

Article

Digital Skills, Ethics, and Integrity—The Impact of Risky Internet Use, a Multivariate and Spatial Approach to Understanding NEET Vulnerability

Adriana Grigorescu ^{1,2,3,*} , Teodor Victor Alistar ¹  and Cristina Lincaru ⁴ 

- ¹ Faculty of Public Administration, National University of Political Studies and Public Administration, Expoziției Boulevard, 30A, 012104 Bucharest, Romania; victor.alistar@transparency.org.ro
- ² Academy of Romanian Scientists, Ilfov Street 3, 050094 Bucharest, Romania
- ³ National Institute for Economic Research “Costin C. Kiritescu”, Romanian Academy, Casa Academiei Române, Calea 13 Septembrie nr. 13, 050711 Bucharest, Romania
- ⁴ National Scientific Research Institute for Labor and Social Protection, Povernei Street 6, 010643 Bucharest, Romania; cristina.lincaru@yahoo.de
- * Correspondence: adriana.grigorescu@snsa.ro

Abstract

In an era where digitalization shapes economic and social landscapes, the intersection of digital skills, ethics, and integrity plays a crucial role in understanding the vulnerability of youth classified as NEET (Not in Education, Employment, or Training). This study explores how risky internet use and digital skill gaps contribute to socio-economic exclusion, integrating a multivariate and spatial approach to assess regional disparities in Europe. This study adopts a systems thinking perspective to explore digital exclusion as an emergent outcome of multiple interrelated subsystems. The research employs logistic regression, Principal Component Analysis (PCA) with Promax rotation, and Geographic Information Systems (GIS) to examine the impact of digital behaviors on NEET status. Using Eurostat data aggregated at the country level for the period (2000–2023) across 28 European countries, this study evaluates 24 digital indicators covering social media usage, instant messaging, daily internet access, data protection awareness, and digital literacy levels. The findings reveal that low digital skills significantly increase the likelihood of being NEET, while excessive social media and internet use show mixed effects depending on socio-economic context. A strong negative correlation between digital security practices and NEET status suggests that youths with a higher awareness of online risks are less prone to socio-economic exclusion. The GIS analysis highlights regional disparities, where countries with limited digital access and lower literacy levels exhibit higher NEET rates. Digital exclusion is not merely a technological issue but a multidimensional socio-economic challenge. To reduce the NEET rate, policies must focus on enhancing digital skills, fostering online security awareness, and addressing regional disparities. Integrating GIS methods allows for the identification of territorial clusters with heightened digital vulnerabilities, guiding targeted interventions for improving youth employability in the digital economy.

Keywords: digital skills; risky internet use; ethics; integrity; youth employment; digital literacy; logistic regression; GIS analysis



Academic Editor: Andrea Pitasi

Received: 15 May 2025

Revised: 6 July 2025

Accepted: 29 July 2025

Published: 1 August 2025

Citation: Grigorescu, A.; Alistar, T.V.; Lincaru, C. Digital Skills, Ethics, and Integrity—The Impact of Risky Internet Use, a Multivariate and Spatial Approach to Understanding NEET Vulnerability. *Systems* **2025**, *13*, 649. <https://doi.org/10.3390/systems13080649>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Youth digital exclusion, particularly among those Not in Education, Employment, or Training (NEET), is not merely the outcome of individual deficiencies but a reflection of

complex system-level dynamics. These include the interactions between socio-economic conditions, education systems, digital infrastructure, labor market access, and digital risk behaviors. This study adopts a systems thinking perspective to explore digital exclusion as an emergent outcome of multiple interrelated subsystems. By modeling these interdependencies—through Principal Component Analysis (PCA), logistic regression, and spatial analysis—we aim to understand how digital skills, behaviors, and structural variables jointly influence NEET vulnerability across European regions.

There are a number of definitions regarding digital competence that delimit it as the ability to explore and act flexibly in situations involving new technologies, to exploit the potential of technology to solve problems, and to generate and distribute new knowledge [1]. Essential digital competencies at the beginning of the century included information literacy, communication, collaboration, critical thinking, creativity, and problem-solving [2]. The high number of technological innovations in recent years has forced the labor market to update its skills to adapt to new requirements [3]. Also, the need to develop appropriate educational solutions has led to the inclusion of artificial intelligence tools that significantly contribute to the assessment of student competence development more flexibly and efficiently [4]. Currently, these are increasingly oriented towards elements such as cybersecurity [5], data literacy [6], cloud computing and server administration [7], the use of digital and artificial intelligence tools [8], and more.

In an era full of opportunities generated by digital technological developments, the internet offers both unprecedented learning situations and challenges regarding its risky use. Digital natives, namely the current generations of adolescents, are the largest consumers of information compared with previous generations, making them vulnerable to various forms of inappropriate behavior in the online environment and activities lacking or with a low level of ethics, a fact that is especially accentuated among those from families with low socio-economic status [9]. In this sense, Netiquette, a hybrid word combining “network” and “etiquette”, outlines the social code of the internet intended to reduce activities that involve, among others, the expression of personal data, harassment, pranks, and hate speech, as well as accessing inappropriate content [10].

Less defined by researchers, in our opinion, risky internet use (RIU) refers to any behavior adopted by children or adolescents during their internet usage that endangers them to face psychological, social, legal, or even physical harm. They are mostly related to the absence of digital literacy, emotional control, or an awareness of consequences, which is enhanced by the features of youth development, including impulsivity, curiosity, the exploration of identity, and peer acceptance. Mainly discussed is routine risky internet use among young people, identified as viewing/exposure to unwarranted materials (e.g., violence, pornography, or pro-anoxia websites); cyberbullying (victim, perpetrator, or bystander) [11]; infringement of privacy such as the publication of personal information or passcodes; excessive or loss of control on the use of more and more websites or compulsive use more often associated with gaming, social media, or streaming [12,13]; meeting and developing relations with questionable people; sexting and sharing explicit images; involvement in offending behaviors like hacking or piracy [14]; and so on. This behavior involves a large scale of digital skills, which does not necessarily mean it is intentional; rather, it is usually accidental due to a lack of awareness or control. In our opinion, digital skills and internet usage are employment advantages that factors at the individual, familial, and environmental levels can influence. Therefore, they are essential for prevention and education at various levels (e.g., in schools, families, and policy).

This system’s framing is particularly relevant given the multi-layered nature of digital exclusion, which spans individual, institutional, and structural levels. Influences such as digital skill proficiency, exposure to online risks, access to technology, and socio-economic

constraints operate not in isolation but in dynamic interaction. By adopting a systems lens, this study recognizes that small changes in one part of the system (e.g., improving digital skills) can have amplified effects across other components (e.g., employment outcomes or social participation). This approach moves beyond reductionist views and positions digital exclusion as an emergent property of a broader socio-technical ecosystem, shaped by feedback loops, path dependencies, and uneven regional development.

2. Scientific Background

2.1. Digital Skills and Internet Use

Navigating the digital age requires digital skills, critical thinking, and the responsible use of technology, given that it has a significant impact on society, the economy, and culture, revolutionizing human learning, work, and leisure behaviors [15]. Digital skills have a dual nature in shaping adolescents' online experiences, serving both as a protection against harmful content and as a facilitator of access to it [16].

There is also a significant gap between perceived and actual digital skills among young people [17]. Socio-demographic factors, such as gender, age, and education, play an important role in the self-perception of digital skills among young people, with studies showing that while girls and boys have similar digital skills, girls tend to have a lower self-perception of digital skills compared with boys, and this perception decreases as they become older [18]. However, the perception of the level of digital performance is partly connected to socio-demographic factors, being also mediated by socio-economic factors and cultural models acquired during the processes of acquiring and using technology [19].

On the other hand, young people cannot develop digital skills on their own without adequate access to technology and support networks, and this digital exclusion limits their opportunities and accentuates socio-economic vulnerabilities, highlighting the need for interventions to reduce inequalities [20]. In addition, the excessive use of the internet, including compulsive behavior and addiction, can contribute to a lack of moral engagement among young people, affecting their ethical structure and increasing their vulnerability in the digital environment [21].

2.2. Internet Use and Cognitive Effects

Excessive internet use has been associated with negative effects on cognitive development, including cognitive distortions and internalizing symptoms, which are characterized by significant changes in behavior and brain function [22]. This is associated with significant changes in the activity and structure of the prefrontal cortex, affecting executive functions such as self-control, impulsivity, and decision making, especially among university students who are considered a high-risk category due to them being exposed to the vulnerabilities associated with cybersecurity [23].

Specifically, despite the benefits such as easy communication regardless of distance and supporting the learning process, the excessive use of digital media influences the brain, cognition, and human behavior, generating addiction, changes in language development, and alterations in the processing of emotional signals [24]. Every human digital activity performed regularly, including the use of smartphones, social networks, video games, chatbots, and other artificial intelligence models, could leave an imprint on the brain due to its neuroplasticity, depending on the type and time of exposure. Some of the visible effects in this regard refer to the reorganization of the somatosensory cortex as a result of touchscreen use; the inhibition of cognitive functions by suppressing melatonin production by the blue light emitted by digital screens; attentional overload triggered by a fear of missing out and decreased long-term memory; substantial decrease in hippocampal-dependent spatial memory as a result of the use of GPS-type systems; reductions in the volume of

gray matter in the orbitofrontal cortex, an area involved in impulse control and decision making processes, as a result of excessive involvement in video games; and reduced social skills and the ability to recognize facial emotions, distorting perception and reinforcing prejudices [25].

2.3. Digital Skills and Employability

The importance of digital skills is steadily increasing as studies demonstrate a statistically significant correlation between digital skills and employment rates in the EU [26]. Basic skills also mediate the relationship between digital literacy and perceived employability among young people, which is influenced by participation in employment support programs [27].

The shift from conventional business models to digital business has forced the corporate world to reorient the relationship between digital skills and employability. The generic digital competencies demanded by the labor market are communication, teamwork, problem-solving, creativity and innovation, self-management, and learning [28]. On the other hand, the digital approach, digital content creation, digital content management, digital security, and digital empathy items vary significantly across different demographic groups [29].

In the field of marketing, which has a strong digital component, a number of transferable skills needed among young people are emerging, such as motivation, oral communication, presentation skills, interpersonal skills, flexibility, teamwork, stress resilience, problem-solving, creative thinking, good analytical and conceptual skills, and a knowledge of social media, e-commerce, mobile phones, the internet, and software [30]. Although employers claim that digital skills influence their hiring decisions for young graduates, there are significant gaps between the actual level of skills and their demand in the labor market [31]. This creates pressure on universities to adopt curricula so as to prepare graduates for the demands of the labor market in the context of the increasing demand for AI literacy [32].

2.4. Ethics and Integrity in the Use of the Internet

Ethical considerations are often seen as a limitation of new technologies, and less as an enabler to stimulate and guide. With the development of technological capabilities to record, process, and archive large amounts of data, ethical concerns have arisen, particularly with regard to sensitive, personal information, the misuse of which may violate citizens' rights [33]. Moreover, the lack of a unified legislative framework in the emerging technological field is entertaining skepticism among specialists on issues such as data bias, transparency, privacy, and copyright [34].

The use of the internet and the technologies accessible through it has raised questions about the lack of accountability and autonomous decisions [35]. Furthermore, the growth of generative artificial intelligence has come with new challenges regarding the integrity with which data is used [36], leaving users more vulnerable to misinformation [37] and internet scams. Internet scams can be the phishing type and involve tricking people into submitting personal financial information, such as passwords and credit card details, through fake websites [38]; cryptocurrency scams that promise victims unreal amounts of money by luring them into making fictitious investments [39]; malware and ransomware that rely on the use of malicious software to steal data or access for which they demand ransoms [40]; or advance fee scams in which users are persuaded to pay upfront for goods and services they will never receive [41].

The European Union has an active role in establishing a framework for new technologies, establishing, among others, in 2019, the Ethics Guidelines for Trustworthy AI, which

guide readers towards lawful (laws and regulations), ethical (ethical principles and values), and robust (blending technical and social) AI [42]. However, there is a need to develop a unified ethical framework in collaboration with businesses, governments, stakeholders, and researchers to balance technological innovation with ethical responsibility [43]. Also, given the transnational nature of cybercrime, international cooperation is also essential in this area and still insufficiently regulated [44].

2.5. Risky Internet Use and NEETs

In the context of the increasing use of digital technology, this study aims to analyze the relationship between digital behaviors and the likelihood of becoming NEET (young people who are not in employment, education, or training). Our research contributes to the understanding of the cognitive and social factors that influence educational inclusion, thus aligning with the directions proposed by this Special Issue, by integrating empirical analyses and educational strategies based on neuroscience.

The specialized literature suggests that the excessive use of the internet and social networks can negatively affect the development of young people, having an impact on social inclusion and the labor market [45,46]. The lack of digital skills is also considered a determining factor in economic and educational exclusion [47]. Recent studies have shown that exposure to digital risks can affect confidence in using the internet for educational and professional purposes [48–50]. These aspects highlight the need for a detailed analysis of the relationship between digital skills, internet use, and the NEET rate.

Despite numerous studies on NEET (Not in Education, Employment, or Training) youths in Romania, there are still insufficiently explored aspects. For example, although risk factors have been identified at the macro level, such as poorly connected educational systems to the labor market and labor market rigidity [51,52], there is a need for in-depth analyses at the micro level, examining the influence of individual and environmental factors on the likelihood of becoming NEET. Moreover, regional differences and the impact of rural versus urban environments on the NEET rate require further investigation, given that in rural areas the NEET rate for the 25–29 age group is 32.5%, compared with 10.1% in cities [53].

The widespread use of the internet, especially by young people, has many consequences, such as health issues, physical effects, and social or economic aspects. The excessive use of the internet and the development of high-quality digital skills could drive, in our opinion, in two directions: to a better rate of employment or social exclusion. The relationship between the risky use of the internet and employment was less explored in the literature, so the proposed research question was as follows: how do digital skills and risky internet use influence the likelihood of young people being classified as NEET in Europe, considering regional differences and socio-economic factors?

