacture in contrast with the rapid technical development of digital domain. (Hirsch-Kreinsen, 2016).

Among the well-established manufacturing sectors, the automotive industry stands out as one of the most affected. Digitalization can impact various operations within the automotive sector, including designing, manufacturing, servicing operations, and even the vehicles themselves (Peters et al., 2016; Llopis-Albert et al., 2021). This paper focuses on the latter two dimensions: digitalization of servicing operations and digitalization of the car itself.

The study was carried out at CUPRA, a newly established sports car Original Equipment Manufacturer (OEM) based in Barcelona, and a spin-off of the historic Spanish carmaker SEAT. At CUPRA, the Digital Business department is responsible for managing a portfolio of digital products related to servicing operations and mobile connectivity with vehicles. By "portfolio of digital products," we refer to a collection of software assets that the company utilizes to provide specific services for customers (servicing operations) or between customers and their vehicles (mobile connectivity). Currently, the portfolio of digital products managed by the department includes the Website, Car Configurator, Stock Locator, eCommerce, Mobile App, Private Area, Connected Car, and Customer Relationship Management (CRM) system. This portfolio is designed to support the business by offering online vehicle sales, digital aftersales services, charging services for Electric Vehicles (EVs), and the sale of car functionality packages. It also generates economic benefits for the company through licensing anonymized data and optimizing internal processes.

Effectively managing multiple digital products with limited economic resources and workforce necessitates prioritizing which new digital requirements to develop. To address this, practitioners can utilize various methodologies for prioritizing digital requirements (e.g., MoSCoW, Planning Poker, Cost-of-Delay, RICE, Kano, etc.). However, these methods are often subjective, relying heavily on experts' knowledge and experience (Münch et al., 2019b; Triefflinger et al., 2021). Additionally, given the involvement of multiple stakeholders in the prioritization process, modeling the interests of the various actors can help generate solutions that benefit the entire department (Liesiö et al., 2021).

The paper's contributions are twofold: practical and theoretical. From a practical standpoint, this study provides the Digital Business and Strategy department of CUPRA with a set of consensual criteria. These criteria are representative of all decision-makers in the department who play an active role in generating and estimating new digital requirements. This set of criteria serves as a tool to mitigate individual human bias in estimating the value of single digital requirements. This common foundation paves the way for a standardized, yet representative, asset for future digital requirement value estimation. The paper's theoretical contribution is also twofold: 1) it proposes a versatile framework for modeling stakeholders' preferences, interests, and beliefs, and 2) it applies this framework to define value delivery for customers and the company in Digital Product Development within the automotive industry. These contributions address specific gaps in current research and practice. Recent developments in Portfolio Decision Analysis (PDA) highlight the need for more sophisticated stakeholders modeling approaches. Specifically, (Liesiö et al., 2021) identify the modeling of stakeholder preferences and beliefs as a crucial research direction, emphasizing the importance of moving beyond single decision-maker models to accommodate multiple par-

ticipants, particularly in complex corporate environments. Our framework directly addresses this research opportunity by providing a systematic approach to capture and model diverse stakeholder perspectives. Moreover, we extend our contribution to address another significant gap in current literature: the challenge of defining and measuring value delivery, which has been consistently highlighted by researchers (Lehtola & Kauppinen, 2006; Münch et al., 2019a). Despite its practical significance, project evaluation often lacks consistent criteria for screening and ranking initiatives (Trieflinger et al., 2021).

Through the application of our framework in the automotive domain, we provide a data-driven approach to prioritization that demonstrates how stakeholder modeling can be effectively used to establish clear value criteria in a specific industry context. To fulfill these objectives, we adopt the proposed framework to extract knowledge about stakeholder beliefs regarding what makes a new digital requirement meaningful in terms of delivering value to the customer or the company. By distilling the experts' interests, we outline a set of relevant criteria that reflects not the preferences of a single individual, but a wider perspective of all the relevant decision-makers of the department. This set of criteria lays the foundation for enabling data-driven requirements prioritization, reflecting department-level priorities.

Herrera and Herrera-Viedma (2000) shows that experts often prefer expressing preferences using linguistic models rather than purely numerical values. Linguistic descriptors help capture the uncertainty in decision-making, and the use of hesitant fuzzy linguistic term sets (HFLTS) reflects the natural hesitancy in human reasoning, supporting a more nuanced group decision-making process (Montes et al., 2015). Additionally, some contributions have analyzed the consensus among participants when adopting HFLTS (Wu & Xu, 2016). In line with the practical contribution of the paper, we adopted an amalgamation of the two concepts to derive the set of relevant and consensual criteria to provide to the Digital Business and Strategy department.

In order to get a realistic view of experts' opinion about relevant criteria, we allow them to include hesitancy along with their perception. This is because hesitancy influence how individuals perceive and interpret information relevant to a decision. We propose a methodology based on Delphi process (Dalkey & Helmer, 1963) and Hesitant Fuzzy Linguistic Term Sets (Rodriguez et al., 2012), to determine the set of consensual criteria to adopt when quantifying the overall value of a new digital requirement.

The remainder of this article is structured as follows. Section 2 covers the literature review while Sect. 3 proposes a review of preliminaries. In Sect. 4 we dives into the proposed framework for defining relevant criteria when estimating the added value of a new digital requirement. Section 5 offers a real case application of the framework, providing insight into its relevance within the industry context. Lastly, Sect. 6 concludes the study and depicts next steps of the research.

2 Literature review

Trieflinger et al. (2021) states that many companies are struggling to identify and establish a prioritization process that focuses on delivering value to the customer and the business. It is not clear what delivering value to the client really means and this, as a consequence, impede to draw a data-driven prioritization framework. A common practice is that product backlog items are prioritized by management or expert opinions, sometimes causing the HiPPO (Highest Paid Person Opinion) effect (Münch et al., 2019b; Kohavi et al., 2007). The risk with this approach is that the wrong priorities are set because the criteria for decision-making

are unclear (Münch et al., 2019b). Trieflinger et al. (2021) reviews a variety of prioritization techniques currently adopted in enterprises operating in dynamic and uncertain environments. Their study concludes that these techniques tend to prioritize new developments for testing purposes. Ultimately, only the test results can identify which new developments are truly worth pursuing. Despite that, in literature there is a little research concerning empirical evaluation of these methods in practice (Agren et al., 2022; Kasauli et al., 2021; Trieflinger et al., 2021). In Münch et al. (2019b), an interviewee mentioned that the “*usage of off-the-shelf approaches without tailoring them to the company context are inappropriate to establish successful roadmapping practices.*”

In such context, becomes relevant the definition of value for customers and business in order to help enterprises to take conscious decision about ranking/prioritization of activities (Lehtola & Kauppinen, 2006). Komssi et al. (2011) conducts a research study within two software companies highlighting that 1) linking business strategy to product development is fundamental and 2) improving understanding of customers’ needs promotes the delivery of the appropriate software features. Münch et al. (2019b) states that the panelists interviewed adopted *development effort, costs, value for customers, feasibility, market relevance, and strategic alignment* as relevant criteria for value generation. Despite participants defining criteria at the core of the prioritization process, estimation still relies on expert opinions rather than empirical facts. Münch et al. (2019a) provides a model for evaluating the maturity level of product roadmapping for organization acting in dynamic market environments with high uncertainties. A set of criteria and maturity stages is proposed and validated by experts. Their model seeks to generalize the maturity status of products; therefore, panelists from different digital sectors were involved. Contrarily, the goal of this paper is to provide a narrow view of a digital department of an automotive organization. Racheva et al. (2010) performs an exploratory research to give answers to different questions about how business value and its creation is perceived in the context of Software Development projects. 11 experts from 8 different European companies (banking, ERP for small businesses, health care management, automotive, content management system, online municipality services system) are interviewed and replies are coded to create base concepts for their grounded theory. Results highlight the importance of satisfying customers’ needs and requests, generating a positive Return on Investment (ROI) and lastly the adoption of the *negative value*. Their research highlights also that small enterprises with limited resources are especially mindful of potential missed opportunities when not implementing specific requirements.

Taking into account these studies, our objective, in this paper, is to go one step ahead, strengthening qualitative results with quantitative figures about group-perception of criteria importance and group-consensus. Business value perception may result context-dependent since every expert perceives business value depending on its job and responsibility (Racheva et al., 2010; Mårtensson et al., 2016). Similarly, Liesiö et al. (2021) highlight the need for portfolio decision analysis models that incorporate multiple stakeholders’ preferences, interests, and beliefs, which aligns closely with the focus of this study on developing a multi-stakeholder approach to prioritization in software development.

