



BANCA D'ITALIA  
EUROSISTEMA

## Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

How do you pay? Consumer profiling through payment diaries  
and Artificial Intelligence

by Giorgia Rocco and Camilla Savarese

October 2026

Number

95



BANCA D'ITALIA  
EUROSISTEMA

## Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

How do you pay? Consumer profiling through payment diaries  
and Artificial Intelligence

by Giorgia Rocco and Camilla Savarese

Number 95 – October 2026

*The papers published in the 'Markets, Infrastructures, Payment Systems' series provide information and analysis on aspects regarding the institutional duties of Banca d'Italia in relation to the monitoring of financial markets and payment systems and the development and management of the corresponding infrastructures in order to foster a better understanding of these issues and stimulate discussion among institutions, economic actors and citizens.*

*The views expressed in the papers are those of the authors and do not necessarily reflect those of Banca d'Italia.*

*The series is available online at [www.bancaditalia.it](http://www.bancaditalia.it).*

*Printed copies can be requested from the Paolo Baffi Library:  
[richieste.pubblicazioni@bancaditalia.it](mailto:richieste.pubblicazioni@bancaditalia.it).*

*Editorial Board:* STEFANO SIVIERO, PAOLO DEL GIOVANE, MASSIMO DORIA,  
GIUSEPPE ZINGRILLO, PAOLO LIBRI, GUERINO ARDIZZI, PAOLO BRAMINI, FRANCESCO COLUMBA,  
LUCA FILIDI, TIZIANA PIETRAFORTE, ALFONSO PUORRO, ANTONIO SPARACINO.

*Secretariat:* YI TERESA WU.

ISSN 2724-6418 (online)  
ISSN 2724-640X (print)

Banca d'Italia  
Via Nazionale, 91 - 00184 Rome - Italy  
+39 06 47921

*Designed and printing by the Printing and Publishing Division of Banca d'Italia*

# HOW DO YOU PAY? CONSUMER PROFILING THROUGH PAYMENT DIARIES AND ARTIFICIAL INTELLIGENCE

by **Giorgia Rocco\*** and **Camilla Savarese\***

## Abstract

This study proposes a classification of Italian consumers based on their payment habits, developed through a hybrid approach using Artificial Intelligence techniques. Five behavioural profiles were defined *ex ante* to represent the full spectrum of payment methods, ranging from a predominant use of cash to the adoption of the most advanced digital solutions. Respondents to the Eurosystem's 2024 SPACE survey were then assigned to these profiles through a procedure combining human assessment, Large Language Models and supervised learning algorithms. Most consumers display 'traditional' behaviour, characterized by the use of cash and cards at physical points of sale and a limited use of remote payment channels. Some 24 per cent of consumers do not use cash, while 15 per cent are well integrated into the digital economy. A minority of the population relies predominantly on cash (7 per cent), while an even smaller share (4 per cent) uses it exclusively in combination with prepaid cards. The socio-demographic characteristics of users in the different behavioural profiles are heterogeneous: for example, younger people are more prevalent among those integrated into the digital ecosystem, whereas older people predominate among those who are cash-oriented. Overall, the findings suggest that the transition towards digital payments is not uniform, adopting alternatives to cash does not necessarily entail using more innovative payment solutions, and cash continues to play a significant role. The approach proposed suggests that payment *policy* initiatives, including those relating to Central Bank Digital Currencies, should take these differences into account.

**JEL Classification:** D12, E42, E58.

**Keywords:** Payment behaviour, consumer profiling, Artificial Intelligence, Large Language Models, financial inclusion, digital payments.

## Sintesi

Questo studio propone una classificazione dei consumatori italiani sulla base delle loro abitudini di pagamento, ottenuta mediante un approccio ibrido basato sull'impiego di tecniche di intelligenza artificiale. Sono stati definiti *ex ante* cinque profili comportamentali, rappresentativi dell'intero spettro delle modalità di pagamento, da un utilizzo prevalente del contante fino all'adozione delle soluzioni digitali più avanzate. I rispondenti dell'indagine SPACE 2024 dell'Eurosistema sono stati successivamente assegnati a questi profili mediante una procedura che combina valutazione umana, Large Language Models e algoritmi di apprendimento supervisionato. La maggioranza dei consumatori adotta un comportamento "tradizionale", caratterizzato dall'utilizzo di contante e carte presso i punti vendita fisici e da un uso limitato dei canali di pagamento a distanza. Il 24% dei consumatori non

---

\* Directorate General for Payments and Market Infrastructures, Banca d'Italia.

usa contante e il 15% è ben integrato nell'economia digitale. Una minoranza della popolazione si affida prevalentemente al contante (7%), mentre una percentuale ancora inferiore (4%) lo utilizza esclusivamente in combinazione con le carte prepagate. Le caratteristiche socio-demografiche degli utenti che ricadono nei diversi profili comportamentali sono eterogenee: ad esempio, tra quelli integrati nell'ecosistema digitale sono presenti più giovani, mentre tra quelli orientati al contante prevalgono gli anziani. In sintesi, i risultati suggeriscono che: la transizione verso i pagamenti digitali non è uniforme; adottare comportamenti alternativi al contante non implica necessariamente il ricorso a soluzioni di pagamento più innovative; il contante continua a mantenere un ruolo rilevante. L'approccio proposto suggerisce che le iniziative di *policy* nel campo dei pagamenti, incluse quelle relative alle Central Bank Digital Currency, dovrebbero tenere conto di queste eterogeneità.

## CONTENTS

|                                                                          |    |
|--------------------------------------------------------------------------|----|
| <b>1. INTRODUCTION</b>                                                   | 7  |
| <b>2. LITERATURE REVIEW ON PAYMENT HABITS</b>                            | 9  |
| <b>3. FROM THEORY TO DATA: THE CLASSIFICATION FRAMEWORK</b>              | 11 |
| 3.1 Define profiles of interest                                          | 11 |
| 3.2 Data                                                                 | 13 |
| 3.3 Assign a profile to each individual: targeting the training dataset  | 15 |
| 3.4 Extension to the full sample: supervised classification              | 20 |
| <b>4. UNDERSTANDING PAYMENT PROFILES IN ITALY</b>                        | 22 |
| 4.1 Behavioural dimensions: indicators underpinning group classification | 23 |
| 4.2 Socio-demographic characteristics                                    | 25 |
| 4.3 Financial profile                                                    | 26 |
| 4.4 Openness to innovation                                               | 27 |
| <b>5. BEYOND THE HARD CLASSIFICATION: A SOFT-LABEL PERSPECTIVE</b>       | 29 |
| 5.1 Construction of soft labels                                          | 29 |
| 5.2 Classification confidence and behavioural overlap                    | 30 |
| <b>6. POLICY IMPLICATIONS</b>                                            | 33 |
| <b>7. CONCLUSIONS</b>                                                    | 35 |
| <b>REFERENCES</b>                                                        | 37 |
| <b>APPENDIX A: LIST OF VARIABLES</b>                                     | 41 |
| <b>APPENDIX B: INTERACTING WITH LARGE LANGUAGE MODELS</b>                | 48 |



---

# 1 Introduction<sup>1</sup>

Digital innovation is reshaping payment systems by introducing new solutions and instruments that respond to the evolving needs of consumers. At the same time, cash continues to play a significant role, particularly in the euro area (Zamora-Pérez *et al.*, 2024). Italy is a paradigmatic example of this dichotomy with consumers very often paying in cash even when cashless instruments are preferred (Di Iorio and Rocco, 2025). In this regard, central banks and policymakers need to fully understand payment behaviour to ensure safer and more efficient payment systems alongside guaranteeing continued access to cash for all segments of the population. A comprehensive understanding of payment behaviour is also relevant for the design of Central Bank Digital Currencies (CBDCs) and for initiatives aimed at enhancing financial and digital literacy.

The existing literature has investigated multiple aspects of payment behaviour, giving rise to several strands of research. The focus has been mainly on payment choice and its determinants, often relying on granular diary data (Bagnall *et al.*, 2016). Other studies have examined the role of incentives and pricing mechanisms, the adoption of digital and cashless alternatives, and the persistence of cash usage as a factor of resilience and inclusion. Recent contributions have explored the potential uptake of CBDCs and the influence of sociodemographic and social factors, including peer effects, gender gaps, and generational differences. A detailed discussion of these contributions is provided in Section 2.

These approaches focus on specific aspects of payment behaviour or sociodemographic population groups. Nevertheless, consumers tend to display coherent payment habits across several dimensions, including the payment instrument used, the channels of purchase and the different use cases (e.g. consumer-to-business and peer-to-peer transactions), as well as the possession of and access to various payment instruments and individual preference. For example, consumers who rely on cash often prefer to shop in physical stores and do not use alternative instruments, and those who use digital tools are also more frequently engaged in online shopping. An approach that considers user profiles, i.e. groups of consumers defined by their payment habits, could therefore provide a more comprehensive basis for policy actions addressing users' unmet payment needs.

Recent studies have included the segmentation of the population through the exploration of clustering (Reijerink, 2024), vulnerability definitions (Falcettoni and Nygaard, 2025; Greene *et al.*, 2024), and theoretical segmentation from policy discussion (Henry *et al.*, 2023). These contributions provide valuable insights into payment behaviour and highlight the relevance of user heterogeneity. However, these approaches have some trade-offs: purely data-driven techniques can generate groups that are not relevant for payment policy when interpreted *ex post*. Conversely, a mere conceptual segmentation of the population based on theoretical considerations may not necessarily be supported by empirical evidence.

In this paper we propose a new hybrid classification framework to classify payment habits that combines theory and Artificial Intelligence (AI) techniques. We

---

<sup>1</sup>We would like to thank the referee, Stefano Siviero, Claudio Impenna, Paola Giucca, Michele Lanotte, Marcello Pericoli, Andrea Nobili and Guerino Ardizzi for their helpful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Italy. All the remaining errors are ours.

---

first define five behavioural profiles *ex ante*, drawing on theoretical considerations, policy needs and knowledge of the Italian payment landscape. Then, we assign a label to a subsample of Italian users through two complementary approaches that run in parallel at the outset: (i) manual labeling, and (ii) automated classification using Large Language Models (LLMs). We then extend this classification to the whole Italian sample using multiple supervised learning algorithms, ensuring scalability and alignment with the original segmentation. The resulting hard classification provides clearly interpretable behavioural groups that facilitate empirical characterisation and policy analysis, even if payment behaviour does not always fit perfectly into discrete categories. To this end, we also use probabilistic soft labels to capture uncertainty and overlapping memberships. The approach illustrates how uncertainty emerging from hybrid human–AI annotation can be exploited to obtain probabilistic measures of behavioural overlap.

The methodology is applied to data from the Study on the Payment Attitudes of Consumers in the Euro Area (SPACE 2024; ECB, 2024). The SPACE survey provides granular information on observed behaviours from payment diaries, and many other information on payment habits, notably preferences and possession of and access to different payment instruments, from the accompanying questionnaire.

The analysis carried out for Italy reveals a heterogeneous distribution across five behavioural profiles. The majority of consumers belongs to a dominant behavioural profile represented by the *Traditionalists*, who combine cash and cards at physical points of sale (POS) and make a limited use of remote channels. Around this core group, payment behaviours diverge along different dimensions. Some consumers are completely independent from cash (*Digital-Only users*) but rely primarily on traditional debit and credit cards, whereas others are fully integrated into the digital economy using different instruments and channels (*Well-Connected users*). At the opposite end of the spectrum, only a relatively small minority remains weakly integrated into the banking system, relying primarily on cash (*Cash lovers*) or prepaid cards (*Bank-tech skeptics*). The soft classification analysis further reveals that the dominant Traditionalist profile is also the broadest behavioural category. Meanwhile many Digital-Only users remain closer to traditional card-based payment habits without necessarily positioning themselves at the frontier of innovation. These findings indicate that the transition towards digital payments is not merely a shift from cash to digital payment methods. Instead, consumers follow heterogeneous pathways towards digitalisation: cash remains an important payment instrument for a large share of the population, while cashless behaviour does not necessarily coincide with the adoption of the most innovative payment solutions.

The remainder of this paper is structured as follows. Section 2 reviews the literature on payment behaviour. Section 3 describes the methodological framework, detailing the data sources and the adopted hybrid approach. Section 4 presents and discusses the findings for Italy with the characteristics of the identified payment profiles. Section 5 complements the hard classification by introducing a probabilistic perspective based on soft labels and behavioural overlap across profiles. Section 6 discusses the policy implications emerging from the results. Finally, Section 7 summarizes the main insights.

---

## 2 Literature review on payment habits

The coexistence of traditional and digital payment habits has given rise to a voluminous literature in the field of payments. A strand of literature explores the persistence of cash usage, highlighting its role in resilience and inclusion. Zamora-Pérez *et al.* (2024) find that many individuals regularly use cash regardless of their access to digital tools, and that the demographic profile of people with limited access to digital instruments is highly heterogeneous rather than representing a single “vulnerable group”, while Shy (2023) reviews the research economists do on consumer holding and use of cash underlining the persistent relevance of cash as a store of value and its unique role in ensuring payment system resilience and financial inclusion. Moreover, van der Crujssen and Reijerink (2024) further explores how structural barriers to digital payments contribute to a persistent digital divide, thus reinforcing the need to preserve access to cash as a matter of financial inclusion.

Other studies have examined the adoption of digital and cashless alternatives, including mobile payments (Felt *et al.*, 2024), e-wallets and BNPL services (Bian *et al.*, 2023). Bounie and Camara (2020), Brown *et al.* (2022), Trütsch (2020) studied the convenience of electronic payments especially using the contactless technology showing that it increases the use of cashless instruments. In line with these findings, Brown *et al.* (2023) show that raising the “tap-and-go” limit significantly increased the use of contactless payments.

Recently, several papers have focused on the adoption of CBDCs. Henry *et al.* (2023) considers whether a CBDC would address unmet payment needs in a cashless society, while Li (2023) assesses the potential demand for a CBDC based on its potential design attributes. Bijlsma *et al.* (2024) and Elsinger *et al.* (2025) explore consumer preferences for the adoption of the digital euro (the CBDC project for the Eurosystem) in the Netherlands and Austria, respectively, assessing the role of privacy, incentives and knowledge.

Focusing on payment behaviour and choice, diary surveys, in which consumers record their payment behaviours over a short period, are considered the gold standard in this field due to the detailed nature of the data they collect. As demonstrated in studies such as Bagnall *et al.* (2016), the choice of instrument at the point of sale (POS) is heavily influenced by transaction characteristics. For instance, the use of cash diminishes with transaction size (Bounie and François, 2006; Klee, 2008; Świecka *et al.*, 2021; Von Kalckreuth *et al.*, 2014), as well as with merchants’ acceptance of cards, particularly in small-value transactions (Bounie *et al.*, 2017; Di Iorio and Rocco, 2025; Huynh *et al.*, 2014; Wakamori and Welte, 2017). Cash-related variables also influence the payment choice. For example, the change received in the form of banknotes and metal coins discourages the use of cash (Chen *et al.*, 2019) and the individual cash attitude measured by the cash-in-hand at the beginning of the diary heightens the probability of paying cash (Alvarez and Lippi, 2017; Arango *et al.*, 2015b; Borzekowski *et al.*, 2008). A more recent work by Lippi and Moracci (2025) adds a structural perspective to this literature, showing that the choice between cash and card is shaped not only by transaction size and cash availability, but also by their interaction.

In line with this strand of research, Rysman *et al.* (2025) use household scanner data to study the long-term shift towards payment cards, showing that idiosyncratic household preferences play a central role in instrument choice. Their results indi-

---

cate that the effect of transaction size, even though it is relevant, is smaller and more heterogeneous across households than previously documented, reinforcing the importance of stable user-specific attitudes in shaping payment behaviour.

Many studies emphasise the role played by incentives, discount and surcharges in the payment choice. Ching and Hayashi (2010) and Carbó-Valverde and Liñares-Zegarra (2011) show that payment card rewards significantly modify payment choice in favour of cards; Arango *et al.* (2015a) find that participation in a credit card rewards program induces a shift towards credit card usage at the expense of both debit cards and cash. Stavins (2018) highlights that cash discounts increase the probability of using cash also by a consumer who prefers other payment methods. Finally, Arango-Arango *et al.* (2018) provide evidence that the relative cost of cards compared with cash and differences in ATM withdrawal costs, are key factors explaining why cash is still extensively used across many developed economies.

Looking at the social context and to specific socio-demographic groups, van der Cruijssen and Knoblen (2021) investigated whether payment behaviour is influenced by peer effect, finding that the people mirror the payment choices of others. The gender gap in using digital payments is discussed in van der Cruijssen *et al.* (2025), highlighting the importance of policies aimed at improving fraud knowledge, digital literacy, and financial literacy, especially among women. The attention is oriented also to payment habits of youngest: Clipal and Zamora-Pérez (2025) show that younger people exhibit average transactional use of cash but play a leading role in precautionary cash storage and express strong view that cash is important.