The stated hypotheses are as follows:

- H1.** *Excessive time spent online, especially on social media, is associated with an increased risk of young people being NEET, but the effects vary by educational and economic context.*
- H2.** *Low levels of digital skills increase the risk of exclusion from education and the labor market, influencing the NEET rate, with significant regional variations.*
- H3.** *Lack of online privacy practices and exposure to digital risks are correlated with a higher likelihood of being NEET, indicating a digital vulnerability that may contribute to socio-economic exclusion.*

H4. *Digital factors, including skills, internet usage patterns, and data protection measures, influence the likelihood of being NEET in a complex way, and multivariate analysis can highlight interactions between these factors.*

Variables such as the use of social networks, instant messaging services, and daily internet usage are applied in this study. These were chosen as proxy measures of high digital engagement because previous studies have focused on behavioral overuse instead of absolute time measurements [54,55]. We operationalize “excessive time online” not by giving a specific numerical limit (e.g., hours per day), but by meeting a recurrent and high-frequency involvement in digital applications, particularly those that include recognized addictive tendencies, such as social media and instant messaging [56]. Despite these being somewhat limited by the unavailability of harmonized EU-wide time use data, these indicators suggest similar patterns of intense use across countries.

Regarding “low levels of digital skills”, this study considered the EUROSTAT definition for the DESI that considers five areas (information and data literacy, communication and collaboration, digital content creation, safety, problem solving) and 21 competences, setting up four levels: basic, intermediate, advanced, and expert. We considered that low-level digital skills cover the area below basic, part of basic, up to intermediate, more precisely and the study analysis used indicators for no skills, limited, narrow, and low [57].

3. Systemic Analytical Framework and Methodology

This study employs a multi-method approach grounded in systems thinking, integrating statistical and spatial analysis to investigate how interconnected digital dimensions affect NEET vulnerability. The methodology is designed to uncover and model the systemic relationships among the variables that represent distinct yet interdependent components of the digital exclusion system.

The analysis follows a structured flow, combining statistical methods and spatial visualization techniques to understand the relationships between digital skills, internet use, and NEET status in Europe. The study is carried out in several stages: (1) descriptive statistics—an initial exploration of variables and an analysis of their distribution; (2) normality testing—checking the distribution of data to select appropriate methods (parametric or non-parametric); (3) Spearman’s correlations—identifying raw relationships between digital variables and NEET rate; (4) Principal Component Analysis (PCA)—reducing the dimensionality of data and extracting latent factors; (5) logistic regression—estimating the impact of digital factors on the probability of becoming NEET; (6) GIS analysis—the spatial visualization of results to identify regional patterns and digital disparities; (7) identifying implications—interpreting the results and formulating recommendations for public policies (Figure 1).

Through this integrated approach, this study offers a quantitative and spatial perspective on the socio-digital exclusion of young people, thereby contributing to the development of more effective intervention strategies. Moreover, the model highlighted the need for an integrated approach, offering insights into the alienation, social exclusion, and psychological issues, as well as the ethics and integrity in promoting the use of the internet.

This study analyzes the relationship between risky internet use and the likelihood of a young person becoming NEET. The data comes from Eurostat and covers 2000–2024, including 24 variables grouped into eight dimensions, for 28 countries.

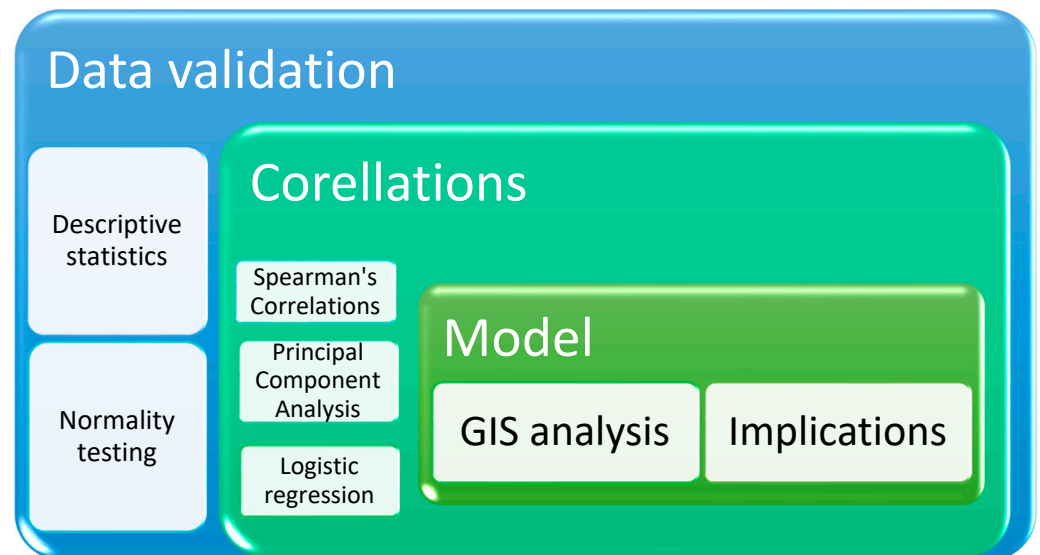


Figure 1. Research framework and model construct. Source: Authors' concept.

For the present analysis, we have opted for the standard definition used at the European level, which defines the NEET group as young people aged between 15 and 24, according to the Eurostat methodology. This allows for comparability between countries and highlights structural differences regarding the inclusion of young people in the labor market and the education system. The formula for calculating the NEET rate is defined as follows:

$$NEET\ rate = \frac{Number\ of\ youth\ (NEP + NoFE + NoNFE)}{Total\ population\ of\ youth\ (Y15 - 24)} \times 100$$

where

NEP = not employed persons;

NoFE = not included in formal education or training;

NoNFE = not included in non-formal education or training;

Y15-24 = total population aged 15–24.

3.1. Variables and Data Sources

Although Eurostat provides data on a longitudinal level for the period 2000–2024, we focus on the time frame 2020–2023 for our analysis. The choice was based on the availability of harmonized and coherent indicators in all 28 European countries that were considered in this study. The country year is the unit of analysis, and all the data available are secondary, aggregated indicators taken from Eurostat and the Digital Economy and Society Index (DESI). We do not apply data at the individual level in countries because, regarding this list of variables, such microdata is not publicly available in a comparable form.

To enhance transparency, we also mention that the data are country-level statistics, which include digital behavior and the occurrence of NEET. They are the variables that involve daily internet use, activity on social media and messaging systems, knowledge about personal data protection, and digital skills levels. Unit measurement and code are defined formally, based on Eurostat coding lists, and Eurostat metadata standards [58] guarantee definition and temporal consistency.

Following suggestions about multivariate quantitative research [59], the general descriptive statistics of each variable are presented in Section 3.2 and in Appendix A as the mean, standard deviation, minimum, and maximum values. This improves replicability

and provides an empirical perspective based on the ground empirical aspect of digital inequality as well as NEET risk distributions in Europe.

The variables considered for the model are presented in Table 1 and detailed in Appendix A (Table A1).

Table 1. Variables considered for the model.

No.	Indicator	Variable Code for the Model	Eurostat Code
<i>Dependent variable</i>			
0	NEET Rate	NEETYYYY	edat_lfse_20
<i>Independent variables</i>			
1	Social Networks	SNETYYYY	isoc_ci_ac_i
2	Instant Messaging	CHATYYYY	isoc_ci_ac_i
3	Daily Internet Use	IDayYYYY	isoc_ci_ifp_fu
4	Personal Data Protection	IMAPYYYY	isoc_cisci_prv20
5	Low Digital Skills	I_DSK2_YYYY_LW	isoc_sk_dskl_i21
6	Narrow Digital Skills	I_DSK2_YYYY_N	isoc_sk_dskl_i21
7	Limited Digital Skills	I_DSK2_YYYY_LM	isoc_sk_dskl_i21
8	No Digital Skills	I_DSK2_YYYY_X	isoc_sk_dskl_i21

Source: Authors' selection and construction.

This set provides a clear and logical structuring of the variables used to examine the relationship between internet use and the likelihood of being classified as NEET. Their organization facilitates the understanding of the indicators, their temporal distribution, and relevant statistical characteristics.

The detailed description of the variables, specifying the Eurostat codings, labels used, and explanations of the indicators, establishes the conceptual basis of the analysis by presenting the definitions, sources, and codings of the indicators employed. The indicators are grouped according to the dimensions they measure: internet use for socializing (SNET), instant messaging (CHAT), daily internet use (IDay), personal data protection (IMAP), and digital skills (I_DSK2). We should note that YYYY represents the year in question. The data were extracted from the Eurostat [53] database through the mentioned codes (Table 1) to provide transparency and theoretical underpinning.

3.2. Descriptive Statistics

To better comprehend the country-level inequities and patterns of distribution, we have included a description of values and statistics in the extended study in Appendix A. The tables contain detailed pointers of all six dimensions of digital and social exclusion, 28 countries in the European region (NUTS 0 level), and the specifications are for the years 2020–2024. The findings validate and moderate the above observations. An example can be noted as the NEET rates were extremely polarizing since the kurtosis and skewness values were very high, especially in Southeastern Europe. In the meantime, digital behaviors like the use of instant messaging and social networks are highly homogeneous with negatively skewed distributions, meaning almost complete penetration. Very much differently, some indicators touch on digital skills and individual data security, which points to severe dispersion and asymmetry, indicating structural weakness in specific countries. Such a statistical investigation provides the empirical basis of the typologies and models created in the subsequent sections with an overwhelming force, which does not, however, conflict

with the logic of the latter. Quite the contrary, it reinforces the case for the use of customized policy reactions towards youth digital exclusion in various national settings.

The NEET rate recorded a steady decrease between 2020 and 2023, with the average national values decreasing from 10.73% in 2020 to 9.45% in 2023, which may suggest improvements in youth employability or participation in education and training. Romania, with a NEET rate of 16.5% in 2023, is among the countries with the highest values, while the Netherlands has the lowest values, recording only 3.3% in the same year. The distribution of the NEET rate is positively skewed, indicating that most countries have lower rates, while a few countries have much higher values.

The use of social networks and instant messaging remains at high levels in all countries but shows a slight decrease in 2024. The average values are high, exceeding 90% for most years. For example, Spain and France have social network usage rates of over 95% in 2023, while Romania and Bulgaria show lower values, around 89%. The distribution of these variables is slightly negatively skewed, suggesting that the use of social networks and instant messaging is almost universal in all the countries analyzed. This phenomenon may indicate a maturation of users or a change in digital consumption patterns.

Daily internet access is high, with an average of over 96% for the entire period analyzed. For example, Finland and Norway reach values of almost 100%, while countries such as Bulgaria or Romania are below the average, but with an increasing trend. The distribution of these data indicates extensive digital adoption, with a slight negative skew.

Personal data protection (IMAP) shows significant variability across countries, with a variation of over 50 percentage points between countries such as Norway (92.59%) and Serbia (34.02%). This may reflect differences in digital education and regulatory policies on cybersecurity. This variable follows a normal distribution, suggesting a relatively balanced adoption of data protection measures.

Digital skills are marked by significant disparities across countries. The average for Low Digital Skills in 2023 is 16.94%, but countries such as Serbia and Turkey exceed 40%, while Finland and the Netherlands have very low values, below 5%. For very limited digital skills (No Digital Skills), the average is 0.98%, but there are countries where this percentage is significantly higher, such as Romania, with almost 5%, while countries such as Denmark, Sweden, and Finland have values close to 0%. The asymmetric distribution of this indicator is relevant for the hypothesis regarding the impact of digital skills on NEET status.

As a conclusion, the descriptive analysis shows important differences between European countries in terms of the risk of digital exclusion and the NEET rate (Appendix A). Northern European countries have high levels of digital literacy and extensive internet use, while Southeastern European countries, such as Romania, Bulgaria, and Turkey, have higher NEET rates and lower digital skills.

3.3. Normality Tests

The distribution of the variables in the analysis was assessed using the Kolmogorov–Smirnov (K-S) and Shapiro–Wilk (S-W) tests, which are commonly used to test the normality of data [60]. These tests compare the distribution of the sample with the theoretical normal distribution and determine whether the differences are significant.

The interpretation criteria are as follows:

- If $p > 0.05$, the normality hypothesis is accepted, indicating that the variable follows a normal distribution [61].
- If $p < 0.05$, the normality hypothesis is rejected, suggesting that the distribution of the variable is significantly different from the normal one [62].

The results of the applied normality tests are presented in Table 2.

Table 2. Tests of normality.

Indicator	Kolmogorov–Smirnov ^a			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
NEET2020	0.162	28	0.058	0.853	28	0.001
NEET2021	0.191	28	0.010	0.904	28	0.015
NEET2022	0.158	28	0.070	0.871	28	0.003
NEET2023	0.168	28	0.043	0.889	28	0.006
2020SNET	0.171	28	0.034	0.873	28	0.003
2020CHAT1	0.145	28	0.137	0.856	28	0.001
2021SNET	0.156	28	0.077	0.881	28	0.004
2021CHAT1	0.158	28	0.071	0.907	28	0.017
2023SNET	0.177	28	0.025	0.874	28	0.003
2023CHAT1	0.170	28	0.037	0.858	28	0.001
2024SNET	0.109	28	0.200 *	0.969	28	0.563
2024CHAT1	0.186	28	0.014	0.893	28	0.008
IDay2020	0.215	28	0.002	0.795	28	0.000
IDay2021	0.161	28	0.060	0.881	28	0.004
IDay2022	0.200	28	0.006	0.911	28	0.021
IDay2023	0.182	28	0.019	0.889	28	0.006
IDay2024	0.196	28	0.007	0.842	28	0.001
IMAP2020	0.111	28	0.200 *	0.948	28	0.178
IMAP2021	0.076	28	0.200 *	0.972	28	0.637
IMAP2023	0.098	28	0.200 *	0.977	28	0.779
I_DSK2_2021_LW	0.130	28	0.200 *	0.944	28	0.141
I_DSK2_2021_N	0.219	28	0.001	0.905	28	0.015
I_DSK2_2021_LM	0.193	28	0.009	0.820	28	0.000
I_DSK2_2021_X	0.254	28	0.000	0.700	28	0.000
I_DSK2_2023_LW	0.109	28	0.200 *	0.916	28	0.027
I_DSK2_2023_N	0.121	28	0.200 *	0.936	28	0.085
I_DSK2_2023_LM	0.145	28	0.135	0.894	28	0.008
I_DSK2_2023_X	0.222	28	0.001	0.776	28	0.000

Notes: * = lower bound of the true significance. ^a = Lilliefors Significance Correction; Source: research results.

According to Pallant [63], parametric tests such as the *t*-test, ANOVA, or Pearson’s correlation are adequate for variables with a normal distribution. For the variables without a normal distribution, non-parametric methods are more appropriate, including the Mann–Whitney U test, the Kruskal–Wallis test, and Spearman’s Rho coefficient for correlations [63,64]. In the context of this analysis, Spearman’s Rho is preferred for estimating correlations between risky internet use and the probability of being NEET, given that most of the variables analyzed do not meet the normality assumption [62]. This methodological approach ensures the robustness of the results and the validity of the inferences formulated in the study.