3 Preliminaries

In this section, we describe the theoretical framework that is needed for the proposed methodology, which is framed within the context of group decision-making scenarios under linguistic assessments. Firstly, we provide an introduction to Delphi technique elucidating its role

as a systematic approach for information elicitation from experts. Secondly, we offer a detailed explanation of fuzzy linguistic models based on Hesitant Fuzzy Linguistic Term Sets (HFLTSS) serving as a means to effectively capture and represent humans' opinions.

3.1 Delphi framework

Delphi was introduced in 1963 by Dalkey and Helmer (1963). This method has commonly been employed to assess or prioritize a range of potential alternatives, uncover implicit assumptions leading to diverse judgments, investigate novel solutions for a particular issue, or establish agreement on a specific subject among a group of experts or stakeholders (Flostrand et al., 2020). The Delphi technique employing several rounds of questionnaires to gather data from a group of experts or panelists is considered as an effective approach for fostering consensus (Hsu & Sandford, 2019). During the initial round, open-ended questions are employed to collect opinions from experts or panelists. The outcomes from this first round are organized into statements, which are then evaluated by the experts in a second round. In subsequent rounds, panelists are presented with the aggregated values of the entire panel and are tasked with re-evaluating their own opinions in light of the collective group perspective. Often, this iterative approach facilitates the achievement of a consensus on key statements within the group (Chalmers & Armour, 2019). A lot of applications of Delphi method can be found in real contexts such as education, medicine, technology, marketing, transportation and management (Chalmers & Armour, 2019; Chan et al., 2016; Hirschhorn, 2019; Ghadami et al., 2021). The Delphi method has some weaknesses and limitations, including the lack of a defined threshold for consensus and the challenge of addressing uncertainties inherent in panelists' opinions (Diamond et al., 2014; Belton et al., 2019). To address these issues, various fuzzy Delphi approaches have been introduced in the literature as potential solutions (Herrera-Viedma et al., 2014).

In this paper, we consider a Delphi method approach using Hesitant Fuzzy Linguistic Terms Sets (HFTLSs). This method allows experts' responses (as observed from the second round) to be expressed using linguistic terms to capture their hesitancy and, in addition, computes a degree of consensus based on these linguistic terms (Yu et al., 2021).

3.2 Hesitant fuzzy linguistic term sets

A summary of the basic concepts related to Hesitant Fuzzy Linguistic Term Sets (HFLTSS) that will be referenced in the methodology are presented in this section. They follow the Hesitant Fuzzy Linguistic Term Sets (HFLTSS) theory introduced by Rodriguez et al. (2012) and its extensions proposed by Montserrat-Adell et al. (2017) and (Porro et al., 2021). The adoption of HFLTSS is tailored for scenarios where linguistic terms are employed to describe perceptions affected by uncertainty (Rodriguez et al., 2012; Du et al., 2023). Linguistic terms increase expressiveness, especially when such terms are an integral part of the assessment. The uncertainty expressed by decision-makers (DMs) or experts is represented by means of HFTLTSs. This allows to capture, in a more flexible way, opinions of experts or DMs and adjust the calculation of central opinion and consensus accordingly.

Let \mathbb{S}_n denote a finite totally ordered set of linguistic terms, $\mathbb{S}_n = \{s_1, \dots, s_n\}$ with $s_1 < \dots < s_n$. The elements of \mathbb{S}_n are considered as the basic linguistic terms, and its cardinal n denotes the granularity of the model.

Definition 1 A HFLTS over the set \mathbb{S}_n , is a subset of consecutive basic linguistic terms of \mathbb{S}_n such that $[s_i, s_j] = \{x \in \mathbb{S}_n | s_i \leq x \leq s_j\}$ with $i, j \in \{1, \dots, n\}$ and $i \leq j$. If $i = j$, $[s_i, s_i]$ is the singleton $\{s_i\}$. Moreover, \mathbb{S}_n and the empty set $\{\} = \emptyset$ are also considered HFLTSs and they are called the full and the empty HFLTSs respectively.

Moving from basic linguistic terms to non-basic ones translates into introducing some degree of uncertainty on the opinions expressed, losing therefore precision. The least precise linguistic term coincides with the full HFLTS, i.e. $[s_1, s_n] = \mathbb{S}_n$. The set of all the possible, non-empty HFLTSs over the original set \mathbb{S}_n is denoted by $\mathbb{H}_{\mathbb{S}_n}$.

In $\mathbb{H}_{\mathbb{S}}$, two partial order relations are introduced, allowing to compare opinions represented by linguistic terms with respect to their precision (\subseteq) or their preference (\leq_P) (Abuasaker et al., 2023). Given $H_1 = [s_i, s_j]$, $H_2 = [s_k, s_l]$ we have that:

$$\begin{aligned} H_1 \text{ is more precise than or equal to } H_2 (H_1 \subseteq H_2) &\iff i \geq k \wedge j \leq l \\ H_1 \text{ is more preferred than or equal to } H_2 (H_1 \geq_P H_2) &\iff i \geq k \wedge j \geq l \end{aligned}$$

Example 1 Let $S = \{s_1, s_2, s_3, s_4, s_5\}$ a set of basic terms, being $s_1 = \text{“Strongly Disagree”}$, $s_2 = \text{“Mildly Disagree”}$, $s_3 = \text{“Unsure”}$, $s_4 = \text{“Mildly Agree”}$, and $s_5 = \text{“Strongly Agree”}$. If we consider the responses given by three DMs when they evaluate a criteria as: $H_1 = \text{“Non agree”}$, $H_2 = \text{“Unsure or Mildly Agree”}$ and $H_3 = \text{“Not strongly disagree”}$, they can be represented in hesitant linguistic terms by: $H_1 = [s_1, s_3]$, $H_2 = [s_3, s_4]$ and $H_3 = [s_2, s_5]$. According to the responses provided in the example and the partial relations introduced, we can state that: 1) H_1 is less preferred than H_2 ; 2) H_2 is more precise than H_3 being included in H_3 but we have no reasons to prefer it to the latter one.

The operator connected union (\sqcup) of two HFLTSs is defined in this context as the least element of $\mathbb{H}_{\mathbb{S}_n} \cup \{\emptyset\}$, based on the subset inclusion relation (\subseteq), that contains both HFLTSs. The connected union together with the intersection provide to the set $\mathbb{H}_{\mathbb{S}_n} \cup \{\emptyset\}$, a lattice structure, as proved in Montserrat-Adell et al. (2017).

Based on $(\mathbb{H}_{\mathbb{S}_n}, \sqcup, \cap)$ lattice structure, a distance between HFLTSs, as reported in Ruiz et al. (2022), is defined in the following way.

Definition 2 Given $H_1, H_2 \in \mathbb{H}_{\mathbb{S}_n}$, the distance between H_1 and H_2 is defined as:

$$d(H_1, H_2) = 2 \cdot \text{card}(H_1 \sqcup H_2) - \text{card}(H_1) - \text{card}(H_2) \quad (1)$$

where $\text{card}(H)$ corresponds to the number of basic elements contained in H .

Example 2 Let us consider the HFLTSs corresponding to the opinions of the two first DMs given in Example 1 and compute the distance between them. First, we must compute $H_1 \sqcup H_2 = [s_1, s_4]$, and then following the previous definition, we obtain:

$$d(H_1, H_2) = 2 \cdot \text{card}([s_1, s_4]) - \text{card}([s_1, s_3]) - \text{card}([s_3, s_4]) = 8 - 3 - 2 = 3$$

This distance between HFLTSs enables the calculation of a central measure or centroid from a set of opinions of a set of DMs or experts in a panel, each of them providing their opinions over a specific alternative in the form of HFLTS. We define the centroid of a set of HFLTSs as the HFLTS which minimizes the sum of distances to all the elements of the set.