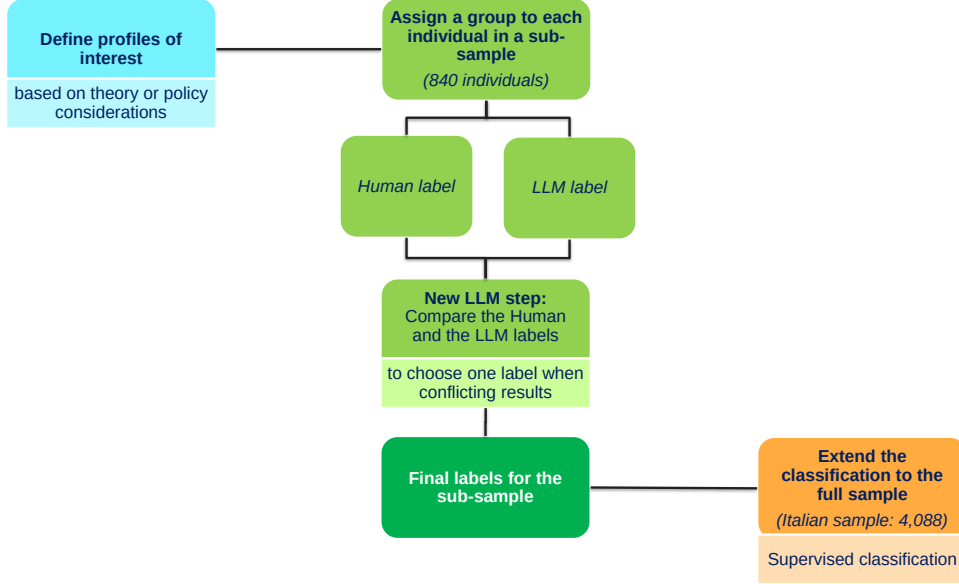
Recent studies have started to classify or discuss users in groups, with the aim of identifying stable behavioural patterns that are relevant to policy design. Reijerink (2024) applies clustering techniques to transaction data to uncover behavioural segments, while Henry *et al.* (2023) discusses the relevance of user profiles in the context of CBDC design. Elsinger *et al.* (2025) explore digital euro adoption among groups defined by their use of cash and cashless instruments, categorised as cash, intermediate, or card payers. In parallel, Greene *et al.* (2024) emphasise the importance of defining underserved users in digital payments, highlighting that vulnerability is not uniform and depends on access and behavioural patterns. This aligns with broader discussions on financial inclusion, which stress the need to distinguish between the “unbanked” and the “out-of-banking population” (Falcettoni and Nygaard, 2025), as these distinctions have direct implications for payment instrument adoption and policy interventions. These studies have either relied on clustering to identify groups from data, refined the definition of vulnerable users, or used theoretical profiles that summarise payment behaviour for policy discussions. We combine these strands by defining a set of predefined behavioural profiles and assigning them to individual respondents using survey data. This allows us not only to interpret payment behaviours in a structured way, but also to quantify the prevalence of these groups and examine the sociodemographic patterns associated with them.

### 3 From theory to data: the classification framework

Our methodology combines theory-driven profile definition, LLM-assisted labelling, and supervised learning. First, we define five policy relevant payment profiles. Second, a subsample of Italian respondents is labelled through a hybrid process combining manual annotation and LLM classification. Third, supervised models extend

the classification to the full sample. Figure 1 summarises the framework. In the following sections, we further detail the role of LLMs, including their definition and their functioning.

**Figure 1: Research flow: the steps of our methodological framework**



Note: The diagram illustrates the three-step methodological approach adopted in the study. First, five policy-relevant profiles are defined based on theoretical considerations. Second, a subsample of respondents is labeled through a hybrid process combining manual annotation and LLM-assisted classification. Finally, supervised machine learning extends the classification to the full sample, enabling cross-country comparison and policy analysis.

### 3.1 Define profiles of interest

In order to define the different categories of users, we draw inspiration from existing policy-oriented segmentation frameworks, with the aim of constructing a taxonomy that is both interpretable and empirically relevant to the Italian payment landscape.

A simple distinction between cash users, cashless users and mixed users, as adopted for example by Elsinger *et al.* (2025) in the context of digital euro adoption, fails to capture important behavioural heterogeneity. Payment habits reflect not only the payment instruments used, but also the degree of participation in formal financial infrastructures, the adoption of digital technologies, and the ways in which different instruments are combined across payment situations.

Our approach is inspired by the conceptual framework proposed by Henry *et al.* (2023) to discuss the adoption of a CBDC, which highlights that consumers may differ substantially in their relationship with financial infrastructures and payment technologies.<sup>2</sup>

<sup>2</sup>Henry *et al.* (2023) discuss heterogeneous payment needs through a set of consumer cohorts defined around a primary “Well-connected” cohort, representing the typical Canadian consumer with access to cash, bank accounts and digital payment services. Additional cohorts capture con-

---

To better characterise payment behaviour, we consider three complementary dimensions. The first is the degree of integration into formal financial and banking systems. The second is the extent of engagement with digital payment technologies. The third concerns payment orientation, namely the prevailing combination of instruments and channels used in everyday transactions. Although these dimensions are closely related, they do not necessarily coincide. For instance, consumers may be fully integrated into the banking system yet still only make limited use of digital payment technologies. Conversely, cashless users are not necessarily the most innovative consumers, as avoiding cash does not automatically imply intensive use of advanced digital payment solutions.

The first dimension is particularly relevant among consumers with limited participation in formal financial infrastructures. Cash-only users are often associated with being unbanked; however, as emphasised by Falcettoni and Nygaard (2025), not all unbanked individuals are alike. Some people choose to disengage from formal financial systems. Others use partial or alternative infrastructures. These people align with the concept of underserved users in digital payments. (Greene *et al.*, 2024). Accordingly, we distinguish between individuals who rely exclusively on cash and those who, despite having access to basic instruments such as prepaid cards, remain only partially integrated into the banking system.

The distinction between traditional and digitally engaged users is equally important in the Italian context. Unlike countries where digital payments have largely displaced cash, payment behaviour in Italy remains strongly anchored to the use of cash and payment cards. As a result, consumers who are fully integrated into the banking system may nevertheless exhibit markedly different degrees of engagement with digital payment ecosystems and different payment orientations.

The interaction of these dimensions gives rise to five behavioural profiles that capture distinct combinations of financial integration, digital engagement and payment habits:

1. **Cash lovers:** Rely exclusively on cash, due either to a lack of access to financial instruments or to a strong preference for its simplicity and anonymity. They represent the most excluded segment from both the banking system and digital payment technologies.
2. **Bank-tech skeptics:** Use cash and prepaid cards, indicating partial access to financial services, but avoid more advanced banking technologies, including debit and credit cards and mobile payments. Their behaviour reflects limited integration into the banking system and low engagement with digital tools due to unfamiliarity or distrust of digital banking.
3. **Traditionalists:** Combine cash and cards in everyday transactions, but show limited use of online or mobile payments. Despite being fully integrated into the banking system, their behaviour reflects persistent reliance on traditional payment habits rather than active engagement with newer digital solutions.

---

sumers whose payment behaviour differs from this benchmark because of specific preferences or constraints, such as stronger reliance on cash (Cash dependent), reluctance to use technology (Technology Averse), or an early adoption of innovative payment technologies (Early adopters). Their framework is primarily conceptual and intended to inform CBDC policy discussions, but provides useful inspiration for thinking about heterogeneity in payment behaviour.

4. **Well-connected users:** Are fully integrated into the digital economy and actively engage with a wide range of payment instruments, including cash, cards, and digital solutions. They combine different channels flexibly and participate extensively in both online and physical transactions.
5. **Digital-only users:** Fully cashless users who rely exclusively on digital instruments, including mobile wallets and e-payments. They represent a behavioural pattern characterised by the absence of cash usage.

Table 1 summarises the positioning of the five profiles along the three conceptual dimensions. The distinction between Well-connected and Digital-only users highlights that digital engagement and cash usage should not be viewed as equivalent concepts. Both groups are highly integrated into the banking system and actively use digital payment instruments, yet they differ substantially in their payment orientation. While Well-connected users adopt an omnichannel approach and flexibly combine different payment methods, Digital-only users are primarily characterised by the complete avoidance of cash.

**Table 1: Conceptual dimensions underlying the proposed payment profiles**

| Profile              | Banking integration | Digital engagement | Payment orientation |
|----------------------|---------------------|--------------------|---------------------|
| Cash lovers          | Low                 | Low                | Cash-dependent      |
| Bank-tech skeptics   | Low / Partial       | Basic              | Cash + prepaid      |
| Traditionalists      | High                | Moderate           | Cash + Cards        |
| Well-connected users | High                | High               | Omnichannel         |
| Digital-only users   | High                | High               | Cashless            |

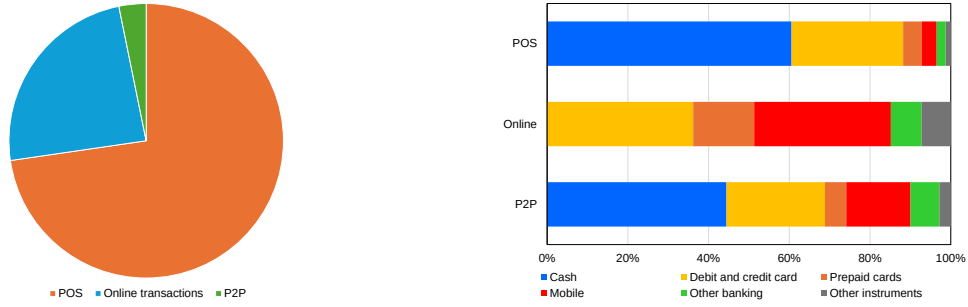
### 3.2 Data

We use data from SPACE 2024, a harmonised survey coordinated by the ECB and conducted across 18 euro area countries. Respondents completed payment diaries and an extensive questionnaire covering payment behaviour, preferences, access to financial services, and attitudes towards payment instruments (Banca d'Italia, 2025; European Central Bank, 2024).

Participants were asked to complete a payment diary alongside an extensive questionnaire on their payment behaviour and attitudes towards cash and cashless instruments. The diary records purchases at the point of sale (POS) and online, recurring payments, as well as peer-to-peer (P2P) transactions. POS, P2P, and online transactions refer to a one-day period. Recurring payments are reported over the previous month. Notably, private P2P payments encompass both physical and online transactions. Additionally, an accompanying questionnaire include sections on payment preferences, access to payment instrument and financial services, cash withdrawal behaviour, and privacy considerations as well as driver for the adoption of new solutions.

The Italian sample includes 4,088 respondents, who reported a total of 8,050 POS transactions, 397 P2P payments, and 2,837 online transactions through their payment diaries.

**Figure 2: Diaries composition.**



(a) One-day diaries distribution

(b) Instruments share per diary

Note: The number of transactions are reported as weighted data and are representative of the final distribution of payment instrument used in the different use cases. Weights are applied to align the sample with population characteristics and mitigate non-response bias. "Mobile" category includes: the use of mobile apps, payment wallet, PayPal, crypto assets and BNPL ; "Other banking" category includes: credit transfers, direct debits and bank cheques; "Other" include gift cards or vouchers and other residuals payment instruments.

We rely on payment diaries using the number of payments made with different instruments, which we categorise into cash, debit and credit cards, prepaid cards, mobile payments, other banking instruments, and other payment methods. Figure 2a shows that POS transactions account for the largest share of recorded payments, followed by online purchases, while P2P payments represent a much smaller fraction. Figure 2b illustrates that payment instrument usage varies substantially across use cases. Although cash remains the dominant payment method at the POS, its role is more limited for online and P2P transactions. Cards (both debit and credit) and mobile payments account for the largest share of online purchases, whereas P2P payments are more evenly distributed across cash, cards, and mobile instruments. Prepaid cards are also used to a non-negligible extent across all contexts. Their popularity in Italy is partly explained by the possibility of obtaining them without opening a bank account. Given this heterogeneity in payment behaviour across use cases and instruments, consumers may exhibit markedly different payment patterns. To capture these differences, Table 2 summarises the expected usage of payment instruments by group across the various contexts, providing a conceptual benchmark for the classification process.

Diaries are essential for assigning labels to consumers, but they capture behaviour only on the survey day and may not fully represent overall payment habits. To address this limitation, we complement diaries with attitudinal data on preferences and ownership of payment instruments. This combination provides a more comprehensive basis for consumer segmentation and allows us to leverage behavioural data for assigning predefined profiles to consumers. The variables of interest used for the labelling and classification process are presented in Table A1 in Appendix A.<sup>3</sup>

<sup>3</sup>For the classification process, we use unweighted data to preserve internal consistency in model training and avoid distortions from survey weights.

**Table 2: Payment instruments used by each group across different use cases**

| Group                       | POS                                      | P2P                                      | Online                                         | Recurring Payments                                    |
|-----------------------------|------------------------------------------|------------------------------------------|------------------------------------------------|-------------------------------------------------------|
| <b>Cash lovers</b>          | Cash only                                | Cash only                                | None                                           | Cash only                                             |
| <b>Bank-tech skeptics</b>   | Cash; prepaid card                       | Cash; prepaid card                       | Prepaid card only; no mobile or e-payments     | Cash; prepaid card                                    |
| <b>Traditionalists</b>      | Cash; debit or credit or prepaid cards   | Cash; debit or credit or prepaid cards   | Limited access; no mobile or e-payments        | Cash; cards; direct debit; credit transfer; no mobile |
| <b>Well-connected users</b> | Cash; cards; mobile or digital wallet    | Cash; cards; mobile or digital wallet    | All instruments including mobile or e-payments | All instruments including mobile or e-payments        |
| <b>Digital-only users</b>   | Cards; mobile or digital wallet; no cash | Cards; mobile or digital wallet; no cash | All instruments including mobile or e-payments | All instruments except cash                           |

Note: Conceptual usage patterns by group across POS, P2P, online purchases, and recurring payments (e.g., rent, utilities).

All these variables of interest used for the labelling and classification process are presented in Table A1 in Appendix A.<sup>4</sup>

In addition, a second set of variables is used ex post for the analysis of group profiles, including sociodemographic characteristics, access to financial services, and attitudes toward new payment solutions. These variables are reported in Table A.2 in Appendix A and are used for descriptive and policy oriented analysis.<sup>5</sup>

### 3.3 Assign a profile to each individual: targeting the training dataset

The initial step involved selecting a sub-sample of 840 Italian respondents and identifying the relevant variables from the SPACE dataset to derive behavioural labels.

#### 3.3.1 Human labelling

We began by manually assigning each user in the subsample to one of the predefined behavioural groups. The classification was carried out through a detailed examina-

<sup>4</sup>For the classification process, we use unweighted data to preserve internal consistency in model training and avoid distortions from survey weights.

<sup>5</sup>Weights are applied to align the sample with population characteristics and mitigate non-response bias.

tion of each individual’s payment records, applying profile-specific criteria derived from the predefined behavioural framework. The assignments were iteratively reviewed to ensure internal consistency across users exhibiting similar patterns in the underlying variables. During this process, implicit decision rules emerged, such as assigning users with no cashless activity to the Cash lovers group, or those avoiding cash entirely to the Digital-only group. Although these rules helped streamline classification, they also revealed ambiguities in borderline cases, where user behaviour did not clearly align with any single profile.<sup>6</sup> This manual approach provided a valuable interpretative starting point, but its scalability and replicability were inherently limited. As Qi *et al.* (2012) point out, manual annotation becomes increasingly impractical in large-scale behavioural studies, where the complexity of user traits and the volume of data can overwhelm annotators and compromise consistency. Moreover, Mehta *et al.* (2021) emphasize that manual classification is often shaped by individual heuristics and subjective interpretations.

In light of these limitations, recent studies have proposed techniques that aim to reduce manual effort without compromising competitive performance. such as active learning approaches (Ilić, 2024), and “uncertainty aware” frameworks like Confident Learning, which estimate and correct label noise in large datasets (Northcutt *et al.*, 2021). Building on these foundations, more recent work advocates hybrid approaches that combine manual expertise with automated tools, or leverage AI assisted annotation to enhance consistency and scalability (Beck, 2023; Klie *et al.*, 2024).