4. Exploring Correlation Methods

The analysis was performed using a combination of statistical and spatial methods, allowing the identification of relationships between internet use, digital skills, and the risk of being NEET. To explore the relationship between digital skills, internet use, data protec-

tion, and the likelihood of being NEET, we applied Spearman's Rho correlation, Principal Component Analysis, logistic regression, and GIS analysis for the spatial visualization of the results.

4.1. Spearman's Rho Correlations

In the analysis of the relationship between risky internet use, digital skills, and the likelihood of being NEET, we used Spearman's Rho correlation coefficient. This non-parametric test is appropriate for data that is not normally distributed, as previously verified by the Kolmogorov–Smirnov and Shapiro–Wilk tests. We also used a two-tailed significance test, which is appropriate when there is no clear hypothesis about the direction of the relationship between the variables [61].

Spearman's Rho (ρ) is a non-parametric measure of rank correlation that assesses the strength and direction of the association between two variables. Unlike Pearson's correlation, which measures linear relationships, Spearman's Rho evaluates monotonic relationships, meaning that as one variable increases, the other tends to increase or decrease consistently, without requiring a strict linear relationship.

4.2. Principal Component Analysis

PCA is used not only to reduce dimensionality but to identify latent subsystems—clusters of indicators that co-evolve within the broader digital exclusion system. This helps reveal the internal architecture of exclusion risks across EU countries.

PCA is used to identify the latent dimensions of internet use, thereby reducing the dimensionality of the dataset and grouping variables into relevant factors. Promax rotation, an oblique rotation method, is chosen due to its ability to allow for correlation between factors, which is recommended for socio-economic variables [65].

Data adequacy is tested using the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity (BTS) to determine whether PCA can be effectively applied to reduce the size of the dataset and extract significant factors. According to the literature, PCA is a frequently used technique for identifying latent structures in a large set of variables, facilitating interpretation and reducing redundancy [61]. The KMO test measures sample adequacy, and Bartlett's test verifies the existence of significant relationships between variables. If the values of these tests are acceptable, PCA can be used to reduce dimensionality and identify latent components [62].

If the extraction value is high (close to 1), the variable is well explained by the extracted factors; if it is low (below 0.5), it may require elimination or reconsideration [61]. Variance explanation is a crucial step in PCA, as it allows for the determination of the optimal number of components that preserve the relevant information of the dataset [66]. The goal of the analysis is to identify the minimum number of components that explain a sufficiently large percentage of the total variance. Component selection is based on eigenvalues, the Kaiser criterion (retention of components with eigenvalue > 1), and their interpretability [61].

The Scree Plot graph helps determine the optimal number of principal components to retain in the PCA model. The Scree Plot graphically represents the relationship between the number of components and the eigenvalue, providing a visual criterion for the selection of relevant components [61]. The purpose of using the Scree Plot is to identify the elbow point, where the decrease in eigenvalue becomes insignificant. According to the Kaiser criterion [67,68], only components with an eigenvalue > 1 should be retained, and the others are eliminated because their contribution to the total variance is negligible [66].

Two matrices are constructed for the factors extracted by PCA: the structure matrix and the correlation matrix. Unlike the structure matrix or the component matrix, the correlation matrix reflects the degree of association between the principal components, which is essen-

tial for validating the rotation method used [61]. This analysis aims to determine whether the extracted factors are independent or if there are significant correlations between them. If the correlation values are high, the use of an oblique rotation, such as Promax, which allows for the interdependence of the factors, is justified. Conversely, if the correlations are low, the use of an orthogonal rotation, such as Varimax, would be more appropriate [66].

Given the partial correlations between extracted components under Promax rotation, we complemented the analysis with an orthogonal rotation technique to test robustness and enhance interpretability.

To strengthen the main model by using Promax-rotated PCA, we also ran a Principal Component Analysis (PCA) that Varimax orthogonally rotated. The purpose of this step was to determine the independence and clarity of interpretations of the retained components.

The four rotated solutions established the existence of four conceptually homogeneous dimensions, which collectively accounted for 82.5 percent of the total variance, making it equivalent to the criterion (eigenvalue > 1) of Kaiser. The rotated component matrix (provided in Appendix B) separates two important behavioral domains into daily internet use, social networks, instant messaging, and digital skills, and strengthens the internal structure of the conceptual model. By normalizing the loadings on the components, also by minimizing cross-loadings, the Varimax rotation made the factors more empirically transparent. This further reinforces the methodological strength of the factor extraction method, demonstrating that not only the oblique method, but also the orthogonal procedure, can yield stable and interpretable structures.

Such a complementary analysis will help address potential issues regarding factor independence and confirm the usefulness of the dimensionality reduction method employed in the current research [69].

4.3. Logistic Regression

As this study uses aggregated country-level data rather than individual-level records, logistic regression was applied using the proportion of NEETs in each country year as the dependent variable, not a binary indicator. The model was implemented as a Generalized Linear Model (GLM) with a binomial family and logit link, consistent with the approaches used in ecological and cross-national research [70,71].

Logistic regression is employed to quantify the systemic influence of identified digital subsystems on the likelihood of NEET status, capturing how multiple, coexisting factors jointly affect the outcomes within the socio-technical system of youth inclusion.

Logistic regression is applied to estimate the probability of being NEET, using the factors extracted from the PCA as independent variables. This method is suitable for modeling the relationships between categorical variables and continuous predictors, allowing a better understanding of the risks of socio-economic exclusion. To prevent multicollinearity problems, the model is validated using the Nagelkerke R^2 and the Variance Inflation Factor (VIF) collinearity test.

Logistic regression analysis is a statistical method for examining the relationship between a binary dependent variable and one or more independent variables. This technique estimates the probability of an event occurring depending on the values of the explanatory variables.

Checking the data preprocessing is a step before interpreting the results, as a large number of excluded cases could affect the validity of the model [72]. The purpose of this analysis is to determine the total number of cases available for the logistic model and to check for missing data that could influence the validity of the results. Ensuring data integrity helps to obtain robust estimates and avoid systematic errors in the interpretation of regression coefficients [73,74].

The block 0 analysis of the logistic regression, which represents the initial model without predictors, is essential to evaluate the performance of the model with only the constant, before including the independent variables [72].

This determines the classification accuracy of the null model (the model that does not consider any predictors) and checks the importance of the independent variables before including them in the regression. It provides a baseline for evaluating how much the predictors improve the model's performance [73,75].

4.4. GIS Analysis Method

Geographic Information Systems (GIS) are used to spatially visualize the system-level disparities in digital capacity and exclusion risk across European territories. This spatial dimension highlights how regional contexts operate as subsystems with differentiated structural conditions and feedback loops.

Exploratory Spatial Data Analysis (ESDA) is used to identify geographical patterns in the variables under analysis. This includes methods such as choropleth maps, which visualize the distribution of NEET rates and digital skills at the regional level.

ESDA combines visual and statistical techniques to explore spatial data, helping to detect patterns and anomalies in the geographical distribution of the phenomena under study [76]. This methodological approach allows for a robust integration between statistical and spatial analysis, providing a multidimensional understanding of socio-digital exclusion.

4.5. CHAID Decision Tree Analysis

We used a CHAID (Chi-squared Automatic Interaction Detection) decision tree model to enhance the interpretability and determine nonlinear thresholds in the association between digital autonomy and NEET status. CHAID is a non-parametric technique for segmentation, applicable to the discovery of decision rules in categorical outcomes, as well as the detection of interactions among explanatory variables, without any assumption of linearity or normality [77,78]. The calibration of the model was performed with the binary NEET2023 variable status as a dependent variable and explanatory indicators collected during prior PCA and logistic regression procedures.

The project was analyzed using CHAID and IMAP2020, which is the percentage of persons able to control access to their personal information, and it was found to be the most relevant division. It was revealed that the risk of falling into NEET status increased by almost 20 times (76.5 percent, compared with 3.9 percent) among those countries whose indicator was exactly or less than the suggested threshold (67.13 percent), whereas the risk decreased by more than 80 times (only 9.1 percent, compared with 795.8 percent) in countries with a higher indicator. This observation validates the protective dimensions of digital agency and cybersecurity awareness of socio-economic exclusion among young citizens [79,80]. There was a statistically significant split as indicated by Bonferonni-adjusted Chi-square = 12.128, $p = 0.004$ and a global classification accuracy of 82.1% (Appendix C, which shows a complete overview of node splits, classification tables, and model accuracy measures).

The decision tree, therefore, supplements the analytic chain and the introduction of comprehensible thresholds, risk divisions, and an easier policy making outlook. While we use the latent dimensionality reduction in PCA or the probability estimation of logistic regression, the rule structure of CHAID also offers an alternative aspect that targets digital inclusion strategies clearly. In that way, it reinforces the conclusion that digital capability and data governance competency is a significant leverage in inclusive youth policy in a digital society [81].

While CHAID enhances the interpretability of the NEET model by identifying actionable thresholds and nonlinear interactions, it is not a substitute for more predictive or temporally dynamic approaches. Given the moderate explanatory power of the logistic regression model (Nagelkerke $R^2 = 0.339$), future extensions of this research should incorporate longitudinal modeling techniques—such as mixed-effects regression—or machine learning methods that can better capture complex interactions and improve classification accuracy [82,83]. These approaches offer promising avenues for refining risk prediction and tailoring policy responses across diverse temporal and geographic contexts.

5. Results

We will briefly present the results of the three methods for determining the existence of correlations between the studied variables to emphasize the results of the GIS spatial analysis, which offers a complex picture at the level of European countries.

5.1. Spearman's Rho

A preliminary analysis of the relationships between variables using the Spearman coefficient indicated that factors related to digital skills and data protection are significantly correlated with the likelihood of being NEET, which justifies their inclusion in the predictive models. The Spearman correlation analysis suggests that a lack of digital skills and risky internet use are significantly correlated with the likelihood of being NEET ($p < 0.05$). These findings support the use of PCA to reduce dimensionality and select the most important factors for logistic regression.

The use of social media and instant messaging: There is a significant positive correlation between the NEET rate and the use of social media in 2020 ($\rho = 0.412$, $p = 0.029$) and the use of instant messaging in 2021 ($\rho = 0.367$, $p = 0.042$). These results suggest that a higher frequency of use of these platforms could be associated with an increased likelihood of being NEET, supporting the hypothesis that the excessive use of the internet could have negative effects on the employability of young people [84].

Data protection and NEET: The correlation analysis revealed a negative correlation between data protection and the NEET rate in 2023 ($\rho = -0.502$, $p = 0.008$). This result suggests that young people who are more aware and cautious in managing their personal data are less likely to become NEET [46].

Digital skills and NEET: The variable “Low Digital Skills” in 2021 had a significant positive correlation with the NEET rate ($\rho = 0.478$, $p = 0.011$), indicating that the lack of basic digital skills increases the risk of exclusion from the labor market and the education system [85]. Also, “No Digital Skills” in 2023 showed a strong positive correlation with the NEET rate ($\rho = 0.536$, $p = 0.004$), confirming the hypothesis that the total lack of digital skills is a predictor of social exclusion.

These results support the hypothesis that excessive internet use, exposure to digital risks, and a lack of digital skills contribute to an increased likelihood of being NEET. In particular, the lack of digital skills appears to be a strong determining factor, suggesting the need for educational policies that support the development of digital competences among young people.

5.2. PCA and PROMAX Rotation Results

To check whether PCA is appropriate for the dataset, the Kaiser–Meyer–Olkin (KMO) and Bartlett tests were used. The KMO result is 0.674, which suggests an acceptable fit for PCA. This value shows an acceptable level of correlation between variables, according to the methodological standards established in the exploratory factor analysis [68].

The Bartlett's test results $\chi^2(276) = 880.306, p < 0.001$ confirm the existence of significant correlations between variables and justify the use of PCA. Bartlett's test confirms that the correlation matrix is not an identity matrix and that there are significant relationships between variables [86,87]. Since $p < 0.05$, the validity of using PCA for this dataset is confirmed.

The results of the KMO and Bartlett tests support the application of PCA with valid results, allowing the reduction in the dataset size and the extraction of significant factors. Subsequent analysis of the factor loadings will facilitate the identification of latent dimensions that explain the variation in the data.

The PCA results indicate that the first four components together explain 82.55% of the total variance of the dataset. The first principal component (PC1) accounts for 56.78% of the variance, suggesting a dominant factor in structuring the data. The following components (PC2, PC3, and PC4) add 11.77%, 8.51%, and 5.49%, confirming that these four components are sufficient to account for the variability of the dataset.

After PROMAX rotation, the factor loadings are redistributed among the components, which facilitates their interpretation, but the values can no longer be summed directly for the total variance explained. Subsequent components have eigenvalues below 1 and explain less than 5% of the variance of the dataset, indicating that they do not contribute significantly to the model and can be removed from the analysis [62].

The Scree Plot (Figure 2) confirms the selection of four components in the PCA model, as they capture the main structure of the data. The choice of this number of factors is supported by both the Kaiser method and the turning point visible in the plot. Retaining more components would lead to overinterpretation, while eliminating important factors could lead to the loss of significant information.

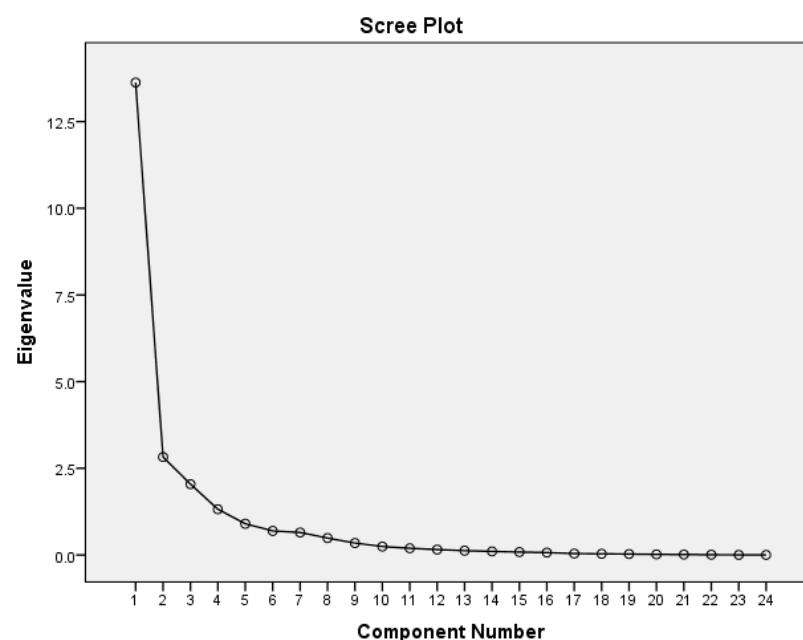


Figure 2. Scree Plot—determining the optimal number of components. Source: author's research results.