Definition 3 Let r be the number of DMs and $G = \{H_1, \dots, H_r\}$ be a set of HFLTSS, corresponding to their opinions, then, the centroid of the set G is defined as:

$$H^C = \arg \min_{H \in \mathbb{H}_S} \sum_{j=1}^r d(H, H_j) \tag{2}$$

Note that the centroid is an interval central measure for ordinal scales with hesitancy, serving as a representation of aggregated opinions. It functions as the “center of gravity” of all expert judgments, accounting for both discrete assessments and interval responses that express uncertainty. The centroid represents the aggregated collective judgment of our expert panel regarding each criterion’s relevance. However, it’s important to note that, similar to the median, the uniqueness of the centroid is not guaranteed. In order to ease the calculation of the centroid, in Porro et al. (2021), it is proved that H^C can be calculated as:

$$H^C = \{[s_L, s_R] \in \mathbb{H}_S \mid L \in \mathbb{M}(s_1^L, \dots, s_k^L), R \in \mathbb{M}(s_1^R, \dots, s_k^R)\} \tag{3}$$

where s_i^L and s_i^R are respectively the left and the right values of the term H_i and $\mathbb{M}(\cdot)$ represents the median qualitative label when r is odd, or the two central values when r is even.

In the latter case, if either s_L, s_R or both contain multiple values, several centroids can be considered as valid options. In this case, in this paper, in order to capture the highest level of hesitancy, the connected union (\sqcup) of the two sets is taken as centroid position.

Example 3 Following Example 1 where the opinions of the three DMs on a criteria are: $H_1 = [s_1, s_3], H_2 = [s_3, s_4]$ and $H_3 = [s_2, s_5]$, we obtain $H^C = [s_2, s_4]$. This hesitant term is the one that minimizes the addition of distances to the three given opinions H_1, H_2 and H_3 .

Finally, evaluating all the distances between DMs opinion and the centroid H^C , and summing them up, it is obtained a measure of *disagreement* among DMs. This allows the definition of the *degree of agreement*, or measure of consensus, δ of the DMs’ panel, with respect to the centroid. The consensus degree aims to quantify the cohesion of expert opinions around the centroid, providing crucial context for interpretation.

Definition 4 Let $G = \{H_1, \dots, H_r\}$, be a set of HLTS, corresponding to the opinions of a group of DMs, then, the degree of agreement (or consensus) among DMs is defined as:

$$\delta_G = 1 - \frac{\sum_{j=1}^k d(H^C, H_j)}{r \cdot (n - 1)} \in [0, 1] \tag{4}$$

where n is the number of basic linguistic terms considered and r is the number of DMs in the panel. Note that $0 < \delta < 1$ due to $r \cdot (n - 1)$ is an upper bound of the addition of distances between the centroid and the HFLTSS DMs opinion (Montserrat-Adell et al., 2017). Note that the measure of agreement, or consensus, is affected by the DMs hesitancy expressed in their opinions. In this measure we consider all the possible values contained in DMs hesitant opinion.

Example 4 Following the previous examples, we can compute the degree of agreement among the answers given by the three DMs:

$$\delta = 1 - \frac{(d(H_1, H^c) + d(H_2, H^c) + d(H_3, H^c))}{(3 \cdot (5 - 1))} = 1 - \frac{(2 + 1 + 1)}{12} = \frac{2}{3}$$

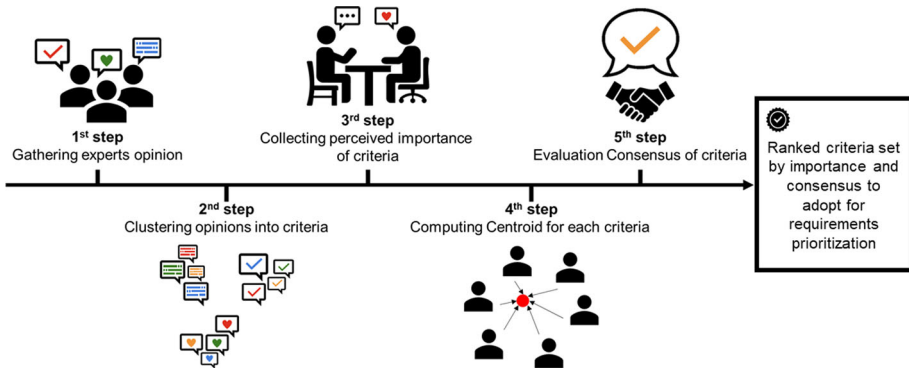


Fig. 1 Visual summary of the five steps in the proposed framework

The centroid information is complemented by the consensus degree measure in refining the final ranking of the obtained criteria. Criteria are primarily ranked by their centroid value, according to the partial order introduced in Subsection 3.2. In cases where centroid intervals are not comparable or are identical, the criterion with higher consensus is given preference. The consensus degree quantifies the cohesion of expert opinions around the centroid, providing crucial context for results interpretation. This dual-metric approach (perceived importance and consensus) ensures we capture not just the central tendency of expert opinions (through centroids) but also their coherence through agreement measure.

4 Criteria definition for digital requirements added-value estimation

In this study, we present the methodology for defining a set of relevant criteria to evaluate the added value that new digital requirements can contribute to the overall portfolio of digital products. Focusing on the automotive sector, specifically on the CUPRA Digital Business and Strategy department, the aim is to extract and organize insights from experts to establish a standardized framework for assessing digital requirements. Prior to this study, decision-makers were independently responsible for creating and scoring their own requirements, leading to a fragmented decision-making process based on individual judgment. This research seeks to unify the evaluation process by providing a consensual set of criteria for future decision-making. The primary objective of this study is to deliver these criteria, which will be used to evaluate digital requirements. In the future, this set of criteria will serve as the foundation for a multi-criteria decision-making (MCDM) algorithm designed to compute scores for each requirement. The output of this study includes two key elements: (1) a ranked and consensual set of evaluation criteria and (2) the ranking itself, which can be used to derive the weights of the criteria to adopt in the future MCDM algorithm. To ensure a robust selection of criteria, relevant professional profiles were identified for interviews, with the Delphi method employed to gather expert opinions on what defines added value. The acquired insights are organized, and group consensus is achieved using Hesitant Fuzzy Linguistic Term Sets (HFLTS).

In the following subsections, we present the methodology, which is structured in five steps (see Fig. 1), along with the framework assumptions and limitations.

4.1 The proposed methodology

First step: Gathering experts opinions We start our study with an exploratory approach, carrying out a first round of 1-to-1 interviews characterized by open-ended questions. The goal is to build a comprehensive understanding of what truly matters at the individual level when evaluating a new proposed digital requirement. Given that panelists come from diverse backgrounds and have different responsibilities, each participant provides their own perspective and judgment on priorities and importance. To achieve this, the first round of interviews uses simple open-ended questions to gather as much information as possible from decision-makers, capturing their unique perceptions and nuances, and identifying the factors that are most important to them when evaluating new digital requirements. The interviews always started with the same question “When evaluating a new digital requirement, which are the relevant factors you are considering?”. This was the starting point for the people interviewed to freely express their opinion and beliefs. Occasionally, when they provided answers with little or no details, they were asked to elaborate a bit more on the reason behind their reply. For this reason, no framework was used to generate open-ended questions. Due to confidentiality requirements, interviews were not audio-recorded, but notes were taken during each session.

Second step: Clustering opinions into criteria. The coding process was conducted through a structured approach involving two independent coders from the Digital Portfolio team (more details about the department structure are given in section 5.1). There are three main reasons for this choice:

1. **Small dataset:** We are handling replies from only 16 decision-makers discussing a very specific topic. The limited number of participants makes the manual coding approach feasible in terms of workload and time required.
2. **Technical content:** The collected replies reflect specific technical context and dynamics of how the department operates both internally and in collaboration with other providers of the Volkswagen Group. Therefore, it was valuable to have internal staff code the answers and categorize them into concepts familiar to the audience.
3. **Confidential information:** For confidentiality reasons, meetings were not recorded; instead, summaries of replies were annotated. The resulting notes lack context, making them difficult to interpret and challenging to code properly. However, Portfolio team employees have an objective yet informed perspective to understand the context and properly cluster opinions into meaningful classes.

Given this context, we determined that internal expertise would be crucial for accurate interpretation of the collected data. Overall, the coding process can be resumed in the 3 following stages:

1. **Preparation:** All interview notes were transferred to virtual sticky notes using Miro online software (Miro (formerly RealtimeBoard), 2024). No pre-defined categories or concepts were established to avoid bias.
2. **Independent Coding Sessions:** Two separate 45-minute coding sessions were conducted with two different employees from the Digital Portfolio team. Each employee independently reviewed the sticky notes and clustered them into emerging concepts. This approach allowed for organic identification of themes from the data.
3. **Criteria Development:** Through an iterative process, initial concepts were refined and consolidated. Similar concepts were merged, and broader themes were identified. This process naturally led to the emergence of our main criteria (Economic Impact, Time Criticality, Beneficiary) and their sub-criteria.