### 3.3.2 LLM-based labelling

The limitations of manual labelling led to the adoption of a more systematic approach that leverages Artificial Intelligence, specifically we employ LLMs: AI algorithms with profound capabilities in understanding and generating natural language. LLMs have been increasingly adopted in marketing and consumer research, supporting tasks such as behavioural prediction, data annotation, and personalized service delivery. For instance, Lee *et al.* (2024) demonstrate how generative AI can be used to develop personalized marketing services by interpreting consumer preferences and automating segmentation strategies. Similarly, Qian *et al.* (2025) provide a comprehensive survey of LLM applications in marketing, highlighting their role in user behaviour learning, content generation, and substitution of human participants.

Interaction with Artificial Intelligence models, such as LLMs, relies on the use of prompts. Prompts are textual inputs that guide the model’s behaviour and may include questions, descriptions, relevant data, and details about the expected output.

Specifically, we used GPT-4.1 (temperature set to 1<sup>7</sup>; OpenAI (2023)) and pro-

---

<sup>6</sup>The rationale behind the choices of human annotators is not directly observable and is unlikely to be fully replicable, as it often relies on implicit rules or heuristics. To approximate this underlying logic, a preliminary comparison was conducted between the subsample’s manual labelling and the classifications produced by the LLM using different prompts. As discussed in the Appendix B, this exercise helped us to understand some of the rules that the human annotator follows to assign labels. For example, they immediately include users in ‘extreme’ groups based on instruments that they do not use. The results of this exercise also informed the selection of the final prompt used for the automated classification with the LLM.

<sup>7</sup>The temperature was treated as a tunable hyperparameter and calibrated through preliminary experiments. In general, temperature values range from 0 to 1 (and can in some implementations exceed 1), controlling the degree of randomness in the model’s output distribution. When the temperature is set to 0, the model behaves in a fully deterministic manner, consistently selecting

---

vided it with a structured prompt. This prompt included (i) the theoretical definitions of the five behavioural groups and (ii) the full set of observed variables for each user, replicating the same information considered during human labeling. The model was instructed to assign each individual to the most appropriate group and to return a short justification for its choice.<sup>8</sup>

To improve robustness, we employed several prompt families. This strategy is conceptually related to the self-consistency approach proposed by ?, according to which agreement across multiple reasoning trajectories can improve reliability and reveal uncertainty in complex classification tasks. Although a single LLM-based label was ultimately retained for the hybrid annotation procedure, the additional prompt formulations provide information on the stability of profile attribution, which is later exploited to construct probabilistic soft labels.

Although the input data are not textual in the traditional sense, the classification task requires semantic interpretation. Diary entries, preferences, and ownership indicators together form a behavioural profile that cannot be reduced to deterministic rules. In this context, the LLM applies contextual reasoning to interpret complex patterns in user behaviour, ensuring consistency and scalability.

Recent work supports the use of LLMs in labeling tasks beyond natural language: for example, Huang and He (2025) demonstrate how LLMs can transform clustering into classification by generating semantic labels, even when working with structured data. A key advantage of our approach lies in the model’s ability to provide justifications for each assignment, enhancing transparency and supporting validation. This interpretability is particularly valuable in our behavioural research, where classification criteria are context dependent and often difficult to formalize. In addition to this qualitative benefit, the LLM-based method offers substantial efficiency gains: processing the subsample of 840 individuals required approximately one hour, a significant improvement over manual annotation.

### 3.3.3 Compare the Human and the LLM labels

The initial agreement between human and LLM labels was 68% (570 cases). As shown in Table 3, Traditionalists account for the largest share of matched observations, Cash lovers exhibit the highest agreement rate. By contrast, Well-connected users show the lowest consistency, indicating that intermediate profiles are more difficult to classify than more extreme behavioural types.

For the mismatches, detailed results are reported in Table 4. The confusion matrix shows that the most frequent disagreements occurred between Traditionalists

---

the most probable class. Although this increases reproducibility, in our setting it led to overly rigid assignments and a systematic concentration of observations within a few dominant behavioural groups, reducing the model’s ability to capture heterogeneity across users. Lower values produced more deterministic classifications but also increased the tendency to concentrate assignments within a limited number of dominant classes. Given the gradual boundaries between some behavioural profiles and the multidimensional nature of the classification task, we selected a temperature of 1 to improve sensitivity to nuanced behavioural patterns and promote a more balanced use of classes. At this level, the model allows for greater exploration of alternative plausible classifications, which helps in capturing subtle differences in user behaviour that would otherwise be suppressed under more deterministic settings. This choice is also consistent with recent evidence suggesting that, in some GPT-4 classification tasks, non-zero temperature settings may improve performance by enabling broader exploration of possible assignments (Li *et al.*, 2025; Sumanathilaka *et al.*, 2025).

<sup>8</sup>A detailed discussion of the prompt design process is provided in the Appendix B.

**Table 3: Group distribution among matched cases**

| Group                | Label | Count | Match with Human Label | Match with LLM Label |
|----------------------|-------|-------|------------------------|----------------------|
| Traditionalists      | 3     | 246   | 82.6%                  | 63.2%                |
| Digital-only users   | 5     | 111   | 63.4%                  | 88.1%                |
| Well-connected users | 4     | 99    | 49.0%                  | 58.9%                |
| Bank-tech skeptics   | 2     | 70    | 72.9%                  | 64.2%                |
| Cash lovers          | 1     | 43    | 63.2%                  | 91.5%                |

Note: The table reports the distribution of user groups among cases where the initial LLM classification matched the human-assigned label. Counts refer to the number of observations in each group. Percentages are computed at the group (row) level and measure agreement across classifiers from two complementary perspectives. Specifically, for each group, “Match with Human Label” indicates the share of observations assigned to that group by the LLM that are confirmed by the human label, “Match with LLM Label” indicates the share of observations assigned to that group by the human classifier that are confirmed by the LLM.

and Well-connected users in both directions: the LLM classified 75 individuals as Traditionalists who were labelled as Well-connected by the human annotator, 34 individuals labelled as Traditionalists by the human were classified as Well-connected by the LLM. This bidirectional confusion highlights the difficulty of distinguishing between these adjacent profiles, which share several behavioural characteristics, particularly the use of cash and cards, but differ in their adoption of mobile and other digital payment methods.

Other notable patterns include cases where the LLM classified as Traditionalists individuals labelled by the human annotator as Cash lovers (24 cases) or Digital-only users (29 cases), as well as cases where Digital-only users were classified as Well-connected.

**Table 4: Confusion matrix of mismatches between LLM and human labels**

| LLM vs Human       | Cash Lovers | Bank-tech skeptics | Traditionalists | Well-connected | Digital-only |
|--------------------|-------------|--------------------|-----------------|----------------|--------------|
| Cash lovers        | 0           | 0                  | 2               | 2              | 0            |
| Bank-tech skeptics | 0           | 0                  | 16              | 18             | 5            |
| Traditionalists    | 24          | 15                 | 0               | 75             | 29           |
| Well-connected     | 1           | 4                  | 34              | 0              | 30           |
| Digital-only       | 0           | 7                  | 0               | 8              | 0            |

Note: Rows represent LLM predictions; columns represent human-assigned labels. Values indicate the number of mismatched cases for each pair of groups.

---

**Managing mismatch: LLM revision considering human labels.** We resubmitted the 270 mismatched cases to the model using a revised prompt<sup>9</sup> that introduced the perspective of a “human expert” offering an alternative classification. This adjustment simulated a dialogical evaluation, encouraging the model to reconsider its initial reasoning in light of a competing interpretation.

Following this second round, the model confirmed approximately 35% of its original decisions (95 out of 270 cases), while adopting the human assigned label in the remaining cases.

Among the retained classifications, confirmations were highly concentrated in the Traditionalists group: 66 out of 95 confirmed cases belonged to this category, and almost all had originally been labelled as Well-connected by the human annotator. The accompanying explanations point to a consistent rationale: these users relied mainly on cash and traditional banking instruments, occasionally using cards for POS transactions, but showed little or no evidence of mobile, online, or advanced digital payments. This behaviour is more consistent with the Traditionalists profile than with the Well-connected profile, which implies broad use of digital and mobile payment instruments. The results therefore show that the revised prompt helped the model articulate and defend its classification decisions more explicitly.

A similar pattern emerges for Bank-tech skeptics, frequently reassigned from the Well-connected group (12 cases). In these cases, the explanations emphasised heavy reliance on cash and prepaid instruments and limited use of mobile or online channels, characteristics that are difficult to reconcile with a genuinely multichannel profile. By contrast, Cash Lovers and Well-connected users were rarely confirmed, suggesting that the LLM applies relatively strict criteria to these categories, particularly regarding the exclusivity of cash usage.<sup>10</sup>

### 3.3.4 Final labels for the sub-sample

Table 5 reports the distribution of user groups across the three labeling stages (LLM, Human, and Final). Traditionalists remain the dominant group (42.8%). Extreme profiles such as Cash lovers and Digital-only users are more represented than in the initial set. This balanced yet differentiated composition is expected to enhance the robustness of subsequent modeling steps.

The iterative refinement process not only enhanced the consistency between human and model classifications, but also provided empirical evidence of the model’s contextual adaptability and capacity for self-correction when confronted with new information. This property is critical for establishing confidence in its practical application, as it demonstrates that the LLM does not rigidly adhere to initial outputs but can revise decisions in light of additional evidence. Furthermore, the inclusion of label justifications (both in the initial and revised rounds) was crucial in validating the underlying reasoning and in identifying edge cases characterized by ambiguous behavioural profiles.

---

<sup>9</sup>In full: *You previously classified users into one of five behavioural groups based on their payment habits. Now, a human expert has reviewed your classifications and provided their own. Your task is to compare your original classification with the human-assigned one and decide whether to confirm your original classification or adopt the human one, providing a brief motivation.*

<sup>10</sup>Building on this observation, we further examined cases where the model retained its original classification. Tables B5 and B6 in the Appendix provide selected examples together with the corresponding motivations.

---

**Table 5: Comparison of group distribution**

| Group                | LLM   |        | Human |        | Final |        |
|----------------------|-------|--------|-------|--------|-------|--------|
|                      | Count | %      | Count | %      | Count | %      |
| Traditionalists      | 389   | 46.3%  | 298   | 35.51% | 359   | 42.78% |
| Digital-only users   | 126   | 15%    | 175   | 20.85% | 182   | 21.69% |
| Well-connected users | 168   | 20%    | 202   | 24.07% | 123   | 14.66% |
| Bank-tech skeptics   | 109   | 12.99% | 96    | 11.44% | 103   | 12.27% |
| Cash lovers          | 47    | 5.71%  | 68    | 8.13%  | 72    | 8.58%  |

Note: The table compares the distribution of user groups across three stages of the labeling process: initial LLM classification, human-assigned labels, and the final reconciled set after review. Counts refer to the number of observations in each group; percentages indicate the share of each group within the respective labelling stage.

Our approach aligns with recent work on hybrid annotation strategies that combine automated labeling with expert review. The Multi-LLM Consensus with Human Review framework proposed by Huang and He (2025) demonstrates how structured interaction between multiple LLMs and human annotators can enhance both accuracy and interpretability significantly reducing the need for manual annotation. This resonates with the human-in-the-loop methodology advocated by Artemova *et al.* (2025), reinforcing the value of iterative and expert guided refinement in complex behavioural classification tasks.

### 3.4 Extension to the full sample: supervised classification

As a final step, we employ supervised learning to extend the classification to the full sample. Recent work shows that LLM-generated labels can effectively train supervised classifiers, achieving performance comparable to models trained on human annotations (Pangakis *et al.*, 2024). In our context, this approach enables the classification of all 4,088 Italian respondents. In our context, this combination enables the classification of all 4,088 Italian respondents without incurring the prohibitive costs of full manual or LLM based annotation but maintaining consistency with the theoretical segmentation framework.

Instead of relying on a single model, we adopt a multi-algorithm strategy to capture the diverse inductive biases and learning mechanisms inherent in different methods. Specifically, four algorithms are considered: Support Vector Machines (SVM), Random Forest, Multilayer Perceptron (a simple neural network with a linear structure), and XGBoost. This practice is well established in supervised learning, where ensemble approaches enhance robustness, generalisation, and interpretability by leveraging model diversity (Breiman, 1996; Dietterich, 2000).

By comparing the performance of these algorithms, our goal was to identify the most suitable technique for extending the classification to the full Italian dataset. To this end, we rely on a set of performance indicators that evaluate both accuracy and the quality of predictions for each behavioural group. We consider four indicators. Accuracy measures the share of correct predictions produced by each algorithm. For each consumer group, we also report Recall (Sensitivity), which captures the proportion of correctly identified positive cases; Precision, which measures the proportion of predicted positive cases that are correctly classified; and the F1 score, a synthetic

index combining precision and recall.

To enhance model performance and ensure fair representation across behavioural groups, we addressed the issue of class imbalance in the training data using the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla *et al.*, 2002). This method generates synthetic examples for underrepresented classes, helping the model learn more balanced decision boundaries and reducing bias toward majority groups. In parallel, we adopted a StratifiedKFold cross-validation strategy with 5 folds to optimize hyperparameter selection for each algorithm. This ensures that the distribution of behavioural groups is preserved across folds, allowing for a more reliable evaluation of model performance and reducing the risk of overfitting.

We reserved 20% of the labeled dataset (168 out of 840 examples) as an independent test set to assess model performance using multiple metrics: accuracy, F1 score, sensitivity and specificity .

The comparative evaluation of the model performance in Table 6 confirms XGBoost as the best performing algorithm with 89% accuracy, although the result is satisfactory for all four models. Random Forest ranks second with 85.7%, and the Neural Network shows a very similar performance, whereas SVM is slightly lower.

**Table 6: Accuracy of the supervised learning algorithms**

| Model          | Accuracy     |
|----------------|--------------|
| SVM            | 0.851        |
| Random Forest  | 0.857        |
| Neural Network | 0.856        |
| XGBoost        | <b>0.890</b> |

Note: Accuracy values are computed on the independent test set (20% of labeled data). Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome.

In Table 7 there are the results for each consumer group across the four supervised models beyond overall accuracy. XGBoost demonstrates strong performance across all behavioural groups, achieving high F1 scores with balanced sensitivity and specificity. It is particularly effective for Digital-only users ( $F1 = 0.950$ ) and maintains good sensitivity for Well-connected users, which is generally the most challenging group to classify.

Random Forest also performs robustly, especially for Cash lovers, where it reaches perfect specificity and high sensitivity ( $F1 = 0.937$ ). However, its performance is less consistent for other groups, with lower sensitivity for Bank-tech skeptics for example. SVM and the Neural Network show greater variability. Both record relatively lower sensitivity for Bank-tech skeptics, and the Neural Network achieves a lower F1 score for Well-connected users (0.808) compared to other groups.

Well-connected users emerge as the most difficult group to classify across all models, likely due to overlapping behavioural signals that make decision boundaries less distinct.

The comparison also highlights a degree of “complementarity” among models: when one algorithm performs less effectively on a specific group, another often compensates. For example, the Neural Network shows lower performance for Well-connected users, Random Forest and XGBoost achieve higher sensitivity for this group. This complementarity motivated the adoption of a multi-model strategy for

the final classification step.

**Table 7: Classification metrics by group across models**

| (a) F1 score         |       |       |       |       |
|----------------------|-------|-------|-------|-------|
| Group                | SVM   | RF    | NN    | XGB   |
| Cash lovers          | 0.848 | 0.967 | 0.823 | 0.903 |
| Bank-tech skeptics   | 0.807 | 0.864 | 0.840 | 0.884 |
| Traditionalists      | 0.833 | 0.844 | 0.845 | 0.868 |
| Well-connected users | 0.816 | 0.777 | 0.808 | 0.823 |
| Digital-only users   | 0.926 | 0.891 | 0.926 | 0.950 |

| (b) Sensitivity (Recall) |       |       |       |              |
|--------------------------|-------|-------|-------|--------------|
| Group                    | SVM   | RF    | NN    | XGB          |
| Cash lovers              | 0.875 | 0.937 | 0.875 | 0.875        |
| Bank-tech skeptics       | 0.724 | 0.758 | 0.724 | 0.793        |
| Traditionalists          | 0.833 | 0.816 | 0.866 | 0.883        |
| Well-connected users     | 0.869 | 0.913 | 0.826 | 0.913        |
| Digital-only users       | 0.950 | 0.925 | 0.950 | <b>0.950</b> |

| (c) Specificity      |       |       |       |       |
|----------------------|-------|-------|-------|-------|
| Group                | SVM   | RF    | NN    | XGB   |
| Cash lovers          | 0.980 | 1.000 | 0.973 | 0.993 |
| Bank-tech skeptics   | 0.985 | 0.992 | 1.000 | 1.000 |
| Traditionalists      | 0.907 | 0.935 | 0.898 | 0.916 |
| Well-connected users | 0.958 | 0.931 | 0.965 | 0.951 |
| Digital-only users   | 0.968 | 0.953 | 0.968 | 0.984 |

Note: The table reports three performance metrics—F1 score, Sensitivity (Recall), and Specificity—for each consumer group across four supervised algorithms: SVM, Random Forest (RF), Neural Network (NN), and XGBoost (XGB). F1 score is the harmonic mean of precision and recall; Sensitivity measures the proportion of correctly identified positive cases; Specificity measures the proportion of correctly identified negative cases.