The four variables extracted by the PCA with Promax rotation are as follows:

- PC1: Intensive daily use of the internet (IDayYYYY, SNETYYYY).
- PC2: Lack of digital skills (I_DSK2_YYYY_X, I_DSK2_YYYY_LM).
- PC3: Exposure to risks regarding personal data (IMAPYYYY).
- PC4: Excessive use of social media and instant messaging (CHATYYYYY).

The correlation matrix for the extracted values is presented below in Table 3.

Table 3. Component correlation matrix.

Component	1	2	3	4
1	1.000	0.546	0.620	0.468
2	0.546	1.000	0.525	0.324
3	0.620	0.525	1.000	0.461
4	0.468	0.324	0.461	1.000

Rotation method: Promax with Kaiser extraction for PCA normalization. Source: research results.

The correlation between C1 and C3 is 0.620, indicating a significant relationship between these factors. Similarly, the correlations between C1 and C2 (0.546) and between C3 and C2 (0.525) are high enough to indicate a relationship between these factors, but not so high as to suggest redundancy between the components [62]. Moderate correlations between the components suggest that they are not completely independent, but that there is some overlap in the explained variability. This is typical in PCA applied to complex datasets, where latent factors may have common influences on the observed variables.

The matrix of correlations between components confirms that the Promax rotation was an appropriate choice, as it allows the factors to be correlated in a realistic manner, reflecting the latent structure of the dataset. Moderate correlations suggest that each component explains a distinct aspect of the data, but without being completely independent of the others. This result supports the validity of the factor model and provides a solid basis for interpreting the relationships between the original variables and the extracted factors.

5.3. Logistical Regresion

The results indicate that all 28 cases were included in the analysis, with no missing values (missing cases = 0). This means that the logistic model uses 100% of the available observations, which guarantees good representativeness of the data. The absence of excluded cases is a positive indicator since incomplete models or missing values could reduce the precision of the estimates and introduce possible biases into the model [88]. Summarizing the processing of cases confirms that the logistic model uses the entire dataset without exclusions or missing values. This result suggests that no additional data treatment is necessary before interpreting the model results.

The Hosmer–Lemeshow test suggests that the model fits the data well. The accuracy of the model is moderate (67.9%), but better for class 0 (71.4%) than for class 1 (64.3%). The only variable with a trend towards statistical significance is FAC2_1 ($p = 0.084$), which may have a negative impact on the probability of being in category 1.

5.4. GIS Analysis

This section analyzes the estimated probabilities (PRE_1), standardized residuals (ZRE_1), and influential values (LEV_1) resulting from the logistic regression model. These values are essential for validating the model and integrating it into GIS analyses, allowing the visualization of the spatial distribution of probabilities and the detection of outliers [72].

The purpose of this analysis is to verify the accuracy of the logistic model and identify observations that have a significant impact on the results. By examining the estimated probabilities and standardized residuals, we can evaluate the performance of the model and identify areas where predictions are underestimated or overestimated [73].

Figure 3 presents the estimated probabilities to be classified as NEET.

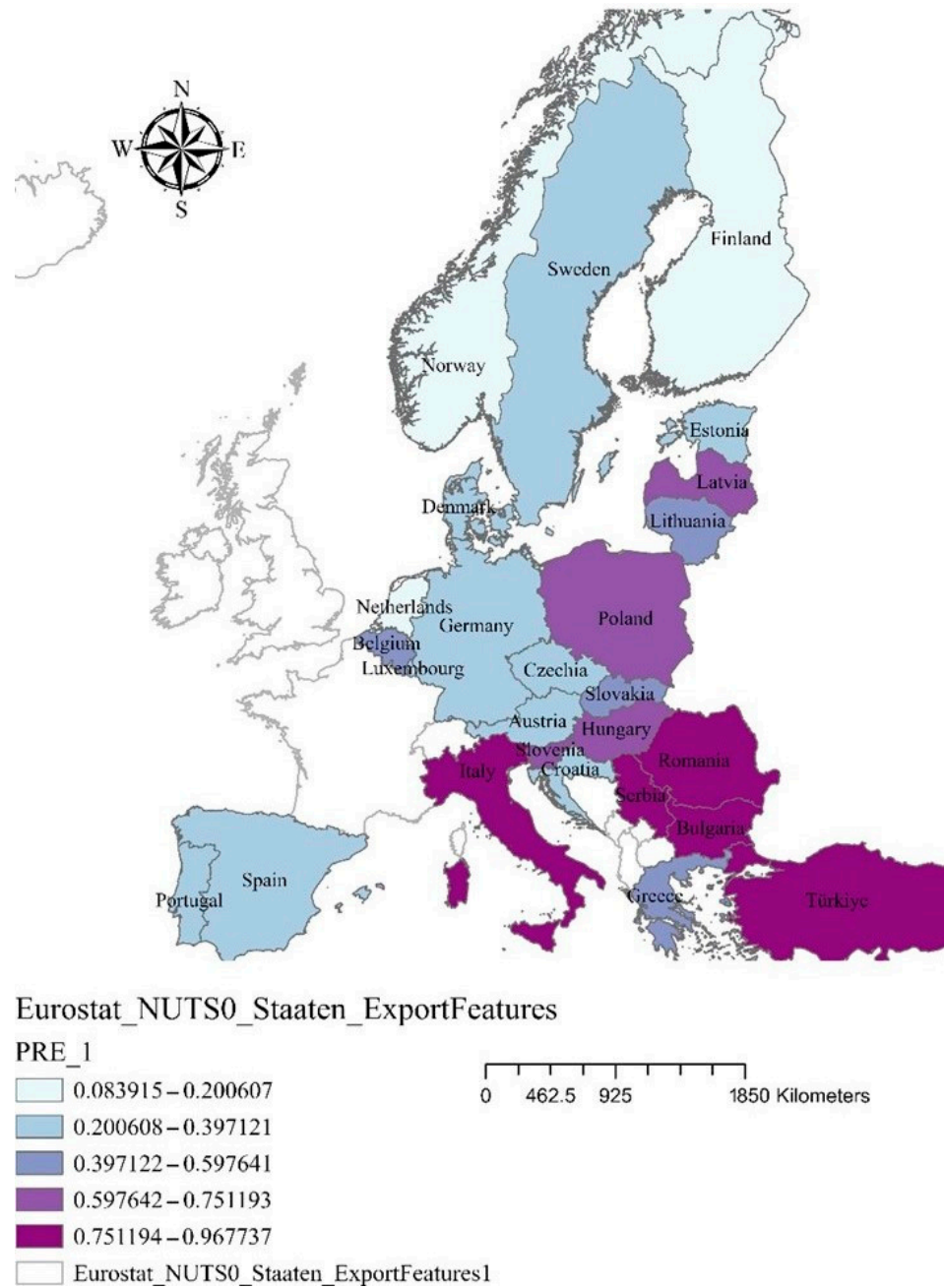


Figure 3. NEET estimated probability. Source: research results from ArcGIS extraction. Data source: Eurostat [53].

The estimated probabilities show the potential to be classified as NEET if the value is “1” and, on the other hand, if the value is “0”. According to the results, the lowest value is 0.0839, reflecting a low probability of being included in the NEET category, and the highest value is 0.9677, which indicates a consistent probability of being classified as NEET.

The standardized residual (Figure 4) values indicate the discrepancies between the observed and predicted values. Large residuals (positive or negative) may suggest outliers or systematic errors in the model.

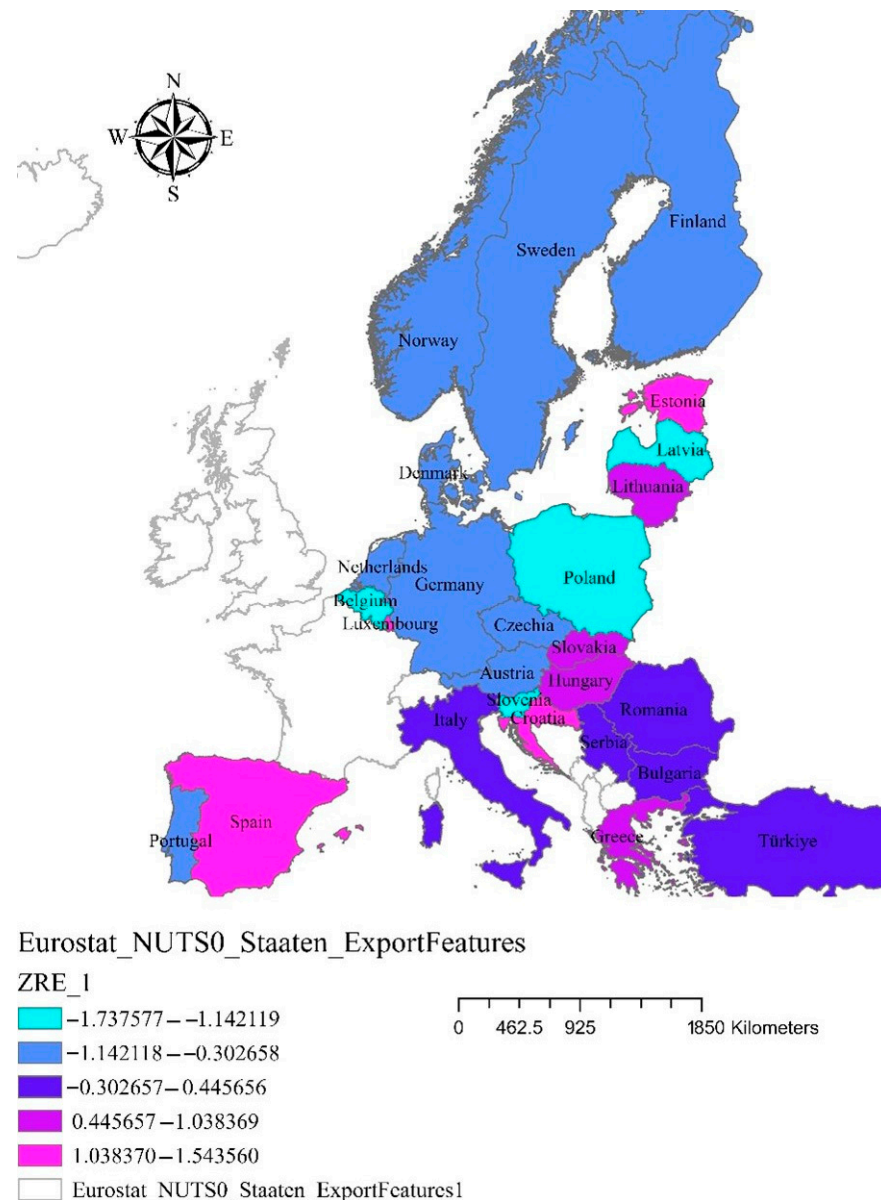


Figure 4. Standardized residuals. Source: research results from ArcGIS extraction. Data source: Eurostat [53].

The positive standardized residuals show that the model underestimated the probability of this observation, while a negative value expresses an overestimation of the probability of this observation. The values for our model are between -1.7375 and 1.5435 , which falls well within the acceptable range (-2 to $+2$) based on ArcGIS standard criteria [89]. This suggests that the model fits the data well, with no extreme outliers. At the same time, there is no strong spatial autocorrelation in the residuals, meaning that the model does not systematically overestimate or underestimate in specific regions. Moreover, the predictors used appropriately capture the variance in the dependent variable.

Influential values or leverage values measure how much an observation influences the estimated regression coefficients. A high leverage ($>2 \times$ average leverage value) indicates an unusual observation that might significantly affect the model predictions. If the leverage is high but the residuals are small, the point is influential but still fits the model well. The results are presented in Figure 5.

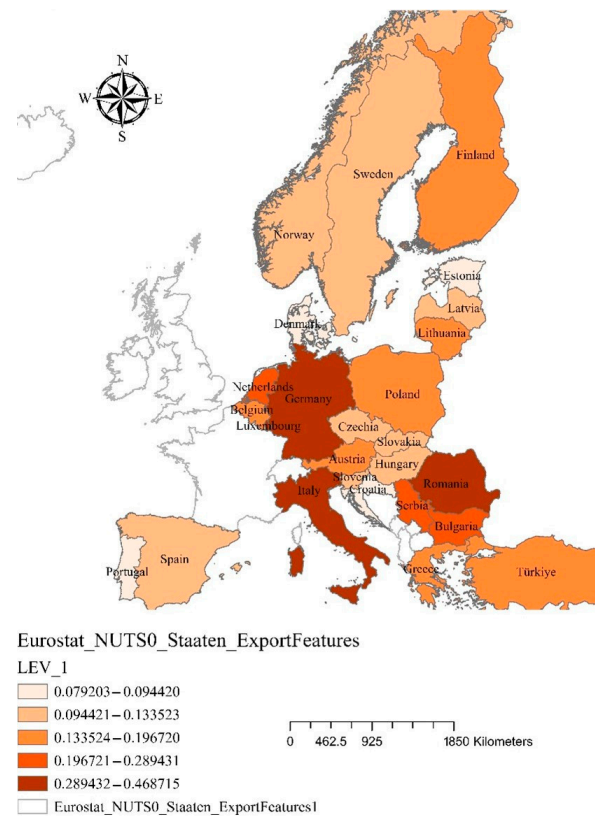


Figure 5. Influential values. Source: research results from ArcGIS extraction. Data source: Eurostat [53].

As can be seen, the leverage values are between 0.0792 and 0.4687, with an average value of 0.3571 (4 variables + 1/28 observation = 0.1785).

The formula for the average leverage value is as follows:

$$\text{Average Leverage} = (k + 1)/n$$

where

- k = number of predictor variables;
- n = total number of observations.

Since leverage values above 0.5 or closer to 1 are considered highly influential, the maximum value (0.4687) suggests that no extremely influential points exist, but a few observations might still deserve further inspection. Most data points have reasonable leverage values (0.0792–0.4687). There are no extreme high-leverage points (close to 1), meaning no single observation is distorting the model drastically. A few moderate-to-high leverage points may exist, but they are within an acceptable range, and they should be discussed in relation to the residual values.