The two independent coding sessions produced remarkably similar results, which were then combined to form the final criteria set presented.

Third step: Collecting perceived importance of criteria The second round of interviews involved a semi-structured conversation lasting approximately 30–40 min each, in which participants were presented with the criteria coded from the first round of interviews. These criteria, which had been previously clustered by two colleagues from the Portfolio team, were presented sequentially in the same order as shown in Tables 2 and 4. The order of criteria presentation was not specifically chosen for any particular reason and it does not reflect any perception of importance; it was simply the output obtained at the end of Step 2 of the proposed framework. After presentation and explanation, DMs were asked to reply to the question “Do you think criterion X is an important factor when evaluating a digital requirement Y?”. The structured reply allows them to express their opinion through a 7-points linguistic scale composed of $\mathbb{S} = \{s_1 = \text{“Strongly Disagree”}, s_2 = \text{“Disagree”}, s_3 = \text{“Somewhat Disagree”}, s_4 = \text{“Neutral”}, s_5 = \text{“Somewhat Agree”}, s_6 = \text{“Agree”}, s_7 = \text{“Strongly Agree”}\}$ and optionally add personal comments to articulate their response. Comments and explanations were welcomed to understand the reasons behind their choices.

Since the objective is to obtain differentiation of attitudes, preferences, or perceptions, a 5 or 7-point would be considered appropriate taking into account what reported in current literature (Revilla et al., 2014). In this specific case, as we were working with a panel of specialized respondents, we selected a 7-point Likert scale because it was a framework they felt more comfortable with. As defined in Definition 1, to collect hesitancy in their opinions, they are allowed to choose an interval of linguistic terms, if convenient. This strategy facilitates the expression of a sentiment (positive, negative, neutral, don’t know) and permits to detect eventual hesitancy in their beliefs.

Fourth step: Computing centroids for each criteria Assuming R being the number of panelists and V the number of criteria to evaluate, the outcome of the third step is the matrix \mathbb{T} of results.

$$\mathbb{T} = \{[s_i, s_j]_k\}_v \text{ for } r = 1, \dots, R; v = 1, \dots, V \quad (5)$$

In other words, for each criterion v we obtain a set of K intervals provided by all panelists. Each interval is a tuple of linguistic terms $[s_i, s_j]$ where s_i, s_j are respectively the *lower* and *upper* bounds of the interval, following Definition 1 of HFLTS. For each criterion v , the centroid is then evaluated. Centroids are calculated by applying the Eq. (3), as defined in the preliminaries.

Fifth step: Evaluation of consensus for each criteria Lastly, after computing the centroids, we evaluate the group consensus of panelists with respect to each criterion by applying the Eq. (4). In the second term of the group-consensus equation, the numerator is obtained using the distance defined in Eq. (1), adopted as the distance measure between the centroid and the panelists’ opinion.

Once the results are obtained, we order the criteria that should be adopted for prioritizing new digital requirements with respect to the business interests. Alternatively, based on the obtained results, we can recommend to discard some criteria and not include them into the set of relevant variables to measure.

4.2 Framework assumptions and limitations

We acknowledge that the Delphi method and HFLTS approach have inherent limitations, such as problems with open-ended questions, the abstract nature of linguistic terms that can be interpreted subjectively, the influence of question formulation, the dependence on proper

term set design, among others. While the main objective of this research was not to address these weaknesses comprehensively, we did attempt to mitigate them where possible:

- **Open-ended questions:** Open-ended questions in Delphi can lead to ambiguous or unfocused responses. Experts may interpret and answer the same question differently, making it difficult to synthesize consistent results. In our study, we mitigated this risk by relying on expert professionals who were interviewed about common aspects of their daily job. This approach helped to ground the questions in the participants' practical experience, somehow limiting the risks typically associated with open-ended inquiries in other contexts.
- **Abstract nature of linguistic terms:** The subjective nature of linguistic terms can generate variability in how participants interpret them. Linguistic labels are weighted differently by each participant based on their experience, background, and interpretation. This can lead to misalignment in responses, undermining the reliability of the consensus. To address this, we encouraged experts to use interval responses when they felt uncertain, allowing for more nuanced expression of their opinions. This approach provided flexibility for experts to convey the degree of certainty in their assessments.
- **Influence of question formulation:** The way questions are formulated can influence responses. We approached this by posing the same question in the same order to all interviewees, aiming to acquire knowledge of their perceived importance across different criteria. This consistency in question formulation helped to minimize variations in interpretation and maintain focus on the criteria being evaluated.

5 Real case application in the automotive industry

Digital transformation in automotive industry affects a variety of operations, like servicing operations and the car itself, among many others. In this section we take a closer look about how digital servicing and the digital side of vehicles shape new business streams for an automotive company, and how such business is managed. At CUPRA, the Digital Business Department can be summarized with 3 main pillars: Portfolio, Business and Products. These pillars are interconnected and their efforts and outcomes are highly correlated. All the work is channeled into two main objectives: 1) Generating business through digital products by selling core products and new services and 2) Maintaining and constantly improving our digital products. Intuitively, Business area is the driver for objective 1, while Product area sponsors objective 2. Portfolio area, on the other hand, serves as enabler for the correct development of the underlying strategy, acting as the treasurer and supervisor of the department's efficiency. Subsection 5.1 dives into the digital business from a managerial point of view in CUPRA, while Subsection 5.2 presents the details about the real case application carried out at the company. Lastly, the obtained results are analyzed in Subsection 5.3.

5.1 Industry context

Nowadays, traditional businesses are exploring new opportunities to broaden their portfolio of profitable activities. In the automotive sector, with the advent of digital technology, there is a concerted effort to monetize new digital services related to the core product: the car. For instance, eCommerce enables the online sale of cars and accessories, while after-sales services such as maintenance and workshop reservations can be managed through the car's infotainment system and mobile app. Bundles of digital functionalities for connected cars

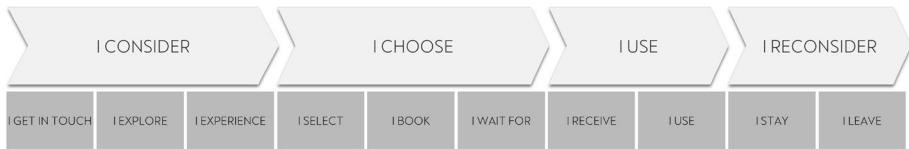


Fig. 2 4-phases customer journey adopted in SEAT CUPRA

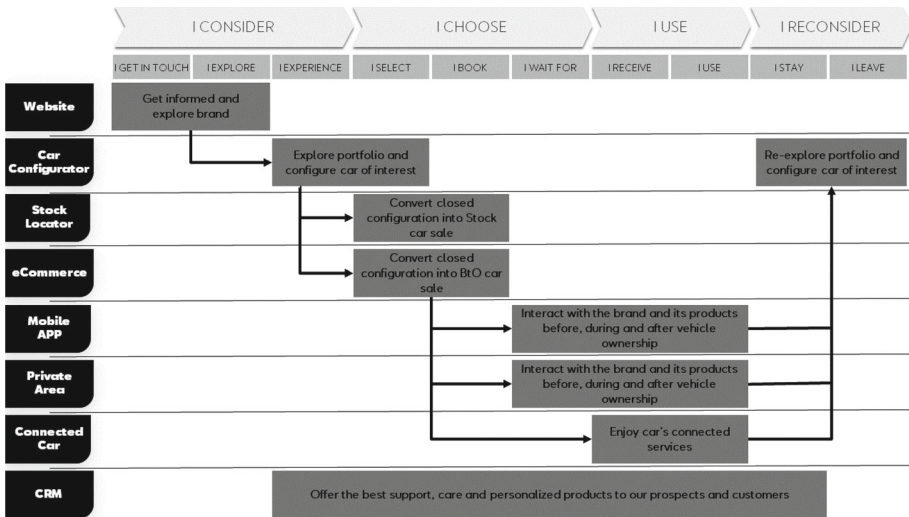


Fig. 3 Relation digital products as customer support along the E2E customer journey

can also be purchased, and charging plans for electric vehicles (EV) can be contracted by customers through their mobile app or website. These activities are not meant to replace the core business but rather to accompany and support it, thereby expanding the portfolio of revenue streams.