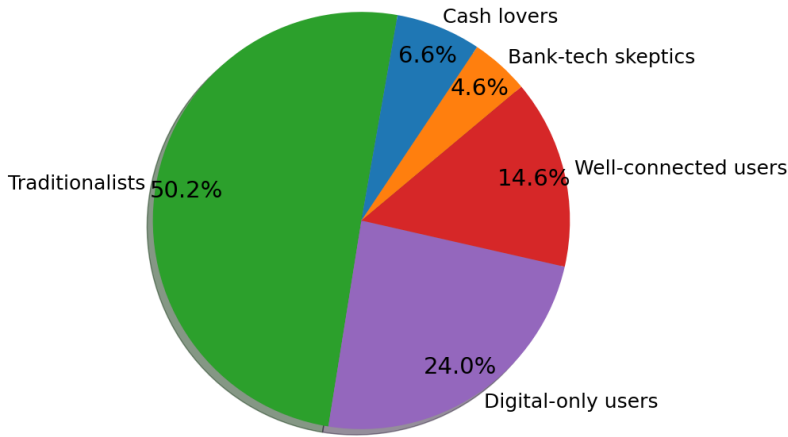
To extend the classification to the entire Italian dataset (4,088 respondents), each of the four trained models (SVM, Random Forest, Neural Network, and XGBoost) was used to independently predict the behavioural group for each individual using its optimal configuration identified in the previous step. Final group labels were assigned through a weighted majority voting scheme, where each model’s contribution was proportional to its classification accuracy on the test set. This approach ensures that models with higher predictive reliability exert greater influence while preserving the diversity of perspectives offered by the ensemble, which improves robustness and generalizability.

## 4 Understanding payment profiles in Italy

Following the supervised classification applied to the full Italian sample (4,088 respondents), we examine how the five predefined user profiles are distributed across

the population. Analyzing both their prevalence and the behavioural patterns that shape these groups provides a unique perspective on payment habits in Italy. Traditionalists represent the largest share (50.2%), followed by Digital-only users (24.0%) and Well-connected users (14.6%). Cash lovers account for 6.6%, Bank-tech skeptics are the smallest group at 4.6%. This distribution confirms that payment habits in Italy remain largely anchored to traditional instruments, despite the growing availability of digital options. The limited presence of groups excluded by digital banking options, such as Cash lovers and Bank-tech skeptics, suggests that financial vulnerability is relatively contained. At the same time, the share of consumers embracing digital innovation is far from negligible, indicating a gradual yet tangible shift toward more diversified and technologically advanced payment practices.

**Figure 3: Distribution of user profiles in Italy**



Note: Group distribution considers weighted data and is representative of the Italian population.

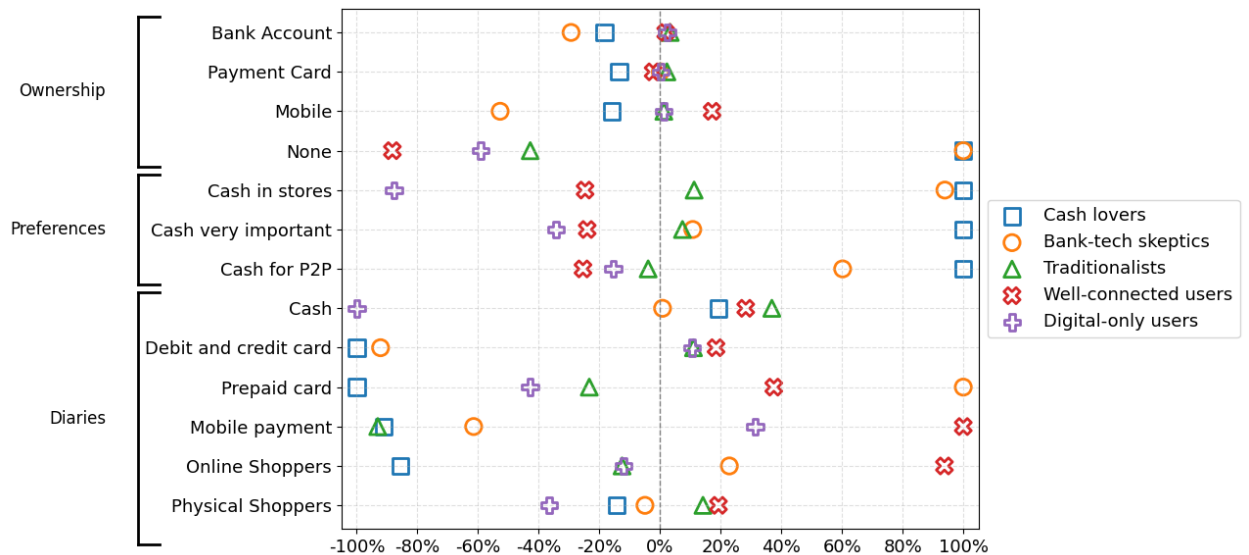
#### 4.1 Behavioural dimensions: indicators underpinning group classification

As discussed in Section 3, the classification of consumers into five profiles relies on a set of behavioural and attitudinal indicators derived from payment diaries, preferences, and access to payment instruments. These indicators differentiate the groups *a priori* in defining the five groups, but examining them *a posteriori* provides additional insight into which variables are most indicative of each profile.

Figure 4 illustrates how the main indicators of payment behaviour deviate from the population average across the five profiles. For Traditionalists, as well as Well-connected and Digital-only users, access to payment instruments is largely in line with the average. Only Well-connected users show slightly higher access to mobile payment options, while access to bank accounts and payment cards remains essentially similar across these three groups. This confirms that differences among them are not driven by potential access to instruments that could enable cash-free or

innovative payment behaviour, but rather by personal attitudes expressed through preferences and actual usage patterns.

**Figure 4: Deviations from the national average for key indicators of payment behaviours by groups**



Note: The figure reports the percentage deviation of each user group from the national average across three dimensions: (i) ownership of payment instruments (bank account, cards, mobile-enabled devices), (ii) attitudinal preferences, and (iii) behavioural indicators from payment diaries (number of transactions by instrument and context). Positive values indicate above-average incidence, while negative values indicate below-average incidence. This representation highlights how classification indicators effectively differentiate the five profiles. For readability, deviations are capped at  $\pm 100\%$  in the figure.

In contrast, Cash lovers and Bank-tech skeptics exhibit markedly lower-than-average access to bank accounts and mobile instruments, with Cash lovers also having fewer cards than other groups. Bank-tech skeptics score even lower on bank account ownership, yet remain aligned with other groups in card ownership due to the widespread availability of prepaid cards, which do not necessarily require a bank account. This pattern shows that the Bank-tech skeptic profile reflects a deliberate choice to avoid banking services, whereas Cash lovers are likely constrained by other factors. For these two groups, access conditions payment choices in addition to preferences.

Preferences reveal a clear divide between groups at the extremes: Cash lovers rate cash as “very important”, whereas Well-connected and Digital-only users express strong preference for digital solutions. Traditionalists, although closer to the population average, still show a tendency toward cash, reinforcing their characterization as conservative users.

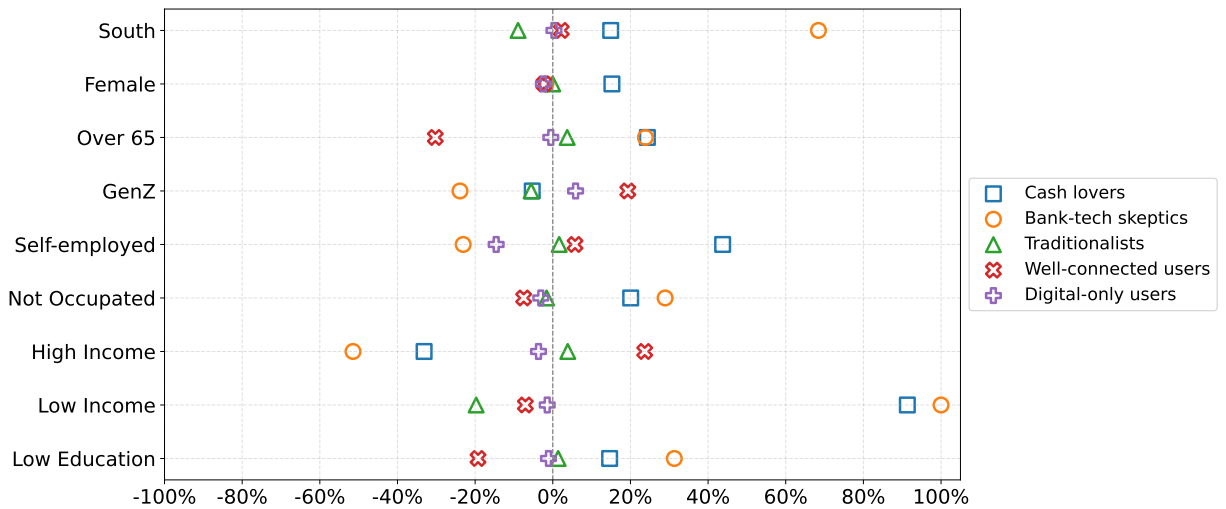
Moving to diary data, both the choice of instruments and the share of individuals making physical versus online purchases display significant dispersion. Digital-only users completely avoid cash, yet their use of other instruments remains below that of Well-connected users, who appear more technologically engaged. This indicates that Digital-only users are cash-free but not necessarily “tech-savvy.” Well-connected users, by contrast, employ all available instruments and demonstrate balanced en-

agement in both physical and online transactions. Cash lovers rely almost exclusively on cash, with negligible card or mobile usage and a very low share of online purchases. Bank-tech skeptics do engage in online shopping but primarily through prepaid cards, combining cash reliance at the POS with avoidance of advanced banking tools.

## 4.2 Socio-demographic characteristics

Socio-demographic factors provide additional insights into the composition of the five profiles and help contextualize payment behaviour within broader structural characteristics. Figure 5 reports deviations from the population average for key attributes, including education, income, age, gender, geographical location, and employment status.<sup>11</sup>

**Figure 5: Deviations from the national average for socio-demographic characteristics by groups**



Note: The figure shows percentage deviations from the national average for key socio-demographic attributes. Deviations are calculated on the share of each attribute within each user group compared to its share in the overall population. Positive values indicate that the attribute is more prevalent within the group than in the population average. Negative values indicate lower prevalence. For readability, deviations are capped at  $\pm 100\%$  in the figure.

The geographical divide and gender gap, widely investigated in Di Iorio and Rocco (2025), persist but become more intuitive when viewed through the lens of user profiles. Bank-tech skeptics show a markedly higher-than-average share of residents in Southern regions, while Cash lovers are close to the average for geography but stand out as the only group with a higher share of women.

The generational divide is even more pronounced: Well-connected and Digital-only users include a higher proportion of Generation Z, suggesting that the transition to digital payments is easier for digital natives, whereas older cohorts (over 65) are concentrated among Cash lovers and Bank-tech skeptics. Traditionalists occupy an intermediate position but lean toward older age groups.

<sup>11</sup>The detail for the variables are presented in Table A4 in Appendix A.

---

Employment status further confirms the vulnerability of Cash lovers and Bank-tech skeptics, who exhibit a higher share of individuals not formally employed. This likely indicates that their payment choices are influenced by broader socio-economic constraints. Interestingly, Cash lovers also include a relatively large share of self-employed individuals, possibly reflecting continued reliance on cash for business transactions.

Income and education, as documented in previous literature, remain key factors in shaping these profiles. High-income individuals are significantly underrepresented among Cash lovers and Bank-tech Skeptics, while they are more common among Well-connected users. Conversely, low-income individuals tend to be more frequent in the first two groups, although Traditionalists appear closer to middle-income segments. Low levels of education are also more concentrated among Cash lovers and Bank-tech Skeptics, indicating that limited digital skills or lower familiarity with technology may contribute to the behavioural patterns observed.

Traditionalists, the largest group, exhibit patterns aligned with the general population, confirming their role as a baseline segment in the Italian payment landscape. Cash lovers and Bank-tech skeptics display more distinctive socio-demographic traits, reinforcing their interpretation as economically and digitally vulnerable. Conversely, Well-connected tend to be younger, more educated, with a higher incidence of high income. These findings confirm that payment behaviour is not only a matter of preferences and access but also reflects underlying economic and generational divides, which have implications for financial inclusion and the design of innovative payment solutions especially if aimed at involving specific vulnerable groups.

### 4.3 Financial profile

Financial characteristics provide a deeper understanding of the structural differences among the five profiles and their potential implications for financial inclusion. Figure 6 reports deviations from the population average for key indicators, including access to banking and non-banking financial products, investment behaviour, and income modalities.<sup>12</sup>

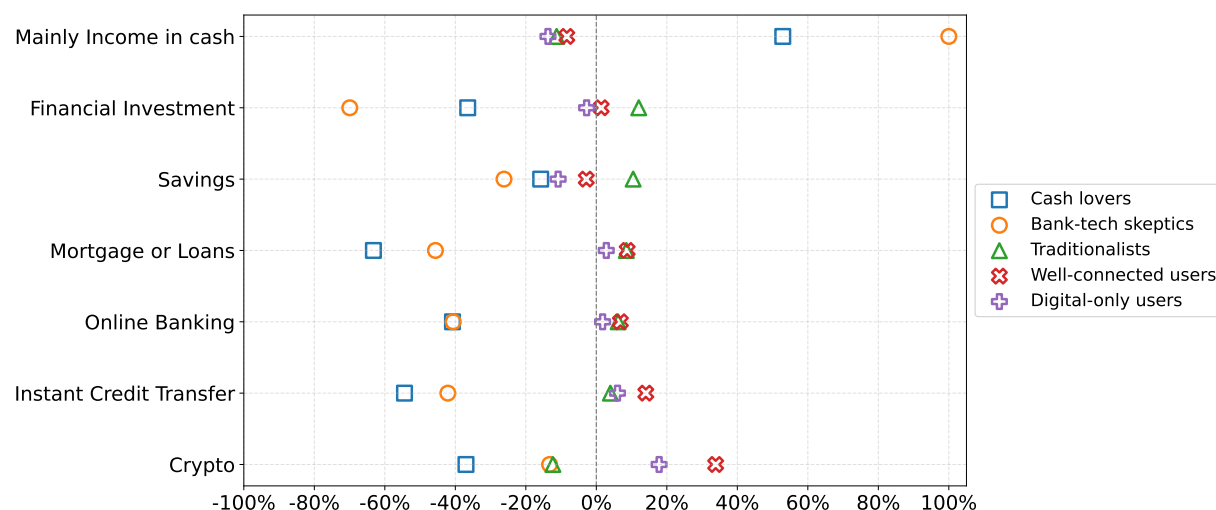
Cash lovers and Bank-tech skeptics stand out as the groups with limited financial inclusion. Both exhibit a markedly higher share of individuals still receiving income primarily in cash and show minimal engagement with advanced financial products. Access to savings accounts, loans, and investment instruments is significantly below the population average, and interest in non-banking innovations such as crypto-assets or in banking innovations such as instant payments is almost absent. These patterns reinforce the interpretation of these groups as structurally excluded from the broader financial ecosystem.

Traditionalists display a profile highly compatible with banking users: they have access to all major banking products, including savings accounts and loans, maintain financial investments, and show openness to banking innovations such as online banking and instant payments. However, they exhibit limited engagement with non-banking innovations as crypto-assets. Therefore, their financial behaviour remains anchored to a logic where financial services are fundamentally linked to traditional banking offerings. Notably, Traditionalists are the only group above the population average for both savings and investments, confirming their conservative and

---

<sup>12</sup>The detail for the variables are presented in Table A5 in Appendix A.

**Figure 6: Deviations from the national average for the financial profile by groups**



Note: The figure shows percentage deviations from the national average for financial characteristics within each user group. Indicators include access to banking products (savings accounts, loans), engagement with non-banking innovations (crypto-assets, instant payments), and income modalities (e.g., prevalence of cash-based income). Positive values indicate that the attribute is more common within the group compared to the total population; negative values indicate lower prevalence. For readability, deviations are capped at  $\pm 100\%$  in the figure.

financially stable profile.