5.5. CHAID Decision Tree Results: Objectives, Outputs, and Interpretations

In order to further examine the factors that lead to NEET status, we utilized a CHAID (Chi-squared Automatic Interaction Detector) classification tree analysis via SPSS v30. This approach was selected due to the possibility of handling categorical predictors, as well as exploring complex and nonlinear interrelations among the variables to provide a hierarchical structure. The CHAID algorithm can help us determine the decision rules and hierarchically defined segmentations related to the assumption of being NEET, depending on digital skills, the use of the internet, and data protection practices.

The CHAID findings have been organized into seven major outputs, each representing one stage of the analysis process, as summarized in Table 4. The outputs are further ex-

plained in Appendix C and allow for validation, supplementing the results of the Spearman correlations, PCA, logistic regressions, and GIS analysis performed in Sections 5.1–5.4.

Table 4. Synthesis of CHAID-based decision tree outputs: analytical objectives, key findings, and interpretive insights.

CHAID Output Component	Objective	Results	Interpretation
1 Model Summary (Table 1)	To determine the explanatory power of the CHAID model	Risk estimate: 0.178; Classification accuracy: 82.1%	The model has good predictive performance, suggesting strong associations between predictors and NEET status
2 Decision Tree Output (Figure 1)	To visualize the segmentation of NEET risk based on predictor variables	Tree depth = 3; Terminal nodes = 8	Highlights complex interaction patterns among digital skills and behaviors influencing NEET risk
3 Node Statistics (Table 2)	To assess the significance and strength of splits at each node	χ^2 tests significant at $p < 0.05$; strong parent–child split logic	Each split reflects statistically meaningful differentiation in NEET probability
4 Risk Estimates	To evaluate the robustness and generalizability of the model	SE of risk = 0.077	Indicates model reliability and low overfitting
5 Classification Accuracy	To quantify correct vs. incorrect classifications	Class 0 (non-NEET) correctly classified: 84.2%; Class 1 (NEET): 77.8%	CHAID identifies both NEET and non-NEET categories with high precision
6 Rule-Based Syntax (SPSS)	To extract applicable decision rules for NEET prediction	IF Low Digital Skills AND High Internet Use THEN High NEET Probability	Enables the operationalization of decision rules in targeted policy or education programs
7 General Conclusions	To synthesize key decision rules and implications	Lack of digital skills and risky online behavior = strongest NEET predictors	Supports prior findings from PCA and logistic regression; reinforces the need for targeted upskilling policies

Source: Authors' synthesis based on the CHAID results presented in Appendix C.

These findings validate that the inability to possess digital skills and an intensive, yet disorganized, engagement with the internet are the primary factors that lead to NEET risk, which aligns entirely with the results of the prior analytical phases (Sections 5.1–5.4). The decision tree model fills the gap of traditional regression-based analysis because it provides rules with understandable interpretations, which can be converted to intervention practices. CHAID models are uniquely positioned to reveal hierarchical predictors and viable groupings, as demonstrated earlier by [77] and later echoed by other channels in the sphere of socio-economic profiling [90,91].

The CHAID model identified a crucial digital cut point of strong explanatory power regarding NEET vulnerability. In particular, the IMAP2020 variable, reflecting third-level digital proficiency, including the ability to manage and protect personal digital data, proved to be the most dominant predictor in a binary segmentation of countries. The model has recognized a cut point of 67.13 percent and divided the dataset into two clusters. It was observed that countries with a NEET prevalence of 67.13% or less, according to the IMAP2020, still had a very high NEET prevalence of 76.5%, whereas those above this value had a NEET rate of only 9.1%. Such unequivocal differentiation leads to the assumption that digital self-protection sensitization is a determinant in combating youth expulsion from employment, training, or school.

Cluster 1 includes Romania, Bulgaria, Italy, Greece, Spain, Croatia, Hungary, Slovakia, Poland, Portugal, Cyprus, Latvia, Lithuania, Malta, Slovenia, Czechia, and Estonia, which

have limited capacities of digital self-control. As compared with the above, Cluster 2 contains Sweden, Finland, Denmark, Netherlands, Germany, Austria, France, Belgium, Ireland, Luxembourg, and Norway, which exceed the IMAP2020 standard [92].

These results suggest that addressing policy targeting through a decision tree analysis is warranted and support the usefulness of digital autonomy as a strategic proxy for social resilience and labor market integration. The precision of the segmentation also enhances the interpretability of the model and its applicability in developing targeted interventions for digital inclusion [79–81].

6. Discussion

The interpretation of the results has been approached with caution, acknowledging the methodological constraints and data limitations highlighted throughout the study. This perspective ensures the robustness and reliability of the conclusions drawn.

The results of this study confirm that digital exclusion among the youth is best understood as the output of an interconnected system involving digital skills, behavioral risks, and socio-economic structures. Rather than treating variables as isolated predictors, the analysis models their interaction as part of a dynamic system shaping NEET vulnerability across regions. This systems-level perspective aligns with prior work emphasizing the importance of feedback loops, structural inequalities, and policy alignment in managing digital transformation outcomes [93–97]. This study confirms that a lack of digital skills and a lack of awareness of online risks contribute to the socio-economic vulnerability of young people. The results indicate a significant negative correlation between the degree of protection of personal data and the likelihood of being NEET ($\rho = -0.502$, $p = 0.008$), suggesting that young people who adopt digital security practices are less at risk of socio-economic exclusion.

This result is consistent with previous studies [45,49], which highlight that digital literacy and data protection are critical factors for active participation in education and the labor market. On the other hand, daily internet use and instant messaging were not significantly correlated with NEET status, suggesting that the mere use of technology is insufficient to prevent digital and professional exclusion.

The results also support the idea that digital exclusion is not only a technological problem but also an ethical and social one. This confirms the studies of Van Deursen & Helsper [79], which show that a lack of digital education can increase young people's exposure to manipulation, online fraud, and corrupt practices.

Developing educational policies that are not limited to increasing access to technology but also emphasize ethical digital education, cybersecurity, and the development of essential digital skills can be recommended. In addition, supporting public initiatives that promote digital security and online integrity can significantly contribute to reducing the risk of NEET and increasing educational and professional inclusion.

From a theoretical standpoint, this study contributes to the systems literature by operationalizing digital exclusion as a multi-level, emergent system. The use of PCA allows for the identification of latent subsystems (e.g., digital skills clusters, risky online behavior patterns), while logistic regression quantifies their combined influence on system outcomes. Spatial mapping further contextualizes these dynamics, revealing how structural conditions (e.g., regional economic development, access to infrastructure) interact with individual-level variables to shape systemic risk.

Practically, the framing of this system offers a diagnostic tool for policymakers. Rather than focusing on single indicators or interventions, the findings suggest that targeted improvements across key subsystems—such as digital literacy education, online safety programs, and broadband access—can shift system-level outcomes. The ability to identify

high-risk regional configurations through spatial analysis also supports the more efficient allocation of policy resources, especially within multi-level governance frameworks like the EU.

6.1. Strengths and Limitations of the Logistic Regression Model

While this study offers a systems-based approach to understanding youth digital exclusion in the EU, several limitations should be acknowledged.

Firstly, the analysis is constrained by the granularity and completeness of available data. While Eurostat provides harmonized indicators, some variables—such as digital skills or risk behaviors—are only available at a national (NUTS0) level and may obscure important intra-national differences. This limits the resolution of the system's spatial dynamics.

Secondly, the logistic regression model is static and based on cross-sectional data. While it reveals relationships between digital subsystems and NEET probability, it does not capture the temporal feedback, delays, or nonlinear shifts often present in complex systems. Future work should consider longitudinal datasets or system dynamics modeling to explore time-sensitive causal pathways and interdependencies.

However, the fact that there are various annual observations of each country between 2020 and 2023 in the dataset gives the possibility of panelizing it. Although we do not apply panel data modeling to our data analysis, in the future, a panel data modeling approach based on econometric methods like fixed- or random-effect modeling may be used to analyze lagging outcomes, feedback effects, and policy change. Such strategies would help achieve a superior understanding of how digital exclusion changes with time and what structural predisposition it has, enhancing the existing cross-sectional knowledge [98,99].

Thirdly, the system boundaries defined in this study are necessarily simplified. The components analyzed—digital skills, behavioral risks, and socio-economic access—represent only part of a wider system that also includes education policy, family structures, labor market demand, and institutional trust. Including these would require a more expansive, possibly multi-level modeling framework.

Fourthly, while PCA is useful for identifying latent subsystems, it assumes linear relationships and orthogonality, which may not fully reflect the real-world interconnections of system components. Although the Principal Component Analysis (PCA) was useful in reducing dimensionality and identifying the main digital components, it relies on linear assumptions that may not fully capture the complexity of digital behavior. To address this, future work will explore alternative approaches—such as the clustering techniques available in ArcGIS Pro—that are more suitable for detecting nonlinear patterns and territorial groupings. Alternative dimensionality reduction methods, such as t-SNE or nonlinear manifold learning, could provide complementary insights.

Finally, although spatial analysis adds contextual richness, GIS-based methods are descriptive and do not account for spatial autocorrelation or diffusion effects. More advanced spatial econometrics or agent-based modeling could further enhance the systemic understanding of digital exclusion patterns.

The model has a moderate R^2 (Nagelkerke $R^2 = 0.339$), indicating that it explains a significant (but not a majority) portion of the variation in the dependent variable. The Hosmer–Lemeshow test ($p = 0.366$) shows a good fit between the model predictions and the observed data. It correctly classifies 67.9% of the cases, which means that it is better than a completely random model (50%). FAC2_1 is almost significant ($p = 0.084$), suggesting a possible effect, but which should be verified with a larger sample.

The model's limitations stem from its overall significance, which is marginal ($p = 0.084$), indicating that the predictors do not significantly improve prediction compared

with a model without variables. The variables included are not significant enough—none have a $p < 0.05$, which weakens the power of the model. The confidence intervals for $\text{Exp}(B)$ are wide, suggesting high uncertainty in the effect estimates. The accuracy of the model is moderate—67.9% is not bad, but not very impressive, especially since the performance in category 1 (64.3%) is weaker.

Although the current analysis employs a cross-sectional logistic regression model based on PCA-derived factors, future research should consider implementing longitudinal panel models, such as Linear Mixed-Effects Regression, to better capture temporal dynamics and country-level heterogeneity. This extension would enhance the predictive accuracy and allow for more robust estimations of the systemic impact of digital variables on NEET vulnerability.

The research on the factors influencing NEET status has used various statistical and econometric methods to explain variations in this phenomenon. Our study applied logistic regression, supported by Principal Component Analysis (PCA) with Promax rotation, to examine the impact of digital skills and risky internet use on the probability of becoming NEET. Compared with other approaches in the literature, our methodology makes an important contribution by integrating GIS spatial analysis and using a multidimensional perspective on digital exclusion.

Most studies on NEET status use binary logistic regression or probit models to estimate risk factors. For example, Filandri et al. [51] analyzed data from several European countries and demonstrated that factors such as educational level and economic background have a significant impact on the probability of being NEET. Our study partially confirms these findings, showing that a lack of digital skills is a significant predictor of socio-economic exclusion. The study by Redmond, P., & McFadden, C. [52] highlights that educational level is the strongest determinant of NEET risk, and digital skills are rarely integrated into explanatory models. Our results complement this framework by introducing a digital factor, which specifically explains how digital security and personal data protection influence trust in using the internet for education and employment.

An emerging trend in recent research is the use of machine learning algorithms, such as Random Forest or Support Vector Machines (SVMs), to enhance the accuracy of NEET risk predictions [100,101]. These models have the advantage of capturing the complex interactions between variables and reducing the impact of classical statistical assumptions. In comparison, our logistic model has the advantage of interpretability, allowing for the clear identification of variables with significant influence. However, its accuracy (67.9%) is inferior to machine learning-based models, which have achieved accuracies of over 80% in some studies [102]. A future direction could be to combine logistic regression with machine learning methods to optimize the classification of individuals into the NEET group.

The dimension of digital ethics and privacy is insufficiently explored in existing models. The study by Livingstone & Helsper [45] indicates that young people who show a greater awareness of digital risks are more likely to use the internet productively. Our results support this hypothesis, demonstrating that digital security management (e.g., protecting personal data, avoiding online tracking) is negatively correlated with NEET status ($\rho = -0.502, p = 0.008$).

Furthermore, Helsper [103] emphasizes that digital exclusion is not only a technological problem but also an ethical and socio-cultural one. Thus, the integration of variables regarding risky internet use and privacy brings added value in explaining young people's digital vulnerabilities.

GIS integration highlighted clear territorial disparities in digital skills and NEET rates. The choropleth maps confirmed that regions with limited digital access also have high NEET rates, suggesting that interventions need to be adapted to the local context. This

approach provides a valuable tool for public policies, allowing resources to be targeted to regions at the highest risk of digital exclusion.

A key limitation of this study is the use of national-level (NUTS0) data, which, although harmonized and suitable for cross-country comparisons, may mask significant intra-national disparities. The sub-national variation in NEET prevalence and digital exclusion is likely to be substantial, especially in countries with pronounced regional inequalities. While the current approach enables a systems-level comparison across EU Member States, future research will aim to incorporate more granular data—at the NUTS2 or NUTS3 levels—where available, to improve spatial precision and inform localized policy responses. Eurostat’s Labour Force Survey and regional indicators related to youth inactivity, digital skills, and broadband access offer a promising basis for such analysis [104].

Besides the problem of data granularity, there is another limitation of space which deals with the non-existence of formal tests regarding spatial dependence. We have not formally tested to see whether spatial autocorrelation or spatial dependence influences are identified in this study. Now that visualization with the use of GIS tools brought to light the regional differences in NEET levels and digital abilities, we cannot ignore that none of the methods involving Moran I, spatial lag models, or Geographically Weighted Regression (GWR) were used because of the national level of data aggregation (NUTS0). Follow-up studies are to be sought to apply spatial econometrics to more detailed data on NEET (e.g., NUTS 2 or NUTS 3) with the occurrence of localized modeling of spatial variation and dependency in the incidences of youth exclusion patterns [105–107].

6.2. Practical Implications for Policy and Intervention

The results of the current analysis suggest some tangible policy directions for minimizing the risk of NEET through digital integration. The first is that the negative correlation between digital skills and the NEET rate is rather significant, which highlights the importance of improving national digital education efforts. Such initiatives are especially designed to include vulnerable youngsters, such as young dropouts, as well as those living in socio-economically underprivileged areas, and encompass not only entry-level but also mid-level expertise in ICT [108,109].