Digital products serve as new touch-points between customers/potential customers and the company. Adopting an End-to-End (E2E) perspective of the Customer Journey (CJ), the objective is to provide support to customers (or potential customers, depending on their position in the journey) through appropriately tailored digital products for each journey phase. The considered CJ includes 4 phases: Consideration, Choice, Usage, and Re-consideration. Such CJ extends from the initial phase in which a client is seeking a new car (Consideration), to the conclusion of their relationship with the car (Re-consideration). This encapsulates the essence of the E2E approach.

A detailed analysis of each phase of the customer journey allows for the identification of general actions that customers are likely to perform within each phase, as illustrated in Figure 2. Guided by this customer journey, the Digital Business Department adopts the customer-centricity principle as the core basis for product development. In simpler terms, digital products are crafted around customers' needs and interests, focusing solely on what is perceived as relevant for them to complete the actions they expect, as highlighted by Agren et al. (2022). Figure 3 provides further details on this principle, illustrating how each digital product is conceived as a supportive tool to enable customers to seamlessly navigate throughout the E2E customer journey.

Digital requirements reflect the underlying strategy. Depending on the goal to be achieved (business-oriented goal, digital product improvement, hybrid), the Business and/or Product areas define new requirements to develop. This is a crucial point where the business strategy intersects with the evolution of the current product portfolio (Trieflinger et al., 2021). The definition of a new digital requirement is an iterative process, continuously refined and adapted with the support of the Portfolio area to ensure the best fit with the department's Objectives and Keys Results (OKRs). However, it is not solely about accomplishing the strategy; critical events such as customer complaints, the entry of new cars into the market, or changes in regulatory policies may shift the focus of the work. Therefore, the establishment of a proper multi-criteria-based prioritization methodology becomes necessary.

5.2 Experimental setup

To construct a general framework, framed within the department's activities and strategy, a total of 16 relevant professional employees were selected to conduct the interviews. Prior to this study, the department's decision-makers were already creating new digital requirements and estimating their values based on their expertise and knowledge. All the people involved in such a process were invited to participate in this research. Additionally, multiple tiers of the vertical hierarchy of the department were involved: Business Manager (Tier 1), Lead Business Owner (Tier 2), and Business Owner (Tier 3). Concerning the Business Team, the related activities include digital sales, digital aftersales, charging business, data-based process optimization, and connected car business. On the other hand, the Products area encompasses the website, private area, car configurator, stock locator, eCommerce, Mobile App, and connected car system. One or more representatives have been selected to cover all the currently running activity streams. A summary of the panelists involved in this study is presented in Table 1.

5.3 Results and discussion

During the initial round of interviews, open-ended questions are asked to the participants. The objective is to establish an informal environment where unfiltered opinions can be shared. Interviews are conducted in person with an average duration of 30 min and answers were annotated. Raw data collected during interviews were then coded internally with the support of two employees of Portfolio area (not involved in interview process), clustering opinions into families. The criteria collected after the first round of interviews and successively internally coded are reported in Table 2.

Successively, the second round of interviews is carried out with the intent of collecting the perceived importance of panelists with respect to the coded criteria. As reported in step 4 of Sect. 4, the interview is semi-structured and perceived importance in shape of linguistic terms is collected. The results are listed in Table 3.

Lastly, by applying Eqs. (3) and (4), centroids of perceived importance and consensus are evaluated for each criteria. As specified in Eq. (3), in case of an even number of panelists K , s^L and s^R can result by two basic linguistic terms. When the two central values coincide, a single label is reported; otherwise the two labels are shown. Consensus and centroids results are reported in Table 4.

Table 1 Overview of profiles selected and interviewed for defining the set of meaningful criteria when estimating the added value of a new digital requirement

Panelist ID	Role	Area	Activity
1	Business manager	Portfolio	Connected car portfolio
2	Business owner	Portfolio	Connected car portfolio
3	Business manager	Business	All business streams
4	Business manager	Business	Online sales and aftersales
5	Lead business owner	Business	Online sales
6	Business owner	Business	Online sales
7	Lead business owner	Business	Online aftersales
8	Business owner	Business	Online aftersales
9	Business owner	Business	Data licencing
10	Business owner	Business	Connected fleet business
11	Business owner	Business	Commercial connected services
12	Business owner	Business	Connected services
13	Business owner	Business	Charging services
14	Product manager	Products	Website and private area
15	Product manager	Products	Car configurator, stock locator, eCommerce
16	Product manager	Products	Mobile apps

To lastly retrieve a single centroid position we have considered incorporating maximum hesitancy. Analyzing the example of Strategy criteria

$$H_{Strategy}^C = [s_{Strategy}^L, s_{Strategy}^R] = [s_5, s_6]$$

Similarly, in the case of *Platforms' usage* we have

$$H_{Platforms'usage}^C = [s_{Platforms'usage}^L, s_{Platforms'usage}^R] = [s_5, s_6]$$

We replicate the same adjustment to all criteria with multiple centroids. Then we sort them considering first their centroid value and secondly by their consensus degree. The results are shown in tabular form in Table 5 and graphically in Fig. 4.

Results demonstrate that experts, relatively to the CUPRA Digital Business department, perceive the economic impact of new digital requirement as the most important criteria to take into account, either in the form of *Direct Income* or *Cost Saving*. Both of this criteria show solid results, having achieved a centroid value of s_7 with a high degree of group-consensus (respectively 95.23% and 87.50%). Another factor of relevance is *Markets/Importers*. This points out how important it is to satisfy Importers needs, in terms of new digital requirements, maximizing the number of Importers involved.

In terms of perceived importance, slightly below the three criteria afore-mentioned, we find the *Related Critical Event* with a group-consensus of 75% and a centroid value of $[s_6, s_7]$. This criteria is perceived as critical in order to incorporate in decision-making processes the sense of urgency that some requirements may inherit from external events like future cars' release, new governmental policies to fulfill, or any other related event with a specific deadline.

A variety of criteria obtained an importance centroid value s_6 , all of them with pretty solid consensus degree. *Customer satisfaction* obtained a 80.20% of consensus demonstrating

Table 2 Criteria families, sub-criteria and brief descriptions as results of the first round of interviews and coding task

Criteria family ID	Criteria	Sub-criteria ID	Sub-criteria	Definition
c ₁	Economic impact	c _{1.1}	Direct income	The new digital requirement enables a new revenue stream capable of generating a direct income
		c _{1.2}	Cost saving	The new digital requirement allows to optimize a process and, thus, cutting costs
		c _{1.3}	Loss prevention	The new digital requirement allows to prevent eventual sale loss
c ₂	Strategy	–	–	The new digital requirement has strong/weak impact on Strategy
c ₃	Platform usage	–	–	The new digital requirement has an impact on Platforms' usage, monitored by specific KPIs
c ₄	User experience	–	–	The new digital requirement has a positive impact on the platforms' user experience
c ₅	Customer satisfaction	–	–	The new digital requirement focuses on solving customers' active complains
c ₆	Time criticality	c _{6.1}	Related critical event	The new digital requirement is related to a relevant event (car's start of production, new policies, etc..)
		c _{6.2}	Margin vs deadline critical event	Time between the delivery of a new digital requirement and the occurrence of a critical event
c ₇	New opportunity enabled	–	–	What kind of new opportunity the new digital requirement is enabling
c ₈	Beneficiary	c _{8.1}	Markets/Importers	How many markets are affected by the new digital requirement. Markets are differentiated by volume sales
		c _{8.2}	Car's owners	How many customers/car owners are seen affected by the new digital requirement
		c _{8.3}	Prospects	How many prospects, in pre-sales funnel, are affected by the new digital requirement
c ₉	Competitor status	–	–	What competitors are doing about the current digital requirement

Table 3 Perceived importance collected from the 16 interviewed for each of the criteria considered. Partial results of the step 4, as defined in Sect. 4