Well-connected and Digital-only users, by contrast, fall below the average for savings but show higher exposure to debt and maintain investment levels close to the population mean. Both groups demonstrate strong engagement with technological innovations, including crypto-assets and instant payments. This pattern shows that technological adoption does not necessarily coincide with a conservative financial profile and may entail some risk exposure. Further research could explore these dynamics in relation to emerging instruments such as BNPL, which often raise concerns for their hybrid nature between payment facilitation and increased indebtedness.

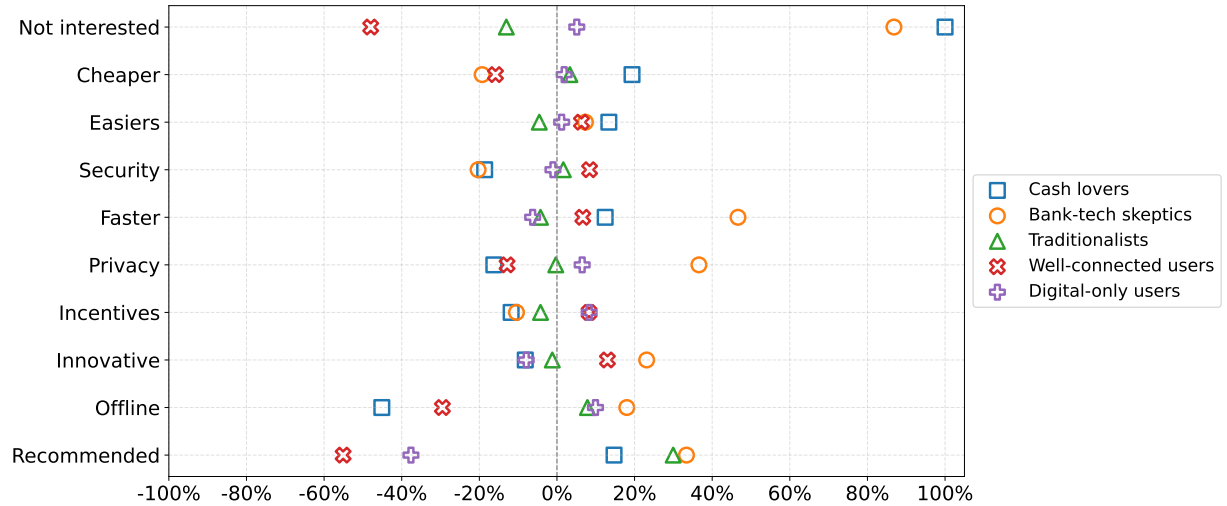
Financial profiles strongly mirror payment behaviours, confirming the principle that “financial inclusion starts with payments” (Bank for International Settlements, 2020). These findings highlight the dual dimension of inclusion: access to banking services and openness to innovation. They emphasise the need for targeted measures to help vulnerable groups and to design solutions that cater for both conservative and digitally savvy users. At the same time, they ensure responsible engagement with technology-driven financial products.

## 4.4 Openness to innovation

Interest in new payment solutions varies significantly across the five profiles, reflecting both structural and behavioural factors. Figure 7 reports deviations from the population average for key drivers of interest, including ease of use, cost, security, privacy, innovative features, speed, incentives, recommendations, and offline functionality, as well as the share of respondents declaring no interest. Factors are ranked

by their total importance in the population (e.g., lower cost is the most cited driver, recommendations are the least) as reported in Banca d'Italia (2025).<sup>13</sup>

**Figure 7: Deviations from the national average in interest in new payment solutions by groups**



Note: The figure shows percentage deviations from the national average for the main drivers of interest in adopting new payment solutions. Deviations are calculated on the share of respondents within each group who indicate a given factor as relevant, compared to its prevalence in the overall population. Factors are ranked by their global importance in the population (e.g., lower cost is the most cited driver, recommendations are the least). Positive values indicate that the factor is more influential for the group than for the average user; negative values indicate lower relevance. For readability, deviations are capped at  $\pm 100\%$  in the figure.

The first striking result is the clear divide in openness. Cash lovers and Bank-tech skeptics exhibit the strongest resistance to innovation, with a markedly higher share of individuals reporting no interest in new payment solutions. When these groups express interest, their priorities differ from those of more digitally engaged users: Cash lovers emphasize simplicity and low cost, Bank-tech skeptics value privacy and offline functionality. These preferences highlight the need for targeted strategies focused on usability and affordability to overcome the barriers faced by these groups.

Interestingly, even among digital-only users, the proportion of respondents who are not interested in new solutions is higher than average, possibly because they are satisfied with the existing options that allow them to avoid using cash. When they are interested, they prioritise offline functionality, privacy and incentives to some extent, indicating a pragmatic approach to innovation rather than enthusiasm for novelty.

Traditionalists display moderate openness to innovation. They do not reject new solutions outright, but their interest is primarily driven by recommendations, indicating that social influence plays a key role in adoption, claiming for a network effect. Offline functionality also emerges as a relevant factor, consistent with their preference for familiar and reliable payment environments.

<sup>13</sup>The detail for the variables are presented in Table A6 in Appendix A.

---

Conversely, Well-connected users demonstrate the highest appetite for innovation. They score above average on most drivers, particularly innovative features, speed, and security, confirming their role as early adopters.

These patterns highlight the heterogeneity of consumer preferences and the importance of understanding how different user profiles respond to new payment solutions. This suggests to policymakers and market participants that, although innovation is generally developed with broad objectives in mind, effective adoption requires recognition of different unmet user needs. For example, tech-savvy users look at advanced features, while simplicity, trust, and cost-effectiveness are important for groups of population characterized by a higher persistence in payment habits.

## 5 Beyond the hard classification: a soft-label perspective

The methodological framework discussed in the previous sections assigns each individual to a single behavioural profile to obtain clearly interpretable groups and facilitate the characterization of payment behaviours and their policy implications. Although this hard classification approach provides a useful representation of the population, it inevitably forces each individual into a single category. The final labels were obtained through a robust annotation procedure combining human evaluation and LLM-based classification; nevertheless, alternative annotation sources did not always assign the same profile to a given individual. As a result, information contained in the degree of agreement across alternative classifications is necessarily lost once a unique label is assigned to each observation. More generally, some consumers may simultaneously exhibit characteristics associated with multiple behavioural profiles, making a purely discrete representation of payment behaviour potentially restrictive.

To further investigate this aspect, we complement the hard classification with a soft-label perspective that captures the degree of proximity of each individual to the alternative behavioural profiles. This exercise is not intended to replace the original classification, but rather to assess its robustness and provide additional insights into behavioural overlap across profiles. Such an approach is also consistent with the recent literature on soft labels in supervised learning. As discussed by Singh *et al.* (2025), collapsing annotation distributions into a single hard label may induce classifiers to overestimate prediction certainty in genuinely ambiguous cases, whereas retaining probabilistic information allows uncertainty to be preserved and avoids imposing rigid boundaries between behavioural profiles.

### 5.1 Construction of soft labels

As discussed in Section 3, the use of multiple prompt families inspired by the self-consistency principle of Wang *et al.* (2022) generates several alternative classifications for each observation. These multiple annotations naturally induce a probabilistic representation of profile membership. To retain this information, we construct soft labels by aggregating the alternative classification outputs generated during the annotation process. Specifically, for each individual ( $i$ ), the probability of belonging to profile ( $j$ ) is computed as

---


$$p_{ij} = \sum_s w_s I_{ij}^{(s)},$$

where  $I_{ij}^{(s)}$  is an indicator equal to one when classification source (s) assigns individual (i) to profile (j), and zero otherwise, ( $w_s$ ) denotes the weight assigned to each classification source, with  $\sum_s w_s = 1$ .

During the construction of the training set, labels were assigned by comparing human annotations with LLM based classifications. Although the hybrid annotation procedure ultimately retained a single LLM evaluation for each observation, the alternative prompting strategies described in Appendix B provide additional information on the stability of profile attribution across different classification settings. The weighting scheme is designed to remain consistent with the hard-label assignment procedure adopted throughout the paper. One third of the total weight is assigned to the human annotation and one third to the benchmark LLM classifier (SynFun). The remaining third is attributed to the alternative prompt-based classifications, which provide supplementary information on classification uncertainty. Within this component, the Hard prompt receives a weight of one half, the Syn, NarrTech, and NarrSoc variants each receive a weight of one sixth. Higher probabilities therefore indicate stronger agreement across classification sources, whereas lower probabilities reflect greater ambiguity in profile attribution.

Once soft labels were defined on the training set, the probabilistic classification framework was extended to the prediction stage through a supervised neural network classifier. Specifically, we trained a multilayer perceptron (MLP) using the behavioural variables described in the previous sections as input features and the soft label probability vectors as target outputs. Rather than optimizing against deterministic one hot labels, the model was trained using a soft cross-entropy loss function, allowing predicted probabilities to be directly compared with the target distributions.

The resulting model produces, for each individual, a probability distribution across the five groups rather than a single deterministic assignment. These probabilities can be interpreted as degrees of proximity to the different profiles and could provide additional information on cases characterized by overlapping or borderline payment behaviours.

## 5.2 Classification confidence and behavioural overlap

The soft-label framework associates each individual with a probability distribution across the five behavioural profiles rather than with a single deterministic assignment. Interpreting these distributions directly is not always straightforward. For this reason, we rely on two complementary representations.

The heatmap in Figure 8 reports the full distribution of membership probabilities across profiles, while Table 8 provides a simplified representation based on alternative assignment thresholds. Together, these two tools allow us to assess both the stability of profile attribution and the extent of behavioural overlap across groups.

Figure 8: Heatmap of soft-label memberships across groups

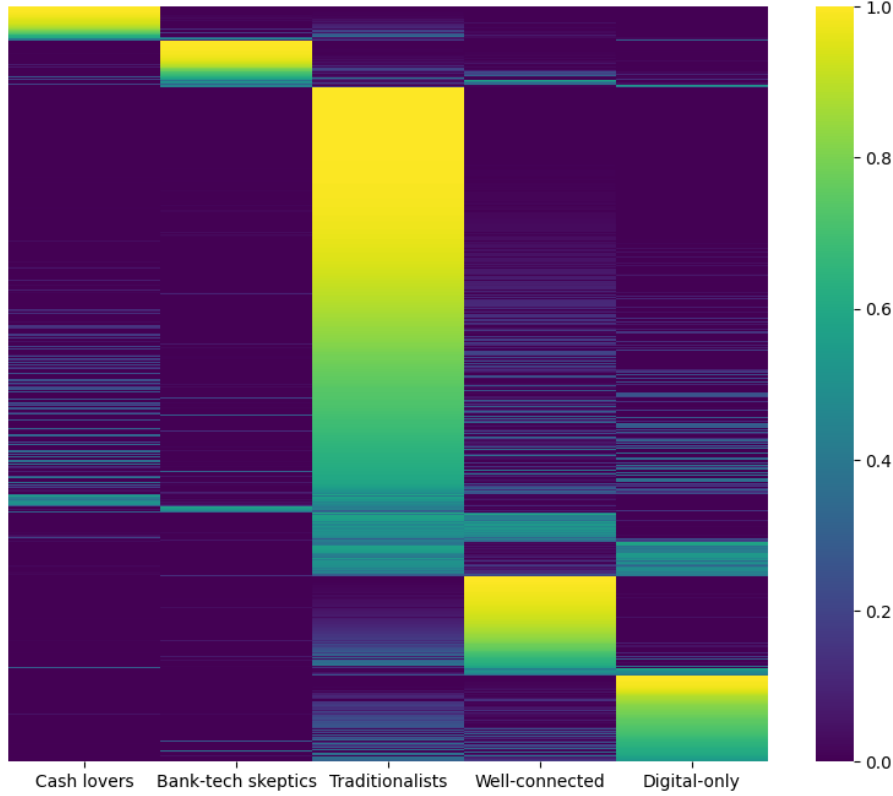


Table 8: Distribution of profile assignments under alternative soft-label thresholds

| Behavioural profile  | $p \geq 0.8$ | $p \geq 0.6$ | $p > 0.5$ | Softmax assignment |
|----------------------|--------------|--------------|-----------|--------------------|
| Cash lovers          | 3.13         | 4.04         | 4.70      | 5.09               |
| Bank-tech skeptics   | 3.74         | 5.01         | 5.75      | 6.19               |
| Traditionalists      | 35.05        | 51.69        | 57.85     | 59.86              |
| Well-connected users | 8.73         | 11.89        | 13.72     | 14.46              |
| Digital-only users   | 4.60         | 10.37        | 13.16     | 14.41              |
| Overlapping profiles | 44.74        | 17.00        | 4.82      | —                  |

Note: The table reports the percentage of individuals assigned to each behavioural profile under alternative probability thresholds based on soft-label memberships. Columns labelled  $p \geq x$  report the share of individuals whose membership probability for a given profile exceeds the specified threshold. “Soft-max assignment” corresponds to assigning each individual to the profile associated with the highest membership probability.

The results show that the classification framework is very stable, even when uncertainty is taken into account. If a membership probability above 0.5 is interpreted as evidence of a prevailing behavioural profile, fewer than 5% of individuals remain unassigned. Consistent with this result, the heatmap reveals that most observations are characterised by a clearly dominant membership, while substantial overlap is limited to a relatively small subset of respondents. The soft-label analysis therefore

supports the validity of the hard classification providing additional information on the degree of proximity between profiles.

Traditionalists emerge as the dominant profile across all thresholds. However, the substantial increase in prevalence under more permissive assignment rules highlights that this group acts as a behavioural centre for individuals displaying mixed or intermediate payment behaviours. In fact, more than one third of individuals are associated with this group with a probability above 0.8, whereas almost 60% are classified as Traditionalists under the soft-max assignment rule. By contrast, Cash lovers and Bank-tech skeptics remain comparatively small groups even under relaxed probability thresholds, but display relatively concentrated memberships. A different pattern emerges for the share of Digital-only users which increases substantially as the probability threshold is relaxed, rising from 4.6% at the 0.8 threshold to more than 14% under soft-max assignment. This suggests a weaker separation profile in contrast with the fully cashless behaviour identified in the characterisation of groups. Similarly, Well-connected users also exhibit increasing prevalence under lower thresholds.

The prevalence patterns emerging from the soft-label framework should not be interpreted as direct alternatives to the hard classification discussed in the previous sections. The hard assignment procedure intentionally imposes discrete boundaries in order to construct interpretable and policy-relevant profiles, whereas the soft-label framework is designed to reveal uncertainty and behavioural proximity across groups. Applying a deterministic assignment rule to soft memberships would ultimately reproduce a new hard classification.

To this aim, it is also useful to assess the extent to which the groups identified through the hard classification remain supported once probabilistic memberships are considered. Table 9 reports the median soft-label probabilities conditional on the original hard assignment.

**Table 9: Median soft-label memberships conditional on hard classification**

| Hard classification | $p_1^{(med)}$ | $p_2^{(med)}$ | $p_3^{(med)}$ | $p_4^{(med)}$ | $p_5^{(med)}$ |
|---------------------|---------------|---------------|---------------|---------------|---------------|
| Cash lovers         | <b>0.750</b>  | 0.005         | 0.058         | 0.001         | 0.000         |
| Bank-tech skeptics  | 0.000         | <b>0.925</b>  | 0.017         | 0.003         | 0.000         |
| Traditionalists     | 0.000         | 0.000         | <b>0.913</b>  | 0.015         | 0.000         |
| Well-connected      | 0.000         | 0.000         | 0.111         | <b>0.836</b>  | 0.000         |
| Digital-only        | 0.000         | 0.000         | 0.308         | 0.041         | <b>0.532</b>  |

Note: The columns  $p_i^{(med)}$  reports median soft-label membership probabilities conditional on the hard classification assignment with  $i = 1$ - Cash lovers, 2- Bank-tech skeptics, 3- Traditionalists, 4- Well connected, 5- Digital Only. Diagonal elements indicate the median probability associated with the profile assigned by the hard classification.

---

The results show a high degree of consistency between the two approaches. Traditionalists, Bank-tech skeptics and, to a lesser extent, Well-connected users display very high median probabilities for their own profile, indicating that these groups remain well identified even when uncertainty is explicitly taken into account. The Traditionalists profile appears remarkably stable despite representing the largest segment of the population.

By contrast, the two behavioural extremes are less strongly confirmed. Cash lovers exhibit lower self-membership probabilities than the intermediate profiles. Digital-only users emerge as the most ambiguous group. For individuals assigned to this category by the hard classification, the median probability of belonging to the Digital-only profile is only slightly above 0.5, while a substantial secondary probability is associated with the Traditionalists profile (0.3). Likely, even Digital-only users are characterised by the absence of cash usage, their broader payment habits often remain relatively close to those of more traditional consumers. In this sense, the soft-label framework provides further evidence that cashless behaviour does not necessarily coincide with a fully innovative payment profile.