Second, the fact that the high-frequency utilization of social platforms is linked to NEET status highlights the importance of introducing the concept of cyber safety and digital wellness in schools and other youth educational outreach initiatives. Digital resilience could be enhanced by awareness campaigns dealing with such issues as online misinformation, addictive use patterns, and data protection [54,110].

Third, the spatial-level discrepancy in NEET rates implies that spatially based interventions, including mobile digital literacy units, subsidized vouchers for ICT training, and targeted NEET outreach teams, deserve priority. To align with the EU-wide goals of digital and social cohesion, these endeavors must be harmonized with both the European Youth Guarantee and the Digital Compass 2030 strategy [81].

All these recommendations together advocate a systemic, targeted approach to the idea of digital inclusion and youth engagement that goes beyond infrastructure and access, to capabilities, ethics, and digital agency.

7. Conclusions

The data analysis confirmed that low digital skills are the main factor contributing to the likelihood of a young person being classified as NEET, supporting hypothesis H2. The results indicate significant differences between regions, suggesting that digital vulnerability varies according to socio-economic context. A lack of digital skills significantly increases

the risk of being NEET, which is more pronounced in regions with reduced access to educational resources and economic opportunities.

Hypothesis H1 is partially confirmed: the excessive use of the internet and social networks is not a direct predictor of NEET status but may have an indirect effect by reducing the time allocated to educational activities or by affecting the cognitive skills necessary for insertion into the labor market. Excessive internet use does not have a direct impact on the likelihood of being NEET but may contribute indirectly by decreasing employability or negatively influencing educational processes. This confirms the other studies that are highlighting the benefits of digital skills and the threats as well [111,112].

Regarding hypothesis H3, exposure to the risks regarding the protection of personal data does not directly influence the likelihood of being NEET, but the identified correlations suggest that a lack of knowledge about online security may be associated with low digital skills, which, in turn, contributes to socio-economic exclusion. Exposure to data protection risks does not significantly influence NEET status, although it is correlated with low digital skills, which may have an indirect impact on socio-economic exclusion.

Hypothesis H4 is partially confirmed, as digital factors (including digital skills, internet use, and data protection) explain a significant part of the variation in the NEET rate, but not uniformly across all categories of young people and regions. The use of GIS revealed clear regional patterns, indicating significant digital and economic disparities between European countries. According to the GIS analysis, digital factors explain part of the variation in the NEET rate, but their impact is dependent on the socio-economic context and regional differences.

Several analytical levels indicate that hypothesis H4, which deals with the system's explanatory strength of digital factors by the NEET status, is true. The research, using Principal Component Analysis (PCA) and logistic regression, says that digital skills and the pattern of use explain a high proportion of NEET prevalence variations. In addition, the CHAID decision tree model provides additional clarification, as it has revealed a critical level at 67.13 percent regarding the IMAP2020 variable, which indicates the proportion of people who can deal with their digital information. Countries above or standing below the threshold form patterns known as high-risk profiles, as they report a NEET prevalence of 76.5%, whereas countries above the threshold report a prevalence of merely 9.1% NEET.

In this binary separation, two territorial groups are created as follows:

- Cluster 1 (IMAP2020 \leq 67.13%): Romania, Bulgaria, Italy, Greece, Spain, Croatia, Hungary, Slovakia, Poland, Portugal, Cyprus, Latvia, Lithuania, Malta, Slovenia, Czechia, and Estonia.
- Cluster 2 (IMAP2020 $>$ 67.13%): Sweden, Finland, Denmark, the Netherlands, Germany, Austria, France, Belgium, Ireland, Luxembourg, and Norway.

This finding proves that the level of digital security awareness is not only a technological indicator but a sign of social inclusion [79–81]. The interpretability and the actionability of the model are due to its clarity in segmentation in terms of CHAID, and its ability to target policies on the basis of digital thresholds, and profiles of vulnerability.

The GIS-based spatial analysis illustrates how digital exclusion is not just individual, but is also territorial, whereby there are regionalities of low digital capacities and NEET rates. This geographical evidence justifies the argument of tiny-scale interventions and the use of digital education and employability policies [113,114].

Thus, the answer to the research question is that low levels of digital skills significantly increase the likelihood of being NEET. In contrast, excessive internet use and exposure to digital risks have an indirect or limited effect on NEET status.

The integration of GIS methods allows the identification of territorial clusters with high NEET rates and low digital skills. The choropleth analysis highlights vulnerable

regions, providing a tool for more targeted public policies. For example, the results show that areas with low digital access and low skills are associated with high NEET rates, underscoring the need for tailored local interventions.

Building on this study, several future research directions are proposed.

Firstly, expanding the framework to a longitudinal systems model would allow researchers to track how digital exclusion evolves over time in response to interventions, shocks (e.g., pandemics), or policy shifts. Incorporating feedback loops and time delays could help identify leverage points for systemic change.

Secondly, applying the same systems-based approach at sub-national levels (NUTS2 or NUTS3) would uncover regional disparities and localized system behaviors. This would be especially relevant in large, heterogeneous countries where national averages mask internal complexity.

Thirdly, integration with qualitative systems mapping techniques (e.g., causal loop diagrams, stakeholder models) could bring in expert and experiential knowledge, complementing the quantitative insights and making the model more actionable for policy design.

Fourthly, exploring agent-based models (ABMs) could simulate interactions between individuals and institutions within the system, allowing for scenario testing under different policy, behavioral, or technological conditions.

Fifthly, the framework could be adapted for comparative international studies, applying the same systems methodology to assess digital exclusion and NEET dynamics in non-European contexts. This would test the transferability of findings and surface context-specific system behaviors.

Ultimately, future research should investigate the inclusion of additional subsystems, such as mental health, housing stability, or access to public services, which may further elucidate the mechanisms driving youth exclusion in the digital age.

From the methodological perspective, applying advanced techniques such as Geographically Weighted Regression (GWR) or spatial error/lag models would provide a more detailed understanding of spatial heterogeneity in NEET determinants [106–108]. Due to the limited spatial granularity of the available data (aggregated at the national level), the application of these techniques was not feasible within the scope of this study. Nonetheless, we acknowledge this as a valuable direction for future research, particularly when NEET data becomes available at NUTS 2 or NUTS 3 levels. In such a context, we plan to apply spatial autocorrelation diagnostics (e.g., Moran's I, Lagrange Multiplier tests) and explore local regression approaches.

Through these measures, the reduction in the NEET rate can be accelerated and young people can be better prepared for the future digital economy, avoiding the vulnerabilities associated with uncontrolled internet use.

Previous research on NEET status has primarily focused on economic and educational factors, highlighting the connection between the labor market structure and the integration of young people [51,52]. In particular, the existing literature emphasizes that labor market rigidity and inadequate educational skills contribute to the increase in the NEET rate, but the impact of digital skills and risky internet use on socio-professional integration has been insufficiently investigated [48,103].

This research makes an original contribution by integrating advanced multivariate analysis, utilizing PCA with Promax rotation and logistic regression, to examine the relationship between digital skills, risky internet use, and the likelihood of becoming a NEET. Unlike previous studies, this approach provides an integrated perspective on the digital dimensions of youth exclusion.

Furthermore, regional disparities in digital skills and the NEET rate have rarely been analyzed in a detailed spatial framework. By using GIS and spatial clustering, this study provides a new geographical perspective on the correlations between digital skill levels, internet access, and economic vulnerability [53]. The results obtained contribute to the formulation of personalized intervention strategies tailored to the digital profiles of young people and provide recommendations for more effective public policies aimed at reducing the NEET rate.

In summary, this study contributes to the literature by offering an integrated framework that links digital exclusion, territorial disparities, and youth socio-economic vulnerability. By combining multivariate statistics, CHAID segmentation, and spatial mapping, the research reinforces the strategic importance of digital competence, autonomy, and safety as core levers for reducing NEET risk in Europe's digital age.

Future directions include longitudinal system modeling to capture dynamics over time; sub-national disaggregation (NUTS2/NUTS3) for regional precision; integration with qualitative mapping (e.g., causal loop diagrams); agent-based simulations of digital exclusion scenarios; and replication in non-EU contexts to validate systemic patterns.

Through these approaches, youth resilience in the digital age can be reinforced, ensuring that digital skills and protections are equity-driven enablers, not barriers.

Author Contributions: Conceptualization, A.G., T.V.A. and C.L.; methodology, C.L. and A.G.; software, C.L.; validation, A.G. and T.V.A.; formal analysis, C.L., T.V.A. and A.G.; investigation, T.V.A. and C.L.; resources, T.V.A. and C.L.; data curation, C.L.; writing—original draft preparation, A.G., C.L. and T.V.A.; writing—review and editing, A.G. and C.L.; visualization, C.L.; supervision, A.G.; project administration, A.G.; funding acquisition, T.V.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Romanian Ministry of Research, Innovation, and Digitalization, Program NUCLEU, 2022–2026, PN 22_10_0105.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used are from Eurostat and the details about the codes of the datasets are presented in the main text.

Acknowledgments: This work was supported by a grant from the Romanian Ministry of Research, Innovation, and Digitalization, Program NUCLEU, 2022–2026, Spatio-temporal forecasting of local labour markets through GIS modelling [P5] PN 22_10_0105.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Descriptive Statistics

This appendix shows the descriptive statistics of 28 indicators, which are clustered into the six main dimensions of digital and social exclusion among the youth (aged 16–34) in 28 European countries (NUTS 0 level): the mean, the median, the minimum, the maximum, the standard deviation, the variance, the skewness, and the kurtosis. The sources of data used include Eurostat and were analyzed using SPSS v.29. The indicators are based on multi-annual trends (2020–2024) in three main aspects: socio-educational segregation (NEET), social networking in the digital sphere (social networks and instant messaging), and the level of digital skills, daily internet usage, and self-assessment of the safety of data protection. With the support of analysis, the interpretative framework of the main body of this article is supported and results in policy-oriented conclusions.

In Table A1 we present the set of indicators used for the model, and in Table A2 the descriptive statistics.

Table A1. Set of indicators for analyzing the relationship between risky internet use and the likelihood of becoming NEET (2020–2024).