Panelist ID	c1		c2	c3	c4	c5	c6		c7	c8		c9
	c1.1	c1.2					c1.3	c6.1		c6.2	c8.1	
1	s7	s7	s7	s5	s5	s5	s6	s6	s3	[s2, s3]	[s2, s3]	[s2, s3]
2	s7	s6	s7	s4	s7	s7	s5	s5	s3	s5	s7	s3
3	s7	s6	s6	s5	s7	s6	s5	[s5, s7]	s2	s6	s6	s4
4	s6	s6	s6	s6	s6	s6	s5	s5	s2	s2	s2	s2
5	s7	s7	s5	[s6, s7]	s4	[s6, s7]	s7	[s6, s7]	s5	s5	s7	s4
6	s5	s6	s6	s5	[s4, s5]	s6	[s6, s7]	[s5, s6]	[s3, s4]	s4	s3	s6
7	s7	s7	s7	s5	s3	s6	s7	s7	s3	s3	s3	s6
8	s7	s7	s7	s7	s6	s7	s7	s7	s6	s7	s7	[s3, s4]
9	s7	s7	[s5, s6]	[s5, s6]	s3	s3	s7	[s5, s6]	s6	,s7	s7	s3
10	s7	s7	s7	s7	s7	s7	s7	s7	s5	s7	s7	s4
11	s7	s7	s7	s2	s5	s5	s7	s7	s4	s7	s7	s7
12	s7	s7	s7	s6	[s3, s5]	[s3, s5]	s6	s6	s5	s5	s5	s6
13	s7	s6	s6	s4	s5	s5	s6	[s6, s7]	s4	s6	s7	[s1, s7]
14	s7	s7	s5	s6	s5	s5	s5	[s5, s6]	s3	s4	s5	s6
15	s7	s6	[s5, s6]	[s4, s5]	s6	s6	s7	s7	s4	s6	s7	[s2, s3]
16	s7	s7	s5	[s5, s7]	s6	s6	s7	s7	s5	s7	s7	s6

Table 4 Centroids (left and right bounds) of perceived importance and consensus evaluated from the opinions collected from the experts

	Centroid s^L	Centroid s^R	Group-consensus
Economic impact			
Direct income	s_7	s_7	95.83%
Cost saving	s_7	s_7	87.50%
Loss prevention	s_6	s_6	77.08%
Strategy	s_5	$\{s_5, s_6\}$	71.88%
Platforms' usage	$\{s_5, s_6\}$	$\{s_5, s_6\}$	69.79%
User experience	s_6	s_6	76.04%
Customer satisfaction	s_6	s_6	80.20%
Time criticality			
Related critical event	s_6	$\{s_6, s_7\}$	75.00%
Margin versus deadline critical event	s_6	s_6	75.00%
New opportunity enabled	s_4	s_4	73.95%
Beneficiary			
Markets/importers	s_7	s_7	84.38%
Car owners/car segment	$\{s_5, s_6\}$	s_6	59.38%
Prospects	s_6	s_6	73.95%
Competitor status	s_4	s_4	71.88%

Table 5 Centroids of perceived importance and consensus evaluated from the opinions collected from the experts

	Centroid	Group-consensus
Direct income	s_7	95.23%
Cost saving	s_7	87.50%
Markets/importers	s_7	84.38%
Related critical event	$[s_6, s_7]$	75.00%
Customer satisfaction	s_6	80.20%
Loss prevention	s_6	77.08%
User experience	s_6	76.04%
Margin versus deadline critical event	s_6	75.00%
Prospects	s_6	73.95%
Strategy	$[s_5, s_6]$	71.88%
Platforms' usage	$[s_5, s_6]$	69.79%
Car owners/car segment	$[s_5, s_6]$	59.38%
New opportunity enabled	s_4	73.95%
Competitor status	s_4	71.88%

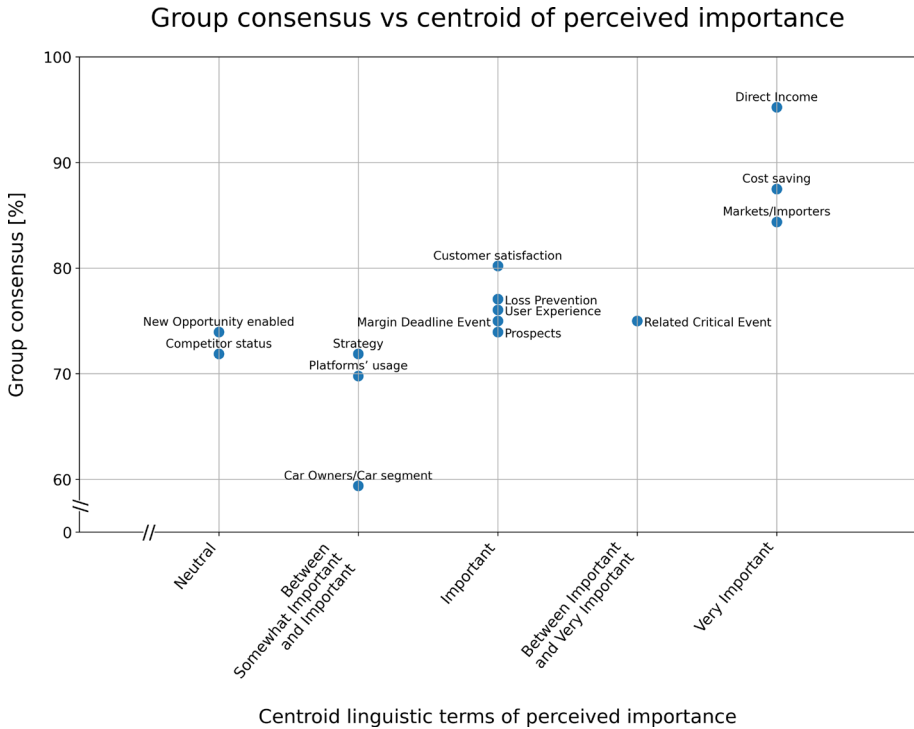


Fig. 4 Scatterplot of centroid positions for criteria perceived importance with respect to group-consensus

the high general perceived relevance that the voice of customer has when developing new digital requirements of introducing in development some bug fix to solve any current issue that is causing dissatisfaction. In just a range of 2 percentage points, from 77% to 75%, we have Loss prevention, User Experience and Margin vs Deadline Event. The first, as already mentioned, stresses the significance of the economic aspect in new digital requirements, even though it is perceived as less fundamental than Direct Income or Cost Saving. This is because it incorporates a hypothesis, thus introducing uncertainty, regarding the sales volume that would not be lost if the digital requirement was to be developed. User Experience (UX), on the other hand try to reward those requirements that are aimed at solving/improving UX aspects (improve accessibility features, foster seamless integration of multiple digital products along the Customer Journey, improve shopping flow based on User Test results, etc.). Margin vs Deadline event is recognized as a dynamic component that could gradually increase its value as we get closer to the deadline, allowing requirements related to any close external events to outstand from the rest. Just below 74% of consensus we find *Prospect*, to give emphasis to those digital requirement affect greater volume of prospects, according to the Digital Funnel towards the sale.

Lowering the centroid importance to the value $[s_5, s_6]$ we obtain 3 criteria, two of which with a consensus around 70% and the third one below the 60%. Strategy is the criterion that achieved the highest degree of consensus (71.88%), This aligns with the emphasis on promoting requirements that support strategic objectives, rather than decisions influenced by the HiPPO (Highest Paid Person's Opinion) effect. Secondly we have *Platform's usage* (69.79%), to weights differently requirements that will have greater/lower impact of platforms

KPIs (Number of Web visits, Car Configurator Conversion Rate, Stock Locator Conversion Rate, Bounce Rate, etc.). It is interesting to observe a lower importance/consensus in this criterion, as a-priori estimations of the impact that new requirements will have on KPIs are difficult to calculate, thus increasing the uncertainty of realization. The last one is *Car Owners/Car segment* with a consensus of 59.38%. It is interesting to see how this criterion, despite having a good value of centroid for importance, received such poor consensus. Many feedback collected pointed into the same direction "requirements that affect the car are supported by a positive economic impact, thus such relevance should already be accounted inside Direct Income. By adding also this criterion would represent a double-counting of the added-value that the requirement would bring overall."

Lastly, two criteria obtained a centroid value of $s_4 = \text{"Neutral"}$, with a discrete support by the majority of panelists: *New opportunity enabled* (73.95%) and *Competitors' status* (71.88%). In the first case, panelists didn't find meaningful to have a set of opportunity enabled by the new digital requirement objectively defined. The majority of them though this criterion would add subjectivity to the whole set variables, scoring it accordingly. For what concern *Competitors' status*, the feedbacks generally pointed into a clear direction: "the brand should not develop new requirements just because other competitors are exploring certain businesses. The company must be consistent with its values and with its own strategy".