At the same time, Traditionalists remain the core profile around which more nuanced behavioural configurations are organised. The soft classification not only strongly confirms individuals originally assigned to this group, but also attracts substantial secondary memberships from several other categories. This finding further reinforces the conclusion that traditional combinations of cash and card payments continue to represent the dominant behavioural pattern in Italy.

Furthermore, the probabilistic framework provides a way to assess the robustness of the proposed behavioural taxonomy itself. The five profiles were defined *ex ante* to maximise interpretability and policy relevance, drawing on three conceptual dimensions: banking integration, digital engagement, and payment orientation. Alternative conceptualisations could emphasise different behavioural dimensions or define different boundaries between groups. The soft-label analysis nevertheless suggests that the proposed segmentation captures the main behavioural structure of the data, as most individuals display a clear dominant membership and only a limited share of observations remains genuinely ambiguous.

The main exception concerns the overlap between Traditionalists and Digital-only users. Rather than indicating a weakness of the classification, this overlap points to the possible existence of an additional intermediate profile composed of consumers who have abandoned cash but are not yet characterised by the broad digital engagement expected for Digital-only users. These consumers could be interpreted as Cashless Traditionalists, maintaining relatively conventional payment habits despite exclusively relying on electronic payment instruments.

## 6 Policy implications

The classification of payment profiles provides actionable insights for policymakers and market participants. Beyond descriptive analysis, these findings inform strategies to promote financial inclusion, design CBDCs, and foster digital literacy.

**The persistence of cash.** Italy remain characterized by a relatively high use of cash compared to other advanced economies. Our approach confirms that the narrative of a sharp divide between fully disadvantaged consumers forced to rely on cash and digitally advanced users is only partially accurate. Instead, cash usage is

---

largely driven by Traditionalists, who represent the majority of Italian consumers, with a profile broadly aligned with the general population. This finding aligns with Zamora-Pérez *et al.* (2024). Therefore, the Eurosystem has committed to preserving access to cash while enhancing the payment ecosystem, ensuring that all payment preferences are taken into consideration.

**Transition toward cashless payments.** As demonstrated in studies such as Bagnall *et al.* (2016), the choice of payment instrument is heavily influenced by transaction characteristics and merchant acceptance (Bounie and François (2006), Huynh *et al.* (2014), Wakamori and Welte (2017), Lippi and Moracci (2025)). However, recent evidence indicates that individual preferences and long-standing behavioural orientations play an equally central role. Using household scanner data, Rysman *et al.* (2025) show that idiosyncratic payment preferences are highly persistent over time which implies that even in an environment where digital payments are universally available and widely accepted, consumers would not necessarily converge toward fully cashless behaviour. Our results confirm this heterogeneity: consumers who are already cashless, like Digital-only users, tend to rely on existing tools rather than actively seeking new solutions. The transition toward a cashless society depends not only on technological aspects but also on cultural and behavioural orientations. From a policy perspective, this means that cashless innovation and the preservation of access to cash are not mutually exclusive. Since most consumers use and value digital instruments, policy can continue to promote digitalisation through usability, security and broad merchant acceptance.

**Financial inclusion.** The analysis highlights the strong link between payments and financial inclusion. While the three groups open to cashless instruments display more advanced financial profiles with access to credit, savings, and investments, Cash lovers and Bank-tech skeptics, show limited engagement with banking products and dependence on cash even for income, thus exposing structural vulnerabilities in an increasingly digital environment. This calls for targeted measures to guarantee access to essential financial services (van der Cruisen and Reijerink, 2024). Preserving access to cash remains essential, but it should be complemented by simplified, low-cost banking solutions and programs that enhance digital literacy among older and low-income segments. Moreover, the higher share of women among Cash lovers suggests that gender gaps in Italy persist and require attention even if limited to a small percentage of people (Di Iorio and Rocco (2025), van der Cruisen *et al.* (2025)).

**Digital literacy and consumer protection.** Policy discussions on payment vulnerability often focus on financially excluded or cash-dependent consumers. However, our results show that even digitally integrated users may face relevant forms of financial vulnerability. The more digitally engaged profiles display higher incidence of loans and greater exposure to hybrid financial instruments that increasingly blur the distinction between payments and credit services. At the same time, Well-connected users show stronger interest in security, technological reliability, and innovative payment features, suggesting greater awareness of the potential risks associated with increasingly complex digital payment ecosystems. This apparent tension highlights that digital integration does not eliminate the need for consumer protection. On the contrary, the expansion of digital and app-based payment systems may increase consumers' exposure to security concerns, opaque credit conditions, and over-indebtedness risks. These findings claim that policies supporting payment digitalisation should be accompanied by strong regulatory safeguards, transparency

---

requirements, and financial education initiatives capable of addressing vulnerabilities that may also emerge among technologically advanced users.

**CBDC design.** The design and adoption of a CBDC in Italy will depend on its ability to meet heterogeneous user needs. Elsinger *et al.* (2025) find financial incentives and security as the strongest adoption factors overall for Austria, which aligns with descriptive evidence on key drivers of interest in new solutions for Italy, complemented by ease of use (Banca d’Italia, 2025). Our group-based analysis allows for a more granular perspective: Well-connected users are likely early adopters, given their strong interest in new solutions, while for other groups adoption will vary according to design features. Offline functionality emerges as a key requirement even among Digital-only users. Traditionalists value recommendations and social influence, suggesting that network effects could play a role. Looking at their current habits, they do not use mobile apps and rarely engage in online payments. However, they have access to both online banking and instant credit transfers, indicating that they are open to a more banking-oriented type of innovation. It will therefore be necessary to see how this mix will translate into adoption of CBDC by this group. Cash lovers and Bank-tech skeptics represent the greatest challenge: they are unlikely adopters unless CBDCs guarantee simplicity, low cost, and privacy. This evidence aligns with recent studies confirming that privacy and security are critical drivers for consumers with low trust in banks (Bijlsma *et al.*, 2024).

These findings highlight that payment behaviour is not only a matter of technological readiness but also reflects structural and socio-economic factors. Policy responses must therefore combine innovation with inclusiveness.

## 7 Conclusions

This paper proposes a novel framework for analysing payment behaviour by combining a conceptually grounded taxonomy of profiles with a hybrid methodology based on manual expertise, LLM-assisted annotation and supervised learning. Unlike approaches that focus on individual aspects of payment behaviour or rely on purely data-driven clustering, the proposed framework produces policy-relevant profiles while remaining directly applicable to large-scale survey data. The probabilistic extension further complements the hard classification by explicitly capturing behavioural overlap and uncertainty, although maintaining the interpretability of the original taxonomy.

Using Italian SPACE 2024 data, we identify five behavioural profiles that differ in payment choices, socio-demographic characteristics and innovation openness. The analysis reveals a strong concentration of Traditionalists, who combine cash and cards without fully embracing innovation. At the extremes, digital-oriented groups (Well-connected and Digital-only) exhibit heterogeneous behaviours ranging from omnichannel payment habits to fully cashless lifestyles. At the opposite end of the spectrum, only a relatively small share of the population remains weakly integrated into the banking system (Cash lovers and Bank-tech skeptics). The soft-label analysis confirms the overall robustness of the proposed taxonomy, showing that boundaries are gradual rather than perfectly discrete, particularly between Traditionalists and Digital-only users.

These findings have direct implications for payment policy. Payment strategies should account for the coexistence of heterogeneous profiles considering the diversity

---

of user needs. While many consumers are well positioned to adopt innovative payment solutions, others continue to rely on cash or combine traditional and digital instruments. Therefore, promoting innovation alongside preserving access to cash remains essential to ensuring an inclusive, resilient and efficient payment ecosystem. Although the primary focus of this paper is not CBDC adoption, the proposed framework also provides a useful basis for evaluating heterogeneous payment needs and the potential adoption of future innovations across user groups.

Several limitations should be acknowledged. First, diary data capture behaviour over a relatively short time horizon, although the integration of payment preferences and ownership indicators helps mitigate this limitation. Second, the proposed taxonomy relies on predefined behavioural profiles. Although this choice facilitates policy analysis, alternative payment ecosystems or institutional settings may require different profile definitions. The probabilistic extension partly addresses this issue by revealing behavioural overlap and uncertainty, suggesting that payment habits may evolve along gradual rather than discrete pathways. Finally, the analysis is based on the 2024 wave of the SPACE survey. Earlier editions differ in diary structure, payment instrument categorisation and questionnaire design, so extending the analysis over time would require careful harmonisation across survey waves.

The study advances the analysis of payment behaviour by introducing a conceptually grounded segmentation framework implemented through a methodology leveraging AI. This approach overcomes the limitations of more traditional statistical techniques (e.g. cluster analysis), which require the groups identified to be interpreted *ex post* and may be difficult to connect to policy needs. It also improves upon purely conceptual classifications that are not necessarily supported by data.

---

## References

- Alvarez, F. and Lippi, F. (2017). Cash burns: An inventory model with a cash-credit choice. *Journal of Monetary Economics*, 90:99–112.
- Arango, C., Huynh, K. P., and Sabetti, L. (2015a). Consumer payment choice: Merchant card acceptance versus pricing incentives. *Journal of Banking & Finance*, 55:130–141.
- Arango, C. A., Hogg, D., and Lee, A. (2015b). Why is cash (still) so entrenched? insights from canadian shopping diaries. *Contemporary Economic Policy*, 33(1):141–158.
- Arango-Arango, C. A., Bouhdaoui, Y., Bounie, D., Eschelbach, M., and Hernandez, L. (2018). Cash remains top-of-wallet! international evidence from payment diaries. *Economic Modelling*, 69:38–48.
- Artemova, E., Tsvigun, A., Schlechtweg, D., Fedorova, N., Chernyshev, K., Tilga, S., and Obmoroshev, B. (2025). Hands-on tutorial: Labeling with llm and human-in-the-loop. *arXiv preprint arXiv:2411.04637*.
- Bagnall, J., Bounie, D., Huynh, K., Kosse, A., Schmidt, T., and Schuh, S. (2016). Consumer cash usage: A cross-country comparison with payment diary survey data. *International Journal of Central Banking*, 12(4):1–61.
- Banca d'Italia (2025). Report on the payment attitudes of consumers in Italy: results from ECB SPACE 2024 survey. Markets, Infrastructures and Payment Systems Working Paper 68, Banca d'Italia.
- Bank for International Settlements (2020). Payment aspects of financial inclusion. Committee on Payments and Market Infrastructures (CPMI).
- Beck, J. (2023). Quality aspects of annotated data. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 17:331–353.
- Bian, W., Cong, L. W., and Ji, Y. (2023). The rise of e-wallets and buy-now-pay-later: Payment competition, credit expansion, and consumer behavior. Technical report, National Bureau of Economic Research.
- Bijlsma, M., van der Crujisen, C., Jonker, N., and Reijerink, J. (2024). What triggers consumer adoption of central bank digital currency? *Journal of Financial Services Research*, 65(1):1–40.
- Borzekowski, R., Elizabeth, K. K., and Shaista, A. (2008). Consumers' use of debit cards: patterns, preferences, and price response. *Journal of money, credit and banking*, 40(1):149–172.
- Bounie, D. and Camara, Y. (2020). Card-sales response to merchant contactless payment acceptance. *Journal of Banking & Finance*, 119:105938.
- Bounie, D. and François, A. (2006). Cash, check or bank card? the effects of transaction characteristics on the use of payment instruments. *Telecom Paris Economics and Social Sciences Working Paper No. ESS-06-05*.
- Bounie, D., François, A., and Van Hove, L. (2017). Consumer payment preferences, network externalities, and merchant card acceptance: An empirical investigation. *Review of Industrial Organization*, 51(3):257–290.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
- Brown, M., Felber, L., and Meyer, C. (2023). Consumer adoption and use of financial technology. *Swiss National Bank Working Paper*.
- Brown, M., Hentschel, N., Mettler, H., and Stix, H. (2022). The convenience of electronic payments and consumer cash demand. *Journal of Monetary Economics*,

- 
- 130:86–102.
- Carbó-Valverde, S. and Liñares-Zegarra, J. M. (2011). How effective are rewards programs in promoting payment card usage? empirical evidence. *Journal of Banking & Finance*, 35(12):3275–3291.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.
- Chen, H., Huynh, K. P., and Shy, O. (2019). Cash versus card: Payment discontinuities and the burden of holding coins. *Journal of Banking & Finance*, 99:192–201.
- Ching, A. T. and Hayashi, F. (2010). Payment card rewards programs and consumer payment choice. *Journal of Banking & Finance*, 34(8):1773–1787.
- Clipal, R. and Zamora-Pérez, A. (2025). Cash is alive... and somewhat young? decoupling age, period and cohort from euro cash use. *Economic Bulletin Articles*, 5.
- Di Iorio, A. and Rocco, G. (2025). Easier said than done: Why italians pay in cash while preferring cashless. *International Journal of Central Banking*, 21(3):155–184.
- Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2):139–157.
- Elsinger, H., Stix, H., and Summer, M. (2025). Consumer preferences for a digital euro: insights from a discrete choice experiment in austria. *Working Papers (Oesterreichische Nationalbank)*, (268).
- European Central Bank (2024). Study on the Payment Attitudes of Consumers in the Euro Area (SPACE) 2024. Technical report, European Central Bank.
- Falcettoni, E. and Nygaard, V. (2025). Refining the definition of the unbanked. Technical report, Board of Governors of the Federal Reserve System (US).
- Felt, M.-H., Welte, A., and Talavera, K. (2024). Untapped potential: Mobile device ownership and mobile payments in canada. Technical report, Bank of Canada Staff Working Paper.
- Greene, C., Hayashi, F., Lloro, A., Shy, O., Stavins, J., and Toh, Y. L. (2024). Defining households that are underserved in digital payment services. *Federal Reserve Bank of Kansas City Working Paper No. RWP*, pages 24–09.
- Henry, C. S., Engert, W., Sutton-Lalani, A., Hernandez, S., McVanel, D., and Huynh, K. P. (2023). Unmet payment needs and a central bank digital currency. *Journal of Digital Banking*, 8(3):242–255.
- Huang, C. and He, G. (2025). Text clustering as classification with llms. *SN Computer Science*, 6(2).
- Huynh, K. P., Schmidt-Dengler, P., and Stix, H. (2014). The role of card acceptance in the transaction demand for money. *Bank of Canada Working Paper*, (2014/44).
- Ilić, V. (2024). Reducing manual labeling effort by identifying the most informative unlabeled data via active learning. *ResearchGate Preprint*.
- Klee, E. (2008). How people pay: Evidence from grocery store data. *Journal of Monetary Economics*, 55(3):526–541.
- Klie, J.-C., de Castilho, R. E., and Gurevych, I. (2024). Analyzing dataset annotation quality management in the wild. *Computational Linguistics*, 50(3):817–

- Lee, G. H., Lee, K. J., Jeong, B., and Kim, T. K. (2024). Developing personalized marketing service using generative ai. In *Proceedings of the 2024 International Conference on Artificial Intelligence and Data Science*. IEEE.
- Li, J. (2023). Predicting the demand for central bank digital currency: A structural analysis with survey data. *Journal of Monetary Economics*, 134:73–85.
- Li, L., Sleem, L., Gentile, N., Nichil, G., and State, R. (2025). Exploring the impact of temperature on large language models: Hot or cold? *Procedia Computer Science*, 264:242–251. International Neural Network Society Workshop on Deep Learning Innovations and Applications 2025.
- Lippi, F. and Moracci, E. (2025). Cash or card? a structural model of payment choices.
- Mehta, R., Ghosh, D., and McCallum, A. (2021). Scalable and interpretable text classification with semi-supervised learning. *Transactions of the Association for Computational Linguistics*, 9:1–15.
- Northcutt, C. G., Jiang, L., and Chuang, I. L. (2021). Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, 70:137–198.
- OpenAI (2023). Gpt-4 technical report. arXiv:2303.08774.
- Pangakis, D., Schroeder, F., and Gurevych, I. (2024). Knowledge distillation in automated annotation: Supervised text classification with llm-generated training labels. In *Proceedings of the Workshop on NLP for Conversational Social Search (NLPCSS)*.
- Qi, X., Tang, J., Chang, Y., and Liu, H. (2012). Community detection with edge content in social media networks. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1817–1821. ACM.
- Qian, Y., Chen, S., Du, Y., Wu, C., Yuan, K., and Jiang, Y. (2025). Large language models for marketing research: A survey and new perspectives. In *Proceedings of the 58th Hawaii International Conference on System Sciences*.
- Reijerink, J. (2024). Distinguishing payment user groups by clustering transaction data. DNB Working Paper 2024, De Nederlandsche Bank.
- Rysman, M., Wang, S., and Wozniak, K. (2025). Understanding preferences for payment cards using household scanner data.
- Shy, O. (2023). Cash is alive: How economists explain holding and use of cash. *Journal of Economic Literature*, 61(4):1465–1520.
- Singh, A., Tiwari, A., Hasanbeig, H., and Gupta, P. (2025). Soft-label training preserves epistemic uncertainty. *arXiv preprint arXiv:2511.14117*.
- Stavins, J. (2018). Consumer preferences for payment methods: Role of discounts and surcharges. *Journal of Banking & Finance*, 94:35–53.
- Sumanathilaka, D., Micallef, N., and Hough, J. (2025). Exploring the impact of temperature on large language models: A case study for classification task based on word sense disambiguation. In *Proceedings of the 2025 7th International Conference on Natural Language Processing (ICNLP)*, pages 178–182.
- Świecka, B., Terefenko, P., and Paprotny, D. (2021). Transaction factors’ influence on the choice of payment by polish consumers. *Journal of Retailing and Consumer Services*, 58:102264.
- Trütsch, T. (2020). The impact of contactless payment on cash usage at an early

- 
- stage of diffusion. *Swiss Journal of Economics and Statistics*, 156(1):1–35.
- van der Crujsen, C., Broekhoff, M.-C., and Knoben, J. (2025). Mind the gap: gender differences in household payment tasks. *De Economist*, pages 1–40.
- van der Crujsen, C. and Knoben, J. (2021). Ctrl+ c ctrl+ pay: Do people mirror electronic payment behavior of their peers? *Journal of Financial Services Research*, 59(1):69–96.
- van der Crujsen, C. and Reijerink, J. (2024). Uncovering the digital payment divide: understanding the importance of cash for groups at risk. *Journal of Consumer Affairs*, 58(2):486–505.
- Von Kalckreuth, U., Schmidt, T., and Stix, H. (2014). Choosing and using payment instruments: evidence from german microdata. *Empirical Economics*, 46(3):1019–1055.
- Wakamori, N. and Welte, A. (2017). Why do shoppers use cash? evidence from shopping diary data. *Journal of Money, Credit and Banking*, 49(1):115–169.
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., and Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.
- Zamora-Pérez, A., Marini, A., and Honkkila, J. (2024). Is there a digital divide in payments? understanding why cash remains important for so many. *Economic Bulletin Articles*, 2.