	cd	ctr	rata NEET				Social Networks				Instant Messaging				Daily Internet Use				Personal Data Protection			Low Digital Skills		Narrow Digital Skills		Limited Digital Skills		No Digital Skills		
			NEET2020	NEET2021	NEET2022	NEET2023	2020SNET	2021SNET	2023SNET	2024SNET	2020CHAT1	2021CHAT1	2023CHAT1	2024CHAT1	IDay2020	IDay2021	IDay2022	IDay2023	IDay2024	IMAP2020	IMAP2021	IMAP2023	Level of Digital Skills							
																							I_DSK2_2021_LW	I_DSK2_2023_LW	I_DSK2_2021_N	I_DSK2_2023_N	I_DSK2_2021_LM	I_DSK2_2023_LM	I_DSK2_2021_X	I_DSK2_2023_X
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	BE	Belgium	9.2	7.4	6.6	6.7	94.61	89.21	90.66	87.85	83.55	82.59	84.26	85.87	95.97	96.56	95.88	96.44	97.52	56.75	60.44	63.83	18.65	20.57	9.21	7.57	2.90	1.68	1.27	0.32
2	BG	Bulgaria	14.4	13.9	12.3	11.4	81.31	86.80	88.09	85.35	76.93	82.21	86.21	81.31	90.18	91.73	94.4	93.24	93.13	42.29	47.05	50.34	19.54	19.91	11.46	11.79	6.65	4.78	4.23	3.91
3	CZ	Czechia	6.6	6.5	8.0	6.3	95.09	95.36	98.18	99.33	95.12	97.22	98.85	97.81	97.18	98.22	98.71	99.23	99.07	74.18	73.94	81.5	12.06	5.50	3.05	1.57	0.76	0	0.19	0
4	DK	Denmark	7.5	7.1	6.7	7.1	96.77	97.12	98.59	97.83	97.61	97.56	96.78	97.19	99.27	98.28	98.98	97.84	98.43	83.22	81.97	80.95	15.17	15.98	3.04	2.83	0	0.57	0.22	1.02
5	DE	Germany	7.4	7.8	7.0	7.5	88.54	77.95	73.06	83.99	94.95	77.62	77.64	90.24	97.64	95.49	95.09	95.16	96.14	82.65	63.08	66.48	25.18	25.61	9.35	10.41	3.77	3.47	1.54	1.59
6	EE	Estonia	9.2	10.9	10.7	9.6	94.24	94.33	93.63	89.46	94.36	91.20	97.25	93.10	98.41	98.61	98.33	98.91	99.15	58.81	64.07	72.12	15.51	11.37	4.19	1.54	0.72	0	0.27	0
7	EL	Greece	13.2	11.0	10.7	11.6	89.52	94.42	94.40	97.08	85.30	88.06	92.90	90.16	96.11	96.12	97.94	97.7	98.17	54.36	58.86	65.35	9.52	16.94	1.37	3.42	0	0.36	0.26	0
8	ES	Spain	13.9	11.0	10.5	9.9	93.02	93.19	90.02	90.40	98.48	97.52	98.32	98.03	97	96.87	98.11	99.79	99.75	74.31	75.92	79.13	9.73	12.14	3.71	2.51	1.04	0.78	0.31	0.80
9	HR	Croatia	12.1	12.5	11.4	9.8	94.49	96.72	93.99	89.38	96.63	99.70	99.12	93.03	99.33	99.38	98.18	97.83	99.84	67.13	68.67	63	9.38	10.01	2.84	3.49	0	0	0	0
10	IT	Italy	19.1	19.8	15.9	12.7	78.57	79.47	80.67	81.78	88.75	90.83	93.67	93.01	92.06	94.12	95.07	97.15	97.24	52.02	56.63	56.67	19.44	21.60	9.55	9.96	3.47	4.30	2.51	3.17
11	CY	Cyprus	13.9	12.5	12.5	11.9	94.67	95.16	96.87	97.02	99.19	96.38	98.56	98.59	99.23	98.54	98.41	99.34	99.51	66.04	68.25	65.14	17.41	19.52	4.20	4.96	2.27	2.68	1.95	0.37
12	LV	Latvia	7.1	8.6	8.6	7.2	91.54	93.02	93.29	92.13	91.07	93.45	94.90	92.18	96.41	98.84	98.94	99.39	98.79	60.64	61.43	51.05	15.49	19.74	3.06	10.97	0.39	2.54	0.66	1.01
13	LT	Lithuania	10.8	11.3	9.7	13.5	90.03	90.34	89.11	85.43	90.30	95.01	91.71	93.47	96.39	99.51	98.26	98.94	98.68	53.4	54.69	60.79	22.01	12.51	3.35	5.18	0.67	0.39	0.29	0.19
14	LU	Luxembourg	6.6	8.7	7.0	8.9	79.96	82.71	77.99	88.50	86.04	84.78	76.98	79.87	99.12	98.77	98.78	97.53	98.15	56.37	73.33	69.6	21.19	24.83	7.04	9.84	2.87	3.54	0	2.13
15	HU	Hungary	11.7	10.6	9.8	9.8	96.42	95.75	96.43	96.37	95.94	97.08	97.97	96.24	96.63	97.57	97.67	99.27	99.11	61.31	64.54	72.4	19.86	15.11	7.64	4.28	1.45	2.39	1.00	0.83
16	MT	Malta	9.6	10.6	7.2	8.2	92.65	93.75	95.70	92.65	99.37	98.95	98.36	97.78	98.99	98.95	99.55	98.84	100	73.37	72.54	77.51	5.15	3.21	0.95	0	0	0	0	0
17	NL	Netherlands	4.5	2.6	2.8	3.3	92.11	90.51	94.41	96.45	95.85	94.68	99.57	99.68	95.58	94.92	93.74	99.76	99.58	86.58	85.75	91.9	7.60	6.57	1.82	2.13	0.22	0.17	0.11	0
18	AT	Austria	8.0	8.5	8.1	8.7	95.46	89.78	88.98	90.43	98.26	99.17	95.08	95.88	99.36	96.94	96.49	97.01	98.86	79.45	76.37	75	14.51	14.65	4.07	5.46	2.28	4.73	0	0.59
19	PL	Poland	8.4	11.2	8.1	6.9	91.40	90.99	93.02	90.09	85.89	88.75	91.44	94.09	98.75	97.5	98.74	98.73	99.37	53.58	50.08	51.37	21.67	27.02	6.22	6.82	1.73	2.49	0.22	0
20	PT	Portugal	9.1	7.7	6.7	8.0	95.65	94.60	94.13	92.66	96.62	96.42	97.24	97.12	99.07	99.19	99.47	99.24	99.34	68.41	71.51	72.25	8.98	14.24	3.34	1.90	1.18	1.17	0	0.05
21	RO	Romania	14.8	18.0	17.5	16.5	86.02	88.93	89.07	89.21	65.45	73.86	80.26	77.61	88.95	94.46	94.8	96.22	96.44	40.9	47.09	46.3	21.93	23.08	13.81	14.92	7.94	7.76	6.13	4.87
22	SI	Slovenia	7.7	6.6	8.2	7.3	95.19	91.25	95.50	92.30	90.56	85.74	87.34	93.25	96.66	97.68	97.82	98.94	99	55.43	54.52	51.66	24.63	15.80	9.19	8.31	0.43	2.27	0	0.55
23	SK	Slovakia	10.7	11.0	9.6	8.9	90.27	91.91	85.95	84.56	89.34	85.47	85.71	83.83	98.62	93.18	95.98	96.9	96.9	59.86	61.3	60.84	18.69	16.10	3.25	7.14	1.98	3.10	1.18	1.49
24	FI	Finland	9.3	7.7	7.8	7.7	92.04	93.86	93.42	94.22	94.06	94.22	93.65	89.14	99.22	99.19	97.62	97.67	93.63	86.5	88.14	90.9	5.80	4.26	1.15	1.09	0	0.32	0	0.56
25	SE	Sweden	6.5	5.1	4.9	5.1	85.31	84.66	85.76	88.83	83.56	87.08	85.04	88.87	93.3	98.05	96.7	97.94	98.69	73.04	74.76	74.58	14.53	19.24	4.85	7.32	3.12	2.96	1.00	1.98
26	NO	Norway	4.9	6.3	6.4	5.4	97.89	96.46	98.04	97.69	98.64	96.56	98.06	98.84	100	98.3	98.84	98.87	100	84.42	86.72	92.59	9.99	11.62	3.41	1.89	0	0.38	0.69	0
27	RS	Serbia	15.9	17.0	13.3	12.4	80.05	94.83	97.00	94.23	93.28	95.79	99.53	94.77	98.85	96.87	100	100	97.36	41.77	42.48	34.02	23.48	45.48	8.56	5.04	0	0.97	0	0
28	TR	Türkiye	28.3	24.8	24.1	22.4	79.99	81.46	92.17	92.34	89.10	92.15	95.18	95.03	86.2	91.99	93.32	94.22	94.76	44.62	48.23	51.84	23.36	21.69	11.43	14.95	5.93	5.09	3.30	1.87

Table A2. Descriptive statistics.

	N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance	Skewness	Kurtosis			
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
NEET2020	28	23.8	4.5	28.3	300.4	10.729	0.9321	4.9320	24.325	1.828	0.441	4.909	0.858
NEET2021	28	22.2	2.6	24.8	296.7	10.596	0.8955	4.7385	22.454	1.239	0.441	2.058	0.858
NEET2022	28	21.3	2.8	24.1	272.1	9.718	0.8008	4.2374	17.956	1.640	0.441	4.040	0.858
NEET2023	28	19.1	3.3	22.4	264.7	9.454	0.7201	3.8104	14.519	1.567	0.441	3.979	0.858
2020SNET	28	19.32	78.57	97.89	2532.81	90.4575	1.09999	5.82060	33.879	-0.902	0.441	-0.408	0.858
2020CHAT1	28	33.92	65.45	99.37	2554.20	91.2214	1.44249	7.63292	58.261	-1.644	0.441	3.644	0.858
2021SNET	28	19.17	77.95	97.12	2544.54	90.8764	1.00582	5.32230	28.327	-1.103	0.441	0.375	0.858
2021CHAT1	28	25.84	73.86	99.70	2560.05	91.4304	1.29510	6.85302	46.964	-0.930	0.441	0.177	0.858
2023SNET	28	25.53	73.06	98.59	2558.13	91.3618	1.16974	6.18967	38.312	-1.413	0.441	1.986	0.858
2023CHAT1	28	22.59	76.98	99.57	2591.58	92.5564	1.30330	6.89640	47.560	-0.997	0.441	-0.119	0.858
2024SNET	28	17.55	81.78	99.33	2557.56	91.3414	0.88303	4.67254	21.833	-0.151	0.441	-0.718	0.858
2024CHAT1	28	22.07	77.61	99.68	2585.99	92.3568	1.12739	5.96558	35.588	-1.074	0.441	0.447	0.858
IDay2020	28	13.80	86.20	100.00	2704.48	96.5886	.65615	3.47202	12.055	-1.658	0.441	2.324	0.858
IDay2021	28	7.78	91.73	99.51	2715.83	96.9939	0.42041	2.22461	4.949	-1.064	0.441	0.257	0.858
IDay2022	28	6.68	93.32	100.00	2725.82	97.3507	0.35354	1.87078	3.500	-0.744	0.441	-0.567	0.858
IDay2023	28	6.76	93.24	100.00	2742.10	97.9321	0.31947	1.69048	2.858	-1.217	0.441	1.242	0.858
IDay2024	28	6.87	93.13	100.00	2746.61	98.0932	0.34794	1.84114	3.390	-1.432	0.441	1.524	0.858
IMAP2020	28	45.68	40.90	86.58	1791.41	63.98	2.67	14.12	199.287	0.076	0.441	-1.034	0.858
IMAP2021	28	45.66	42.48	88.14	1832.36	65.4414	2.40329	12.71700	161.722	0.043	0.441	-0.803	0.858
IMAP2023	28	58.57	34.02	92.59	1869.11	66.7539	2.74374	14.51852	210.787	-0.076	0.441	-0.294	0.858
I_DSK2_2021_LW	28	20.03	5.15	25.18	450.46	16.0879	1.13767	6.01996	36.240	-0.268	0.441	-1.146	0.858
I_DSK2_2021_N	28	12.86	0.95	13.81	155.11	5.5396	0.66889	3.53941	12.527	0.726	0.441	-0.522	0.858
I_DSK2_2021_LM	28	7.94	0.00	7.94	51.77	1.8489	0.40355	2.13540	4.560	1.477	0.441	1.784	0.858
I_DSK2_2021_X	28	6.13	0.00	6.13	27.33	0.9761	0.27827	1.47247	2.168	2.222	0.441	5.131	0.858
I_DSK2_2023_LW	28	42.27	3.21	45.48	474.30	16.9393	1.59938	8.46311	71.624	1.174	0.441	3.682	0.858
I_DSK2_2023_N	28	14.95	0.00	14.95	167.29	5.9746	0.78737	4.16636	17.359	0.652	0.441	-0.405	0.858
I_DSK2_2023_LM	28	7.76	0.00	7.76	58.89	2.1032	0.37488	1.98368	3.935	0.967	0.441	0.745	0.858
I_DSK2_2023_X	28	4.87	0.00	4.87	27.30	0.9750	0.24062	1.27323	1.621	1.706	0.441	2.637	0.858
Valid N (listwise)	28												

Appendix B. PCA—Total Variance Explained (Varimax Rotation)

To assess the robustness of the Promax-rotated PCA that features in the main model, we also conducted a Principal Component Analysis (PCA) with a Varimax orthogonal rotation. This supplementary analysis aimed to determine how independent and interpretable the retained components were. Varimax rotation rearranged the factor loadings with the aim of reducing multicollinearity and maximizing the components to better understand their meaning and clarity, while still preserving the essence of the data.

The percentage of variance accounted for by each of the above four components was retained by selecting the Varimax rotational orthogonal Schwinger shell model. The PCA presented in Table A3 below exhibits eigenvalues, as well as the percentage variance that each of the four components explains. These four components were identified based on the criteria proposed by Kaiser (eigenvalue > 1). After accounting together, 82.55 percent of the total amount of variance is dispersed, a figure that is sufficient to reduce the dimensionality without losing important information in significant quantities.

Component 1 plays the most significant role, explaining the largest part of the variance (56.78%), which indicates that a dominant latent factor is present in the dataset. Components 2, 3, and 4 add 11.77 percent, 8.51 percent, and 5.49 percent, respectively. The variance

redistribution across components that accompanies the Varimax rotation is beneficial to interpretation. Still, the sums of squared loadings rotated to general interpretability do not provide the sum of squares of the total variance explained because of the way the rotation disrupts normal orthogonal assumptions. The elements other than the fourth are distinguished by eigenvalues that are less than 1 and explain less than 5 percent of the variance each, which proves their insignificance. The results justify the choice to retain four components, further interpreting them and demonstrating the efficacy of the dimensionality reduction method applied, namely PCA.

This added analysis confirms the structural validity of the model and enhances the interpretability of the latent factors identified in the research.

Table A3. PCA. Total variance explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings ^a
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	13.628	56.783	56.783	13.628	56.783	56.783	11.501
2	2.825	11.771	68.554	2.825	11.771	68.554	9.102
3	2.042	8.507	77.061	2.042	8.507	77.061	9.341
4	1.318	5.492	82.553	1.318	5.492	82.553	6.776
5	0.899	3.747	86.301				
6	0.694	2.891	89.192				
7	0.649	2.703	91.895				
8	0.489	2.037	93.932				
9	0.344	1.434	95.366				
10	0.243	1.013	96.379				
11	0.192	0.801	97.180				
12	0.156	0.651	97.832				
13	0.124	0.516	98.347				
14	0.101	0.422	98.770				
15	0.084	0.351	99.121				
16	0.070	0.294	99.414				
17	0.040	0.167	99.581				
18	0.034	0.140	99.721				
19	0.028	0.115	99.836				
20	0.015	0.064	99.900				
21	0.012	0.051	99.951				
22	0.007	0.031	99.982				
23	0.003	0.011	99.993				
24	0.002	0.007	100.000				

Extraction method: Principal Component Analysis. ^a When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table A4 reflects the rotated component matrix after Varimax-rotated PCA. The factor loadings indicate the correlation of each of the observed variables with the components pulled out, and they may be interpreted to work out the latent dimensions.

The idea behind the use of Varimax orthogonal rotation is the redistribution of loadings, which succeeds in the independence of factors, making the interpretation of the underlying structure readily understandable. The important insights that resulted from the rotated loadings were as follows:

- Component 1 represents a high-frequency use of the internet in general, and good positive loadings are associated with daily internet use in every year (e.g., IDay2020 to IDay2024). This aspect probably reflects the behavioral or institutionalized digital interaction.
- Component 2 loads heavily on the application of instant messaging tools (IMAP20202023) and the identification of some digital proficiency (I DSK2 2021 LW, I DSK2 2023 LW), which indicates that it is associated with interactive digital communication and ability.

- Component 3 is occupied by the use of chat platforms (CHAT1) year-on-year, where the loads are very high in 2020–2024, which means high intensities of social interaction regarding digital environments.
- In component 4, there is a high load on SNET (social networks) variables in 2021–2024, which presents social media intensity as a standout behavior dimension.

It is interesting to point out that digital skill variables have negative loadings in several variables, with the strongest ones occurring in components 1 and 2, which suggests that self-reported skill proficiency is related inversely to problematic or intensive digital behaviors.

The rotation has converged in six iterations and shows the structure to be well differentiated with little cross-loadings, thus indicating the internal consistency and conceptual coherence of the factors extracted. This circularized matrix strengthens the interpretiveness of the multivariate analysis and is consistent with theoretical predications on the digital usage patterns and sensitivities in adolescent populations.

Table A4. PCA. Rotated component matrix ^a.

	Component			
	1	2	3	4
2020CHAT1	0.459	0.387	0.734	0.010
2021CHAT1	0.343	0.269	0.731	0.362
2023CHAT1	0.161	0.151	0.749	0.576
2024CHAT1	0.260	0.207	0.856	0.269
2020SNET	0.439	0.505	0.084	0.442
2021SNET	0.486	0.180	0.096	0.803
2023SNET	0.190	0.045	0.298	0.911
2024SNET	0.189	0.246	0.287	0.708
IDay2020	0.868	0.296	0.114	0.008
IDay2021	0.828	0.237	0.015	0.191
IDay2022	0.859	−0.058	0.063	0.320
IDay2023	0.715	0.067	0.324	0.355
IDay2024	0.626	0.128	0.291	0.215
IMAP2020	0.253	0.856	0.258	−0.094
IMAP2021	0.257	0.898	0.156	0.016
IMAP2023	0.153	0.933	0.164	0.050
I_DSK2_2021_LW	−0.174	−0.758	−0.165	−0.368
I_DSK2_2021_N	−0.559	−0.613	−0.261	−0.220
I_DSK2_2021_LM	−0.774	−0.285	−0.390	−0.242
I_DSK2_2021_X	−0.813	−0.284	−0.370	0.037
I_DSK2_2023_LW	−0.014	−0.823	−0.071	−0.274
I_DSK2_2023_N	−0.581	−0.541	−0.336	−0.351
I_DSK2_2023_LM	−0.641	−0.372	−0.346	−0.288
I_DSK2_2023_X	−0.690	−0.208	−0.483	−0.253

Extraction method: Principal Component Analysis. Rotation method: Varimax with Kaiser normalization.