The outcome of this research points out that the economic impact of a new digital requirements along with the Markets' interests are the most relevant and solid criteria to consider. Also, aspects related to Customers satisfaction and time criticality of new developments are perceived as relevant. On the contrary, we identified some criteria that were mentioned during the first round of interviews but finally didn't get enough consensus or importance from the panelists. Therefore we can suggest discarding criteria *Competitors' status*, *Opportunity type* and *Car Owner/Car segment* from the criteria basis. For the first two, there was a solid consensus about their perception as *Neutral*, among all the other criteria. Considering that one of the objective of the study is to identify meaningful criteria, we suggest to not include these two criteria into the criteria basis. For what concern the criterion *Car Owner/Car segment*, the low consensus degree and the feedback received suggest us to not recommend this criterion, to avoid possible conflicts or disagreement in future prioritization outcomes.

The predominance of DMs from the Business domain reflects the department's primary objective: generating profit by offering services. This panel composition may have influenced the centroid and consensus results, but it also mirrors the department's priorities. If our set of consensual criteria is meant to represent the department's composition and priorities, then this distribution is expected and appropriate. Given the practical contribution of the study, a collective meeting was organized with the stakeholders who participated in it. The objective of the meeting was to collect stakeholders' feedback about the results, as well as facilitate the exchange of different perspectives and perceptions among them. Overall, the study results were very well received, and participants were particularly enthusiastic about the systematic structure of the study and the output obtained. Results were presented via Image 4 and Table 5, and numbers were first introduced before moving to the feedback session. They found the top 5 criteria particularly interesting, along with their respective positions in the ranking. Most participants saw their beliefs reflected in the ranking, highlighting how the top tier of the list encompasses the principal customers benefiting from the CUPRA Digital Department's deliveries: the company itself, the markets and final customers. The first two criteria, Direct Income and Cost Savings, reflect the company's economic interests; the Markets/Importers criterion points out the relevance that national companies have in the deliveries that the central company makes. "Importers are the gateway to reach the final customers in every market we operate in, so we should carefully listen to their needs and requests," stated a stakeholder

during the collective meeting. The Related Critical Event criterion is relevant for setting the department's roadmap and being prepared to face every milestone the company establishes, without missing important deadlines that can affect operations (new regulations, policies, interdependencies with other providers, new car launches, etc.). Finally, the Customer Satisfaction criterion in position 5 clearly states the relevance that end users have in the company's priorities. "Word of mouth is a very powerful tool, and we cannot allow unsatisfied customers out there. A happy customer will potentially buy a new CUPRA when the time to change cars comes" a participant stated. Additionally, an interesting exchange of opinions emerged regarding the Strategy criterion in position 10. Some participants wondered why the ranking position was so low, given that Strategy is a fundamental priority to consider. On the other hand, other participants stated that their importance perception was low since every new future development should always be consistent with the set strategy, consequently making such criterion irrelevant. Despite the constructive exchange of opinions, they finally all agreed that having such a criterion in the set can benefit the department in further promoting those new developments that have a strong fit with the company strategy.

6 Conclusion, limitation and future Work

In this paper, we propose a new framework for defining relevant criteria to take into account when estimating the added-value of a new digital requirement to develop, addressing in a small part the research gap highlighted in Trieflinger et al. (2021). We improve current digital requirements development by providing a methodology that incorporate multiple stakeholders' preferences, interests, and beliefs, as suggested in Liesiö et al. (2021), and the inclusion of uncertainty in experts' opinions which pave the way for a more realistic and fair data-driven prioritization process. Despite the results shown are tailored for a specific company where the study was carried out, they could be a reference for the automotive sector. However, they should not necessarily be generalized for the entire sector, since different automotive companies can adopt different strategies according to the historic context and environment they're operating.

The new framework demonstrates its effectiveness as a tool for extracting expert knowledge and building a relevant set of criteria. Moreover, through its twofold perspective-importance and consensus-it effectively highlights which criteria should not be marked as relevant for the prioritization process. While the real-case application results are specific, the proposed framework can be replicated in various contexts to extract knowledge from expert groups.

Despite its utility, we acknowledge certain limitations, such as the abstract nature of linguistic terms in the HFLTS approach, leading to subjective interpretations, and the influence of question formulation on responses. Future research could focus on standardizing linguistic terms in HFLTS, exploring innovative question formulation techniques, and improving the analysis of complex, open-ended responses. The use of specialized software, such as Le Sphinx, for textual analysis could also enhance the process and enable a more objective coding of responses. Additionally, increasing the number of rounds of expert interviews may further ensure a higher consensus in final prioritization. Also, comparing this framework with other methodologies and replicating it in different sectors would validate its effectiveness and improve generalizability. We also recognize the absence of a detailed comparison between our proposed methodology and existing frameworks. While this was beyond the scope of the current study, we agree that such a comparative analysis-focusing on factors like the number

of rounds, challenges faced by panelists, or variations of the Delphi method-would provide valuable insights. However, the time constraints and availability of the Decision Makers limited the feasibility of such comparisons in this study. This will be addressed in future research, where a comprehensive analysis of different methodologies will be conducted to highlight key differences in outputs rather than definitive measures of efficiency. These future research directions aim to address the current limitations while expanding the approach's applicability and robustness across various decision-making contexts.

Finally, while the current study focuses on defining and ranking the criteria, the selection and implementation of the appropriate MCDM algorithm will be the next step. The Multi-Attribute Value Theory (MAVT) framework is considered a strong candidate for future implementation due to its interpretability and flexibility, making it well-suited for the problem and the intended users.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study is funded by Agencia de Gestión de Ayudas Universitarias y de Investigación (AGAUR) - Generalitat de Catalunya - within the framework *Doctorats Industrials* with grant number 2021-DI-012. It has been also partially supported by the PERCEPTIONS Research Project (PID2020-114247GB-I00), funded by the Spanish Ministry of Science and Information Technology.

Declarations

Conflict of interest The authors declare that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abuasaker, W., Nguyen, J., Ruiz, F. J., Sánchez, M., & Agell, N. (2023). Perceptual maps to aggregate assessments from different rating profiles: A hesitant fuzzy linguistic approach. *Applied Soft Computing*, *147*, 110803.
- Agren, S. M., Heldal, R., Knauss, E., & Pelliccione, P. (2022). Agile beyond teams and feedback beyond software in automotive systems. *IEEE Transactions on Engineering Management*, *69*, 3459–3475. <https://doi.org/10.1109/TEM.2022.3146139>
- Belton, I., MacDonald, A., Wright, G., & Hamlin, I. (2019). Improving the practical application of the delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technological Forecasting and Social Change*, *147*, 72–82. <https://doi.org/10.1016/j.techfore.2019.07.002>
- Chalmers, J., & Armour, M. (2019). The delphi technique. In P. Liangputtong (Ed.), *Handbook of research methods in health social sciences* (pp. 715–735). Singapore: Springer Singapore.
- Chan, J., Ghose, A., Seamans, R. (2016). The internet and racial hate crime: Offline spillovers from online access. *MIS Quarterly*, *40*(2), 381–404, [2023-12-26]<https://www.jstor.org/stable/26628911>