---

# Appendices

## A List of variables

This appendix reports the complete list of variables employed in the study.

Table A1 presents the indicators used for the classification process, excluding behavioural data from payment diaries, but including attitudinal variables on preferences and ownership of instruments. Table A2 illustrates the complete prompt structure provided to the LLM classifier. It includes the aggregation of diary transactions across payment contexts, the attitudinal variables on preferences and ownership, and the full instruction block used to derive behavioural group assignments. Table A3 presents the aggregated behavioural inputs used for classification, reporting user level totals across payment contexts together with the relevant attitudinal indicators.

Tables A4, A5 and A6 provide the variables used ex post to profile consumer groups, covering socio-demographic characteristics, financial access, and openness to innovation, respectively. Variable names and codes correspond to the internal conventions adopted in this work, which are based on the original SPACE 2024 questionnaire structure.

**Table A1: Variables used for labelling consumers through manual annotation and LLM**

|             | Variable            | Description                                                                                                                                 | Type                                                                                           |
|-------------|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| Preferences | POS preference      | Categorical variable indicating the preferred payment method at the POS                                                                     | 1 = Cash<br>2 = Card or cashless<br>3 = No clear preference                                    |
|             | Cashless preference | Categorical variable indicating the preferred cashless payment method at the POS in case the variable POS preference is different from cash | 1 = Card<br>2 = Mobile payments<br>3 = Bank cheque<br>4 = Prepaid card                         |
|             | Cash Importance     | Categorical variable indicating the importance of having the option to use cash                                                             | 1 = Very important<br>2 = Fairly important<br>3 = Not so important<br>4 = Not important at all |
|             | P2P preference      | Categorical variable indicating the preferred method for P2P transfers (friends/family)                                                     | 1 = Cash<br>2 = Bank transfer<br>3 = Mobile payment<br>4 = Other                               |
|             |                     |                                                                                                                                             |                                                                                                |
| Ownership   | Bank Account        | Binary variable indicating if the respondent has a bank account                                                                             | 0 = No<br>1 = Yes                                                                              |
|             | Payment Card        | Binary variable indicating if the respondent has a payment card                                                                             | 0 = No<br>1 = Yes                                                                              |
|             | Mobile              | Binary variable indicating if the respondent has a mobile phone that can be used to make payments                                           | 0 = No<br>1 = Yes                                                                              |
|             | None                | Binary variable indicating if the respondent does not have any of the above instruments                                                     | 0 = No<br>1 = Yes                                                                              |

---

**Table A2: Example of full prompt and aggregated input supplied to the LLM**

This classification divides users into five behavioural groups based on their payment habits. The data are structured as tot\_INSTRUMENT\_USECASES. The data report total number of transactions for 5 main groups of instruments in the different use cases.

INSTRUMENT: cash: cash dccard: debit and credit card pcard: prepaid cards mobile: digital wallets, crypto, epayments and mobile in general other\_bank (cheques, direct debits and credit transfers) other (other instruments as vouchers and residuals)

USECASES: Point of Sale (POS), Person to Person (P2P), Online Purchases (REM), Recurring Payments (BILL).

Moreover, consider also the following additional variables related to declared preferences and ownership of payment tools:

Preferences:

QQ3: Indicates the user's preferred payment method in physical stores (1 = Cash, 2 = Digital, 3 = No clear preference).

QQ3A: Indicates the user's preference for cashless payment instruments (1 = Card, 2 = Mobile including wearables, 3 = Checks, 4 = Prepaid cards, 6 = No preference, 999998 = Refuse, 999999 = Don't know).

QQ4C: Measures how important it is for the user to have the option to use cash (1 = Very important, 4 = Not important at all).

QQ17: Indicates the user's preferred method for sending money to friends/family (1 = Cash, 2 = Bank transfer, 3 = Mobile/online, 4 = No clear preference).

Ownership:

QQ1A 1: Possession of a bank account that allows payments. QQ1A 2: Possession of a payment card (debit or credit). QQ1A 6: Possession of a mobile device or wearable that supports app-based payments. QQ1A 4: Indicates no ownership of any of the above tools.

-- Description of the groups:

1. Cash Lover Rely exclusively on cash. These individuals lack access to or intentionally avoid cards and digital tools. They strongly prefer the anonymity and simplicity of cash.
2. Bank-tech Skeptics Use cash and prepaid cards but reject more advanced banking technologies such as debit/credit cards and mobile payments.
3. Traditionalists Combine cash with cards for POS transactions, but rarely engage in online or mobile payments.
4. Well-Connected Consumers Well-integrated into the digital economy, using cash, cards, digital wallets, and mobile solutions across all contexts.
5. Digital-Only Lifestyle Fully cashless users who rely exclusively on digital instruments, including mobile wallets and e-payments.

--

*Behavioural table omitted for brevity.*

Use this document as context when classifying users based on their payment behaviour. Please always choose one among the five predefined behavioural groups and return the answer in the format: "Number. Name".

USER DATA:LLM\_input\_text

---

**Table A3: Example of aggregated behavioural input supplied to the LLM**

USER DATA

User 11000056:

Point of Sale: tot cash POS = 3, tot dccard POS = 0, tot pcard POS = 0, tot mobile POS = 0, tot other bank POS = 0, tot other POS = 0, tot POS = 3

Online / Remote Payments: tot dccard REM = 0, tot pcard REM = 0, tot mobile REM = 0, tot other bank REM = 0, tot other REM = 0, tot REM = 0

Person to Person: tot cash P2P = 0, tot pcard P2P f = 0, tot dccard P2P f = 0, tot mobile P2P f = 0, tot other P2P f = 0, tot dccard P2P r = 0, tot pcard P2P r = 0, tot mobile P2P r = 0, tot other bank P2P r = 0, tot other P2P r = 0, tot pcard P2P = 0, tot dccard P2P = 0, tot mobile P2P = 0, tot other bank P2P = 0, tot other P2P = 0, tot P2P = 0

Recurring Payments: tot cash BILL = 0, tot dccard BILL = 0, tot pcard BILL = 4, tot other bank BILL = 1, tot other BILL = 0, tot BILL = 5

Overall totals: tot cash = 3, tot dccard = 0, tot pcard = 4, tot mobile = 0, tot other bank = 1, tot other = 0

Declared preferences: No clear preference in stores; Prefers prepaid cards; Cash not very important; No clear preference for P2P

Owned instruments/devices: Has a payment card.

---

**Table A4: Variables for group characterization: Socio-Demographic**

| Variable      | Description                                                                                                                         | Type              |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| Low Education | Binary variable indicating if the highest level of education is primary/lower secondary education                                   | 0 = No<br>1 = Yes |
| Low Income    | Binary variable indicating if the income is less than 1.500 euro                                                                    | 0 = No<br>1 = Yes |
| High Income   | Binary variable indicating if the income is more than 4.000 euro                                                                    | 0 = No<br>1 = Yes |
| Not Occupated | Binary variable indicating if the respondent declares that he does not carry out any professional activity or is a student          | 0 = No<br>1 = Yes |
| Self Employed | Binary variable indicating if the respondent declares himself self-employed                                                         | 0 = No<br>1 = Yes |
| GenZ          | Binary variable indicating if the age of the respondent is less than 29                                                             | 0 = No<br>1 = Yes |
| Over 65       | Binary variable indicating if the age of the respondent is more than 65                                                             | 0 = No<br>1 = Yes |
| Female        | Binary variable indicating if the respondent is female                                                                              | 0 = No<br>1 = Yes |
| South         | Binary variable indicating if the respondent resides in Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicily or Sardinia | 0 = No<br>1 = Yes |

---

---

**Table A5: Variables for group characterization: Financial characteristics**

| Variable                | Description                                                                                                    | Type              |
|-------------------------|----------------------------------------------------------------------------------------------------------------|-------------------|
| Mainly Income in cash   | Binary variable indicating if the amount of regular income received in cash is between half and three-quarters | 0 = No<br>1 = Yes |
| Financial Investment    | Binary variable indicating if the respondent has financial investment                                          | 0 = No<br>1 = Yes |
| Savings                 | Binary variable indicating if the respondent has savings account                                               | 0 = No<br>1 = Yes |
| Mortgage or Loans       | Binary variable indicating if the respondent has mortgage or loans                                             | 0 = No<br>1 = Yes |
| Online Banking          | Binary variable indicating if the respondent has online banking services                                       | 0 = No<br>1 = Yes |
| Instant Credit transfer | Binary variable indicating if the respondent has access and use instant payments                               | 0 = No<br>1 = Yes |
| Crypto                  | Binary variable indicating if the respondent has crypto assets                                                 | 0 = No<br>1 = Yes |

---

**Table A6: Variables for group characterization: interest in new payment Solution**

| Variable       | Description                                                                                                                               | Type              |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| Not interested | Binary variable indicating if the respondent is not interested in trying a new solution                                                   | 0 = No<br>1 = Yes |
| Cheaper        | Binary variable indicating if cheaper is one of the main factors that would convince the respondent to try out a new payment solution     | 0 = No<br>1 = Yes |
| Easier         | Binary variable indicating if easier is one of the main factors that would convince the respondent to try out a new payment solution      | 0 = No<br>1 = Yes |
| Security       | Binary variable indicating if security is one of the main factors that would convince the respondent to try out a new payment solution    | 0 = No<br>1 = Yes |
| Faster         | Binary variable indicating if faster is one of the main factors that would convince the respondent to try out a new payment solution      | 0 = No<br>1 = Yes |
| Privacy        | Binary variable indicating if privacy is one of the main factors that would convince the respondent to try out a new payment solution     | 0 = No<br>1 = Yes |
| Incentives     | Binary variable indicating if incentives is one of the main factors that would convince the respondent to try out a new payment solution  | 0 = No<br>1 = Yes |
| Innovative     | Binary variable indicating if innovative is one of the main factors that would convince the respondent to try out a new payment solution  | 0 = No<br>1 = Yes |
| Offline        | Binary variable indicating if offline is one of the main factors that would convince the respondent to try out a new payment solution     | 0 = No<br>1 = Yes |
| Recommended    | Binary variable indicating if recommended is one of the main factors that would convince the respondent to try out a new payment solution | 0 = No<br>1 = Yes |

---

## B Interacting with Large Language Models

The process of selecting the instruction text to be provided as input to the LLM is not standardized and requires considerable time and experimentation, a process commonly referred to as *prompt engineering*. This appendix discusses the various attempts carried out in order to identify the “most effective” prompt for our task and to assess its alignment with human reasoning. Initially, the prompts were constructed following a rule based logic, providing the model with a set of explicit criteria to be applied through concrete *Logical Rules*. This approach was also motivated by patterns made during the observation of the manual labeling of the subsample, where a subconscious application of such rules became evident. It is natural for a human annotator to seek specific features to categorize a user, for example by immediately excluding individuals from the more “extreme” groups (Cash lovers and Digital-only users in our case) based on clear behavioural signals. The following is an example of the rule set used to identify users belonging to the Digital-only users :

- Cash usage is zero across all contexts: `tot_cash == 0`.
- At least one digital instrument is used in any context (POS, REM, P2P, BILL) such as `tot_mobile > 0`, `tot_dccard > 0` or `tot_pcard > 0`.

This rule-based approach (hereinafter referred to as Hard) achieved a match rate of over 70% with the human assigned labels. Specifically, the match was respectively 100% and 88% for Cash Lovers and Digital-only. The Bank-tech skeptics groups also showed very high alignment, with match rates close to 85%. In contrast, for the more intermediate groups of Traditionalists and Well-connected users, the match rate dropped to 68% and 58%. Although these values still represent a clear majority of correct matches, they are markedly lower than those observed for the more extreme groups. This pattern reinforces the idea that Traditionalists and Well-connected users exhibit more heterogeneous and less sharply defined behaviours, making them harder to capture through rigid rule-based criteria.

On the one hand, rule-based prompts provide stability and interpretability; on the other hand, they impose constraints: when the prompt uses strong and highly explicit phrasing, the model almost invariably follows the given instructions. This enhances consistency, but it also reduces the opportunity to observe how the model interprets the data autonomously, thereby limiting the exploration of its reasoning capabilities and the inherent subjectivity of classification.

Given that human assigned labels were intended as a starting point rather than a definitive benchmark, we gradually shifted toward a more flexible approach. In subsequent attempts, we removed the explicit rules, allowing the model to interpret the user descriptions and assess how they aligned with the underlying behavioural patterns. This transition enabled a more nuanced exploration of the model’s reasoning and its ability to capture complex and context dependent behaviour.

Table B1 presents an example of the different descriptions used to characterize the Cash lovers group.

As we worked through these different prompt families, we systematically compared the results with both the human assigned labels and the initial hard rule version of the prompt. Throughout this process, we kept a critical perspective and embraced iterative refinements aimed at achieving a balance between stability and

---

**Table B1: Comparison of Cash lovers Group across prompt versions**

| Style                             | Group Description                                                                                                                                                                |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Synthetic and functional</b>   | <i>Rely exclusively on cash. Lack access to or intentionally avoid cards and digital tools.</i>                                                                                  |
| <b>Synthetic</b>                  | <i>Rely exclusively on cash. Avoid cards and digital tools.</i>                                                                                                                  |
| <b>Narrative and sociological</b> | <i>They typically do not have a bank account or choose to avoid banking services altogether. Their preference for cash is driven by its simplicity and anonymity.</i>            |
| <b>Narrative and technical</b>    | <i>They are either unbanked or intentionally disconnected from formal financial systems. They tend to reject digital tools and prefer cash for its simplicity and anonymity.</i> |

---

interpretability. Our goal was not to replicate human labels mechanically, but to identify a formulation capable of capturing behavioural nuances while maintaining internal coherence across groups.

Among the tested alternatives, the Synthetic and Functional version (hereafter SynFunc) emerged as the most robust configuration.

Figure B9 illustrates the distribution of labels generated by the different prompt families compared with the human annotation highlighting substantial differences in the distribution of groups across prompts. Narrative formulations tend to concentrate observations in the Traditionalists category and reduce the share of Well-connected users, whereas the Hard configuration produces sharper distinctions for the most extreme behavioural profiles (Cash lovers and Digital-only users).

By contrast, SynFunc generates a distribution that remains comparatively close both to the human annotation and to the other prompt families, without excessively concentrating observations in specific groups.

**Figure B9: Comparison of group distribution across human annotation and multiple LLM versions**

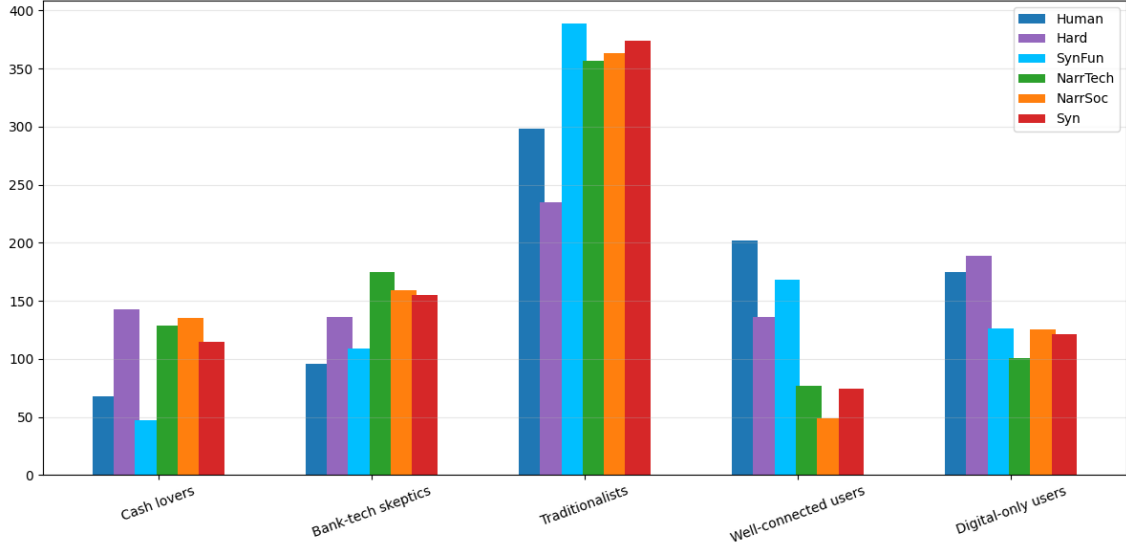


Table B2 provides a quantitative comparison across all prompt versions. For clarity, the abbreviations used in the table are: SynFunc = Synthetic and functional (bold), NarrTech = Narrative and technical, NarrSoc = Narrative and sociological, Syn = Synthetic.<sup>14</sup>

The comparison confirms that the choice of SynFunc is motivated not by its maximum agreement with a single benchmark, but by its relatively stable alignment across all prompting strategies. While other prompts achieved stronger agreement with either the Hard labels or specific narrative formulations, they also produced more polarized classification patterns. SynFunc acts as an intermediate configuration, achieving substantial agreement with both human labels (67.8%) and the Hard configuration (60.9%), while remaining consistently aligned with the other prompt families, with values ranging between 61% and 66%. For these reasons, SynFunc was selected as the final configuration, as it provided the most convincing balance between cross prompt consistency and the preservation of a heterogeneous and behaviourally plausible group distribution.

It might be interesting to note that prompts such as NarrSoc and Syn exhibited extremely high mutual agreement (above 90%), suggesting convergence toward very similar classification behaviours. But these formulations also generated less balanced group distributions, particularly underrepresenting the Well-connected users category, as said before.

To further evaluate the robustness of the selected configuration, Tables B3 and B4 report the confusion matrices comparing the final SynFunc prompt with, respectively, the Hard labels and the human annotation.<sup>15</sup>

<sup>14</sup>See Table B1 for examples of these styles.

<sup>15</sup>These figures correspond to the prompt configuration that focused exclusively on diary data; in its final iteration this version was extended to incorporate both preferences and ownership variables. Furthermore, these comparisons do not include the second step of review that led to the labels used for training.

**Table B2: Match between different prompts**

|                | Human | Hard  | SynFunc | NarrTech | NarrSoc | Syn |
|----------------|-------|-------|---------|----------|---------|-----|
| Human          |       |       |         |          |         |     |
| Hard           | 0.744 |       |         |          |         |     |
| <b>SynFunc</b> | 0.678 | 0.609 |         |          |         |     |
| NarrTech       | 0.589 | 0.744 | 0.631   |          |         |     |
| NarrSoc        | 0.602 | 0.770 | 0.616   | 0.909    |         |     |
| Syn            | 0.631 | 0.764 | 0.662   | 0.905    | 0.912   |     |

The matrices show that disagreements are concentrated primarily in the intermediate groups, particularly between Traditionalists and Well-connected users, where behavioural boundaries appear less clearly defined. At the same time, some discrepancies also emerge for the more extreme categories when comparing the final prompt with the Hard configuration. This reflects the fact that the rule-based approach tends to enforce sharper distinctions, whereas SynFunc allows for more flexible interpretations of behavioural patterns. As a result, the final prompt partially smooths the rigid boundaries imposed by deterministic rules while preserving substantial overall agreement with both human and rule-based classifications.

The confusion matrices support the validity of the selected configuration, as disagreements are concentrated in profiles characterized by overlapping behavioural patterns rather than in clearly distinct cases.

**Table B3: Hard (row) vs Final Prompt**

| Label | 1  | 2   | 3   | 4   | 5   |     |
|-------|----|-----|-----|-----|-----|-----|
| 1     | 47 |     | 91  | 5   |     | 143 |
| 2     |    | 94  | 36  | 5   | 1   | 136 |
| 3     |    | 3   | 181 | 50  | 1   | 235 |
| 4     |    | 9   | 38  | 77  | 12  | 136 |
| 5     |    | 3   | 43  | 31  | 112 | 189 |
|       | 47 | 109 | 389 | 168 | 126 | 839 |

*Total: 0.61%*

**Table B4: Human (row) vs Final Prompt**

| Label | 1  | 2   | 3   | 4   | 5   |     |
|-------|----|-----|-----|-----|-----|-----|
| 1     | 43 |     | 24  | 1   |     | 68  |
| 2     |    | 70  | 15  | 4   | 7   | 96  |
| 3     | 2  | 16  | 246 | 34  |     | 298 |
| 4     | 2  | 18  | 75  | 99  | 8   | 202 |
| 5     |    | 5   | 29  | 30  | 111 | 175 |
|       | 47 | 109 | 389 | 168 | 126 | 839 |

*Total: 0.68%*

---

To conclude, prompt engineering should not be interpreted as a purely mechanical optimization exercise aimed at maximizing agreement with a predefined benchmark. Rather, it represents an iterative process of calibration in which different formulations emphasize different dimensions of the data and, consequently, produce different behavioural interpretations. In this context, the objective of the analysis was to identify a prompt capable not only of reproducing human annotations, but also of preserving the complexity and heterogeneity of the observed payment behaviours. The selected SynFunc formulation emerged as the most suitable solution because it combined relatively high agreement with human and rule based classifications with a more balanced and behaviourally plausible distribution of groups.

More broadly, these results highlight that the formulation of prompts can substantially influence classification outcomes, particularly in contexts characterized by nuanced and overlapping categories. This means that prompt engineering should be treated as a substantive methodological component of the analysis rather than a purely technical preprocessing step.

---

**Table B5: Examples of cases “LLM Confirmations”**

| LLM Group          | Human Group          | Final Decision | Motivation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|--------------------|----------------------|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cash lovers        | Traditionalists      | LLM            | The user uses exclusively cash (both for POS and recurring payments) and traditional banking instruments only for recurring payments (other banks, typically bank transfers or direct debits for bills). There are no card, mobile, or online transactions. The complete absence of digital instruments and the predominance of cash best positions this user as a Cash Lover, as Traditionalists also expect the use of cards (which are completely absent here).                                                                       |
| Bank-tech skeptics | Well-connected users | LLM            | The user doesn’t use cash or mobile/advanced digital tools, but only cards (especially prepaid cards) and some traditional banking tools. They don’t display the typical diversity of Well-connected users (no mobile phone, wallet, or combined use of digital/advanced tools); instead, they better fit the Bank-tech skeptics profile, which shuns the most innovative tools and favors more basic and secure options (prepaid cards and some use of simple banking tools).                                                           |
| Digital-only users | Bank-tech skeptics   | LLM            | The user has not made any cash, card, or prepaid transactions; he uses exclusively digital tools (mobile and online, 7 transactions) and “other” tools for recurring payments, never using cash or traditional tools. This behaviour is consistent with the definition of a “Digital-only ” (no cash, only digital tools). No signs of skepticism toward banking technology emerge, such as the preferential use of cash or prepaid cards or the refusal of mobile devices. Therefore, I confirm the classification proposed by the LLM. |

---

Note: Motivations provided by the model for cases in which it retained its original label after the review round and comparison with the human label.

**Table B6: Examples of cases “Human Confirmations”**

| LLM Group          | Human Group          | Final Decision | Motivation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------------------|----------------------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Digital-only users | Well-connected users | Human          | The user uses various instruments (cash, cards, other banking methods) but does not demonstrate a total avoidance of cash (at least one cash transaction) nor a predominant use of exclusively digital instruments. The absence of digital payments (mobile, pre-paid) and the presence of both cash and card/recurring banking instruments better fits the “Well-connected users” profile rather than the “Digital-only ” profile, which presupposes the absence of cash. I confirm the human classification. |
| Traditionalists    | Cash lovers          | Human          | All transactions are exclusively in cash, without the use of cards or other digital tools in any usage scenario. According to the reference, this is a typical characteristic of the “Cash lovers” group. The human classification is therefore more appropriate than “Traditionalists,” which would require a mix of cash and cards.                                                                                                                                                                          |

Note: Motivations provided by the model for cases in which it switched from its original label to adopt the human label after the review round and comparison with the human label.

## RECENTLY PUBLISHED PAPERS IN THE 'MARKETS, INFRASTRUCTURES, PAYMENT SYSTEMS' SERIES

- n. 55 Rating the Raters. A Central Bank Perspective, *by Francesco Columba, Federica Orsini and Stefano Tranquillo*
- n. 56 A general framework to assess the smooth implementation of monetary policy: an application to the introduction of the digital euro, *by Annalisa De Nicola and Michelina Lo Russo*
- n. 57 The German and Italian Government Bond Markets: The Role of Banks versus Non-Banks. A joint study by Banca d'Italia and Bundesbank, *by Puriya Abbassi, Michele Leonardo Bianchi, Daniela Della Gatta, Raffaele Gallo, Hanna Gohlke, Daniel Krause, Arianna Miglietta, Luca Moller, Jens Orben, Onofrio Panzarino, Dario Ruzzi, Willy Scherrieble and Michael Schmidt*
- n. 58 Chat Bankman-Fried? An Exploration of LLM Alignment in Finance, *by Claudia Biancotti, Carolina Camassa, Andrea Coletta, Oliver Giudice and Aldo Glielmo*
- n. 59 Modelling transition risk-adjusted probability of default, *by Manuel Cugliari, Alessandra Iannamorelli and Federica Vassalli*
- n. 60 The use of Banca d'Italia's credit assessment system for Italian non-financial firms within the Eurosystem's collateral framework, *by Stefano Di Virgilio, Alessandra Iannamorelli, Francesco Monterisi and Simone Narizzano*
- n. 61 Fintech Classification Methodology, *by Alessandro Lentini, Daniela Elena Munteanu and Fabrizio Zennaro*
- n. 62 The Rise of Climate Risks: Evidence from Expected Default Frequencies for Firms, *by Matilde Faralli and Francesco Ruggiero*
- n. 63 Exploratory survey of the Italian market for cybersecurity testing services, *by Anna Barcheri, Luca Bastianelli, Tommaso Curcio, Luca De Angelis, Paolo De Joannon, Gianluca Ralli and Diego Ruggeri*
- n. 64 A practical implementation of a quantum-safe PKI in a payment systems environment, *by Luca Buccella and Stefano Massi*
- n. 65 Stewardship Policies. A Survey of the Main Issues, *by Marco Fanari, Enrico Bernardini, Elisabetta Cecchet, Francesco Columba, Johnny Di Giampaolo, Gabriele Fraboni, Donatella La Licata, Simone Letta, Gianluca Mango and Roberta Occhilupo*
- n. 66 Is there an equity greenium in the euro area?, *by Marco Fanari, Marianna Caccavaio, Davide Di Zio, Simone Letta and Ciriaco Milano*
- n. 67 Open Banking in Italy: A Comprehensive Report, *by Carlo Cafarotti and Ravenio Parrini*
- n. 68 Report on the payment attitudes of consumers in Italy: results from ECB SPACE 2024 survey, *by Gabriele Coletti, Marialucia Longo, Laura Painelli, Emanuele Pimpini and Giorgia Rocco*
- n. 69 A solution for cross-border and cross-currency interoperability of instant payment systems, *by Domenico Di Giulio, Vitangelo Lasorella, Pietro Tiberi*
- n. 70 Do firms care about climate change risks? Survey evidence from Italy, *by Francesca Colletti, Francesco Columba, Manuel Cugliari, Alessandra Iannamorelli, Paolo Parlamento and Laura Tozzi*
- n. 71 Demand and supply of Italian government bonds during the exit from expansionary monetary policy, *by Fabio Capasso, Francesco Musto, Michele Pagano, Onofrio Panzarino, Alfonso Puorro and Vittorio Siracusa*
- n. 72 Statistics on tokenized financial instruments: A challenge for central banks, *by Riccardo Colantonio, Massimo Coletta, Riccardo Renzi*
- n. 73 Credit Risk Assessment with Stacked Machine Learning, *by Francesco Columba, Manuel Cugliari, Stefano Di Virgilio*

- n. 74 What if Ether Goes to Zero? How Market Risk Becomes Infrastructure Risk in Crypto, *by Claudia Biancotti*
- n. 75 The Cyber Risk of Non-Financial Firms, *by Francesco Columba, Manuel Cugliari, Marco Orlandi, Federica Vassalli*
- n. 76 Sustainability and financial innovation: The emerging role of Fintech for Good (F4G), *by Alessandro Lentini and Daniela Elena Munteanu*
- n. 77 Hydrogeological and credit risk: the Italian firms' physical risk-adjusted probability of default, *by Manuel Cugliari, Simone Narizzano and Federica Vassalli*
- n. 78 Liquidity Optimization in Gross Settlement Systems with Quantum Reordering: Application to TARGET2, *by Valerio Astuti, Adriano Baldeschi, Luca Bastianelli, Giuseppe Bruno, Ajit Desai, Danica Marsden and Riccardo Russo*
- n. 79 The expert assessment within Banca d'Italia's in-house credit assessment system, *by Lorenzo Esposito, Massimo Guglielmi, Francesco Monterisi, Simone Narizzano and Marco Orlandi*
- n. 80 Digital payments and economic performance: evidence from Italy, *by Guerino Ardizzi, Niccolò Lippi Boncambi, Cristina Demma and Alberto Leorati*
- n. 81 The euro-area repo market: structure, participants and interlinkages, *by Lorenzo Caverni, Annalisa De Nicola and Mattia Persico*
- n. 82 Siamese Networks for AI-powered Automated Banknote Quality Control, *by Salvatore Gentile, Andrea Luciani, Sabina Marchetti, Domenico Pansini and Marco Viticoli*
- n. 83 Private Equity and Innovation in the Euro Area, *by Emanuele Degani, Simone Letta and Tommaso Perez*
- n. 84 Social and Governance Factors in Default Risk Assessment, *by Manuel Cugliari, Stefano Di Virgilio, Enrico Foscolo and Alessandra Iannamorelli*
- n. 85 Household Allocation of the Next Generation of Digital Money, *by Marcello Pericoli*
- n. 86 Are Stablecoins Efficient for Remittances? Evidence from a Mystery Shopping Exercise by Banca d'Italia, *by Alberto Di Iorio, Enrica Di Stefano, Michele Mascioli and Giorgio Trebeschi*
- n. 87 Who uses 'Buy Now, Pay Later' in Italy? Evidence from Household Survey Data, *by Daniele Figoli and Serena Palazzo*
- n. 88 Fee-cap Models for Payment Instruments: a Policy Perspective, *by Nicola Branzoli, Stefano Pietrosanti and Alessia Vita*
- n. 89 Send me money ASAP: Fast Payment Systems and the Cost of Remittances, *by Marco Brandi, Alberto Di Iorio and Andrea Nobili*
- n. 90 Hedging at speed in sovereign bond markets, *by Onofrio Panzarino and Antonio Perrella*
- n. 91 Sentiment analysis with large language models for credit risk assessment, *by Manuel Cugliari and Giulio Gariano*
- n. 92 The Architecture of Cross-Border Finance: Fifty Years of Evolution from National Segmentation to Fragmented Globalization, *by Cristina Di Luigi and Antonio Perrella*
- n. 93 Privacy-Enhancing Technologies and Auditability in Digital Payment System, *by Matteo Nardelli, Francesco De Sclavis and Michela Iezzi*
- n. 94 The Credit Assessment of Large Firms, *by Lorenzo Esposito, Marisa Franzoni, Massimo Guglielmi, Alessandra Iannamorelli and Raffaella Sorgente*