- Chanias, S., Myers, M. D., & Hess, T. (2019). Digital transformation strategy making in pre-digital organizations: The case of a financial services provider. *The Journal of Strategic Information Systems*, 28(1), 17–33. <https://doi.org/10.1016/j.jsis.2018.11.003>
- Dalkey, N., & Helmer, O. (1963). An experimental application of the delphi method to the use of experts. *Management Science*, 9(3), 458–467.
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401–409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>
- Dremel, C. (2017). *Barriers to the adoption of big data analytics in the automotive sector*. <https://www.alexandria.unisg.ch/handle/20.500.14171/102119>
- Du, J., Xie, N., Liu, S., & Goh, M. (2023). Grey linguistic term sets for decision-making. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05319-0>
- Flostrand, A., Pitt, L., & Bridson, S. (2020). The delphi technique in forecasting- a 42-year bibliographic analysis (1975–2017). *Technological Forecasting and Social Change*, 150, 119773. <https://doi.org/10.1016/j.techfore.2019.119773>
- Ghadami, N., Gheibi, M., Kian, Z., Faramarz, M. G., Naghedi, R., Eftekhari, M., & Tian, G. (2021). Implementation of solar energy in smart cities using an integration of artificial neural network, photovoltaic system and classical delphi methods. *Sustainable Cities and Society*, 74, 103149. <https://doi.org/10.1016/j.scs.2021.103149>
- Herrera, F., & Herrera-Viedma, E. (2000). Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and systems*, 115(1), 67–82.
- Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., & Pedrycz, W. (2014). A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17, 4–13. <https://doi.org/10.1016/j.inffus.2013.04.002>
- Hirschhorn, F. (2019). Reflections on the application of the delphi method: lessons from a case in public transport research. *International Journal of Social Research Methodology*, 22(3), 309–322. <https://doi.org/10.1080/13645579.2018.1543841>
- Hirsch-Kreinsen, H. (2016). Digitization of industrial work: development paths and prospects. *Journal for Labour Market Research*, 49(1), 1–14. <https://doi.org/10.1007/s12651-016-0200-6>
- Hsu, C.-C., & Sandford, B. A. (2019). The delphi technique: making sense of consensus. *Practical assessment, research, and evaluation*. <https://doi.org/10.7275/pdz9-th90>
- Imran, F., Shahzad, K., Butt, A., & Kantola, J. (2021). Digital transformation of industrial organizations: Toward an integrated framework. *Journal of change management*, 21(4), 451–479. <https://doi.org/10.1080/14697017.2021.1929406>
- Kasauli, R., Knauss, E., Horkoff, J., Liebel, G., & de Oliveira Neto, F. G. (2021). Requirements engineering challenges and practices in large-scale agile system development. *Journal of Systems and Software*, 172, 110851. <https://doi.org/10.1016/j.jss.2020.110851>
- Kohavi, R., Henne, R.M., & Sommerfield, D. (2007). Practical guide to controlled experiments on the web: listen to your customers not to the hippo. *Proceedings of the 13th acm sigkdd international conference on knowledge discovery and data mining* (pp. 959–967). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1281192.1281295>
- Komssi, M., Kauppinen, M., Töhönen, H., Lehtola, L., & Davis, A.M. (2011). Integrating analysis of customers' processes into roadmapping: The value-creation perspective. *2011 IEEE 19th international requirements engineering conference* (pp. 57–66). <https://doi.org/10.1109/RE.2011.6051662>
- Lehtola, L., & Kauppinen, M. (2006). Suitability of requirements prioritization methods for market-driven software product development. *Software Process: Improvement and Practice*, 11(1), 7–19. <https://doi.org/10.1002/spip.249>
- Li, S., Gao, L., Han, C., Gupta, B., Alhalabi, W., & Almakdi, S. (2023). Exploring the effect of digital transformation on firms' innovation performance. *Journal of Innovation & Knowledge*, 8(1), 100317. <https://doi.org/10.1016/j.jik.2023.100317>
- Liesjö, J., Salo, A., Keisler, J. M., & Morton, A. (2021). Portfolio decision analysis: Recent developments and future prospects. *European Journal of Operational Research*, 293(3), 811–825.
- Llopis-Albert, C., Rubio, F., & Valero, F. (2021). Impact of digital transformation on the automotive industry. *Technological Forecasting and Social Change*, 162, 120343. <https://doi.org/10.1016/j.techfore.2020.120343>
- Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for digital transformation: Lessons for information systems research from the study of ict and societal challenges. *MIS Quarterly*, 40(2), 267–278.
- Mårtensson, T., Ståhl, D., & Bosch, J. (2016). Continuous integration applied to software-intensive embedded systems—problems and experiences. P. Abrahamsson, A. Jedlitschka, A. Nguyen Duc, M. Felderer,

- S. Amasaki, and T. Mikkonen (Eds.). *Product-focused software process improvement: 17th international conference, profes 2016, trondheim, norway, november 22-24, 2016, proceedings 17* (pp. 448–457). Springer International Publishing.
- Matt, C., Hess, T., & Benlian, A. (2015). Digital transformation strategies. *Business & information systems engineering*, 57, 339–343. <https://doi.org/10.1007/s12599-015-0401-5>
- Miro (formerly RealtimeBoard). (2024). *Miro*. San Francisco, CA. <https://miro.com> (Online collaborative whiteboard platform, version: Online)
- Montes, R., Sánchez, A. M., Villar, P., & Herrera, F. (2015). A web tool to support decision making in the housing market using hesitant fuzzy linguistic term sets. *Applied Soft Computing*, 35, 949–957.
- Montserrat-Adell, J., Agell, N., Sánchez, M., Prats, F., & Ruiz, F. J. (2017). Modeling group assessments by means of hesitant fuzzy linguistic term sets. *Journal of Applied Logic*, 23, 40–50. <https://doi.org/10.1016/j.jal.2016.11.005>
- Münch, J., Triefflinger, S., & Lang, D. (2019). The product roadmap maturity model deep: Validation of a method for assessing the product roadmap capabilities of organizations. S. Hyrynsalmi, M. Suoranta, A. Nguyen-Duc, P. Tyrväinen, and P. Abrahamsson (Eds.), *Software business: 10th international conference, icsob 2019, jyväskylä, finland, november 18–20, 2019, proceedings 10* (pp. 97–113).
- Münch, J., Triefflinger, S., & Lang, D. (2019). What’s hot in product roadmapping? key practices and success factors. X. Franch, T. Männistö, and S. Martínez-Fernández (Eds.), *Product-focused software process improvement: 20th international conference, profes 2019, barcelona, spain, november 27–29, 2019, proceedings 20* (pp. 401–416).
- Peters, S., Chun, J.-H., & Lanza, G. (2016). Digitalization of automotive industry-scenarios for future manufacturing. *Manufacturing Review*, 3, 1. <https://doi.org/10.1051/mfreview/2015030>
- Porro, O., Agell, N., Sánchez, M., & Ruiz, F. J. (2021). A multi-attribute group decision model based on unbalanced and multi-granular linguistic information: An application to assess entrepreneurial competencies in secondary schools. *Applied Soft Computing*, 111, 107662. <https://doi.org/10.1016/j.asoc.2021.107662>
- Racheva, Z., Daneva, M., Sikkel, K., & Buglione, L. (2010). Business value is not only dollars—results from case study research on agile software projects. M. Ali Babar, M. Vierimaa, and M. Oivo (Eds.), *Product-focused software process improvement: 11th international conference, profes 2010, limerick, ireland, june 21-23, 2010. proceedings 11* (pp. 131–145). Springer Berlin Heidelberg.
- Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree-disagree scales. *Sociological methods & research*, 43(1), 73–97.
- Riasanow, T., Galic, G., & Bohm, M. (2017). Digital transformation in the automotive industry: Towards a generic value network. In proceedings of the 25th european conference on information systems (ecis), guimaraes, portugal, june 5-10 (p.3191-3201).
- Rodriguez, R. M., Martinez, L., & Herrera, F. (2012). Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, 20(1), 109–119. <https://doi.org/10.1109/TFUZZ.2011.2170076>
- Ruiz, F., Agell, N., & Sanchez, M. (2022). A qualitative approach for aggregating people’s perceptions. In: Proceedings of the qr 2022 35th international workshop on qualitative reasoning: co-located at international conference on artificial intelligence (ijcai’22): Vienna, july 23rd, 2022 (p.41–44). Escola Superior d’Administració i Direcció d’Empreses (ESADE).<http://hdl.handle.net/2117/379372>
- Triefflinger, S., Munch, J., Bogazkoy, E., Eissler, P., Schneider, J., & Roling, B. (2021). How to prioritize your product roadmap when everything feels important: A grey literature review. *2021 ieee international conference on engineering, technology and innovation (ice/itmc)* (p.1-9). IEEE. <https://ieeexplore.ieee.org/document/9570243>
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>. (SI: Review issue).
- Wu, Z., & Xu, J. (2016). An interactive consensus reaching model for decision making under hesitation linguistic environment. *Journal of Intelligent & Fuzzy Systems*, 31(3), 1635–1644.
- Yu, W., Zhang, Z., & Zhong, Q. (2021). Consensus reaching for magdm with multi-granular hesitant fuzzy linguistic term sets: a minimum adjustment-based approach. *Annals of Operations Research*, 300(2), 443–466. <https://doi.org/10.1007/s10479-019-03432-7>