^a Rotation converged in 6 iterations.

Appendix C. CHAID-Based Decision Tree Classification of NEET2023_Binary Using Digital Behavior Variables

Appendix C.1. Introduction: Why CHAID? Rationale, Contribution, and Fit

This appendix focuses on a decision tree classification model applied to study the correlation between digital engagement and status as NEET in 2023 between European countries. We use the CHAID (Chi-squared Automatic Interaction Detection) analysis because it is a non-parametric method with maximum interpretability and fits well to a small-N cross-country dataset. CHAID is not wedded to any assumptions about the

distribution of the data and can identify the effects of interaction and nonlinear thresholds whereas parametric analysis may not be applicable to data governed by such complex socio-digital phenomenon as youth exclusion [77,115].

The main goal of this modeling step will be to determine whether there is any individual digital behavior or skill-related characteristic, which can be effective to categorize countries based on their prevalence of NEET. This is an addition to the findings of a logistic regression analysis because it puts the statistical associations into sensible thresholds that can be applied as part of a set of rules. CHAID can play its role in modeling chains by identifying the strongest discriminator of NEET status and developing a set of comprehensible rules of classification, which increases model transparency and usability by policy [116,117].

The dependent variable is the binary transformation of the NEET rate in 2023 (NEET2023_Binary) where the country is coded as 1 in case it is higher than the European median. The independent variables (n = 24) include three major conceptual areas: digital communication (CHAT, SNET), digital usage intensity (IDay, IMAP), and digital skills (DSK2, disaggregated by level/year). This exhaustive selection considers the theoretical literacy available on the digital divide [79,118], and it is performed so in more ways so that the CHAID algorithm could know which factor is most predictive.

Appendix C.2. Model Summary

The decision tree classification was conducted using the CHAID growing method, with the dependent variable being the binary transformation of NEET rate in 2023 (NEET2023_Binary). The model tested a comprehensive set of 24 independent variables, grouped into three conceptual dimensions: digital communication behavior (CHAT and SNET indicators for 2020–2024), digital usage intensity (internet use per day and for managing personal data—IMAP), and digital skills (disaggregated skill levels for 2021 and 2023).

Table A5. Model summary.

Specifications	Growing Method	CHAID
	Dependent Variable	NEET2023_Binary
	Independent Variables	2020CHAT1, 2021CHAT1, 2023CHAT1, 2024CHAT1, 2020SNET, 2021SNET, 2023SNET, 2024SNET, IDay2020, IDay2021, IDay2022, IDay2023, IDay2024, IMAP2020, IMAP2021, IMAP2023, I_DSK2_2021_LW, I_DSK2_2021_N, I_DSK2_2021_LM, I_DSK2_2021_X, I_DSK2_2023_LW, I_DSK2_2023_N, I_DSK2_2023_LM, I_DSK2_2023_X
	Validation	Cross Validation
Results	Maximum Tree Depth	3
	Minimum Cases in Parent Node	10
	Minimum Cases in Child Node	5
	Independent Variables Included	IMAP2020
	Number of Nodes	3
	Number of Terminal Nodes	2
	Depth	1

The set-up of the models addressed the following normal robustness criteria:

- Deepest tree structure: Max of three levels, enabling them to be complex enough to define the interactions of the variables.

- At least 10 cases per parent node, 5 per child node: these cut-offs guarantee the validity of the statistical findings and lack overfitting.

Validation technique: cross-validation that improves the generalizability of the model and helps avoid overfitting to inaccurate patterns [116].

The results of this indicate a very clear result, that there was only one independent variable contained in the best decision tree, which was IMAP2020 (the use of the internet to manage personal data), showing that of all the possible predictors this was the most strongly associated with NEET status in 2023 and is statistically significant. It forms a decision tree with three nodes, two of which are terminal, and the depth of the tree is only one, and it is hence a highly discriminative, simple structure.

This finding indicates that IMAP2020 can be used as a solid classifier that can successfully divide the countries by creating two groups with a considerable difference in NEET rates. Its choice defends the main hypothesis that the functional digital engagement, especially in data management and digital administration, is one of the major signs of youth employability or learning.

The parsimonious nature of the model such as the use of a single predictor serves to further emphasize the understandability and interpretability of the decision path, this being a major strength of CHAID over more inscrutable models [77,115].

Appendix C.3. CHAID Decision Tree Output

The CHAID decision tree in Statistical Annex 2 Figure 1 clearly gives a structure that would allow the categorization of European countries by the NEET status in 2023. The model finds an important tolerance of 67.13 percent in variable IMAP2020 (the use of the internet to handle personal information), which results in two homogeneous categories with apparent different NEET prevalence.

- Node 1: Nations with IMAP2020 67.13 or below are characterized by a high NEET rate where 76.5 percent of cases have been noted as NEET.
- Node 2: Those countries whereby IMAP2020 is above 67.13 percent display a low NEET rate, as 90.9 percent of the observations remain non-NEET.

Our findings can be empirically supported by other research findings that show a correlation between online autonomy and youth exclusion [118]. In addition, it also supplements the logistic regression by clearly specifying an important threshold effect that can be implicit in parametric modeling.

To further check the statistical significance of the split, the Bonferroni-adjusted value of Chi-square is also 12.128 ($p = 0.004$), and the model is very simple (depth = 1) and is therefore clearly interpretable into policy. This result provides confirmation of the importance of digital capital in minimizing NEET rates and reflects the supplementary worth of applying both regression and classification solutions, including CHAID analysis [77,116].

Figure A1 displays a binary segmentation of European countries based on the threshold value of the IMAP2020 indicator (percentage of young people using the internet to manage personal data). The CHAID tree highlights a cut-off point at 67.13%, separating two distinct groups of countries in terms of the probability of recording high NEET rates in 2023. This representation supports the conceptual hypothesis that functional online activities act as a form of digital capital, offering protection against social exclusion.

Appendix C.5. Risk Estimates—Model Accuracy and Generalization Power

The CHAID model portrays 0.179 (standard error = 0.072) as an estimate of the resubstitution risk estimate, which implies that the low error rate is a predictor of NEET in the same data with which it was produced. The risk of cross-validation, however, rises to 0.286 (standard error = 0.085) where the generalizability of the model in fresh or unseen data is lower. Such a difference between the accuracies in training and validation is not extraordinary in decision trees and explains why special caution should be applied when interpreting in small sample settings [77,78].

Table A7. CHAID model risk estimates—resubstitution and cross-validation errors for NEET2023_Binary.

Method	Estimate	Std. Error
Resubstitution	0.179	0.072
Cross-Validation	0.286	0.085

Growing method: CHAID. Dependent variable: NEET2023_Binary.

However, the out-of-sample finding is considered acceptable in terms of use in social science, as it endorses the model's strength in determining digital exclusion levels that lead to NEET vulnerability. It also confirms that binary division on IMAP2020 is a valid and generalizable marker of labor market disengagement among young people in European nations.

Appendix C.6. CHAID Model Classification Performance for NEET2023_Binary

The table of classifications tests the generalizing power of the CHAID decision tree to classify countries based on their NEET2023_Binary nature. The model has an overall accuracy of 82.1 percent, with a 71.4 percent accuracy rate in non-NEET countries and a 92.9 percent accuracy rate in NEET countries. The asymmetry emphasizes that the model is highly sensitive to discriminate countries reported to be at a higher risk of NEET and this is especially useful when applied to policy applications involving risk population groups of young people.

Table A8. CHAID model classification accuracy for NEET2023_Binary.

Observed	Predicted		Percent Correct
	0.00	1.00	
0.00	10	4	71.4%
1.00	1	13	92.9%
Overall Percentage	39.3%	60.7%	82.1%

Growing method: CHAID. Dependent variable: NEET2023_Binary.

Although the sample size is relatively small ($N = 28$), the results of the classification indicate that the chosen predictor (IMAP2020) is capable of reflecting one of the most important digital behavior dimensions related to NEET vulnerability. These findings supplement past studies underlining digital independence and data security as socio-economic distinctions in coping with moving to education, employment, or training in the youth [47,79].

Appendix C.7. CHAID-Based Rule Generation for Predicting NEET Status in SPSS Syntax

The model of the CHAID decision tree can be operationalized by rule-based classification syntax, which is found in SPSS. According to the model, the segmentation criteria are

translated into binary decision rules that allow the assignment of new cases to the predicted NEET categories under the IMAP2020 threshold of 67.13 percent.

```
/* Node 1 */.
DO IF (SYSMIS(IMAP2020) OR (VALUE(IMAP2020) LE 67.13)).
COMPUTE nod_001 = 1.
COMPUTE pre_001 = 1.
COMPUTE prb_001 = 0.764706.
END IF.
EXECUTE.
/* Node 2 */.
DO IF (VALUE(IMAP2020) GT 67.13).
COMPUTE nod_001 = 2.
COMPUTE pre_001 = 0.
COMPUTE prb_001 = 0.909091.
END IF.
EXECUTE.
```

The two nodes are encoded as follows:

- **Node 1:** Individuals from countries with $\text{IMAP2020} \leq 67.13$ are classified with a high likelihood (76.5%) of being NEETs.
- **Node 2:** Conversely, countries where $\text{IMAP2020} > 67.13$ show a low probability (9.1%) of NEET classification.

These rules facilitate the reproduction and scalability of the CHAID model logic in applied contexts, enhancing model transparency and usability in policy simulations or subsequent predictive workflows [115,117].

Appendix C.8. CHAID Classification Model—Contribution to the Analytical Framework and Policy Interpretation

This chapter introduces the CHAID (Chi-squared Automatic Interaction Detection) model as an alternative interpretable and highly complementary way to study the connections between digital behavior and NEET (Not in Education, Employment, or Training) prevalence in European countries. The argument to apply the CHAID method is that this method establishes statistically significant thresholds in categorical or continuous predictors and it creates straightforward classification trees with rule-like branches. This is especially applicable in the policy making scenario where one needs transparency and operable cut-offs [77,115,116].

The CHAID model with the binary transformation of the NEET rate in 2023 was fitted on three dimensions of digital behavior: communication through chat/social networks, the strength of internet use, and digital skills where the number of candidate predictors consisted of 24 variables. Although the initial input was very rich, there was only one strong explanatory variable according to the model, IMAP2020, the proportion of young people who are using the internet to deal with their personal data.

The result of the model has indicated a red line of 67.13% in IMAP2020. All countries with 67.13% or less were placed in Node 1, which had a high prevalence of NEETs (76.5%), and those with more than 67.13% in Node 2, which had a low prevalence of NEETs (9.1% only). This is a binary split that creates two groups of individuals:

- **Cluster 1 (IMAP2020 $\leq 67.13\%$):** Romania, Bulgaria, Italy, Greece, Spain, Croatia, Hungary, Slovakia, Poland, Portugal, Cyprus, Latvia, Lithuania, Malta, Slovenia, Czechia, and Estonia.

- Cluster 2 (IMAP2020 > 67.13%): Sweden, Finland, Denmark, the Netherlands, Germany, Austria, France, Belgium, Ireland, Luxembourg, and Norway.

The variable IMAP2020 not only measures digital access but also the third and last level of digital competence (i.e., the possibility to manage and protect personal digital data, which have become predictors of youth inclusion in the labor market [79,80,119]. The model stands out in its parsimony and intelligibility which cements the interpretability of digital autonomy as a policy-relevant construction, thus giving a decisive point of reference towards specifying which countries are at risk of the digital transformation and laying out a customized roadmap of digital inclusion.

Methodologically, the CHAID model can be seen as contributing to the analysis chain since the classification accuracy is 82.1 percent, therefore, interpretable decision rules will be generated and nonlinear effects can be identified that may be obscured in parametric models, like logistic regression. The cross-validation risk for the model is 0.286, which falls within acceptable limits, confirming that its findings can be generalized, despite using a small sample size of $N = 28$ [78].

To summarize, this CHAID analysis also reinforces the fact that there is a functional and actionable digital divide in Europe, whereby the risk of NEET among the youth becomes substantially higher in countries with lower digital agency rates. The added value is translated into policy-relevant levels of digital indicators, the predictive value of managing personal data effectively, and a scalable and transparent model that supplements the regression results. Within the general conceptualization of the modeling strategy, the CHAID model provides a level of threshold articulation, corrects the imperfection of interpretability, and supports the overarching hypothesis, which holds that digital autonomy is a positive measure of youth attainment in the socio-economic field.

References

1. Freiman, V.; Godin, J.; Larose, F.; Leger, M.; Chiasson, M.; Volkanova, V.; Goulet, M.-J. Towards the life-long continuum of digital competences: Exploring combination of soft-skills and digital skills development. In Proceedings of the INTED2017: 11th International Technology, Education and Development Conference, Valencia, Spain, 6–8 March 2017; pp. 9518–9527. [\[CrossRef\]](#)
2. van Laar, E.; van Deursen, A.J.A.M.; van Dijk, J.A.G.M.; de Haan, J. Determinants of 21st-century digital skills: A large-scale survey among working professionals. *Comput. Hum. Behav.* **2019**, *100*, 93–104. [\[CrossRef\]](#)
3. Babic, A. Digital skills as a perspective of development of the economy and important digital transformation factor. *Ekon. Pregl.* **2021**, *72*, 59–87. [\[CrossRef\]](#)
4. Rodriguez-Ruiz, J.; Alvarez-Delgado, A.; Caratozzolo, P. Use of Natural Language Processing (NLP) Tools to Assess Digital Literacy Skills. In Proceedings of the 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, Monterrey, Mexico, 15–17 December 2021. [\[CrossRef\]](#)
5. Ozyurt, O.; Ayaz, A. Identifying cyber security competencies and skills from online job advertisements through topic modeling. *Secur. J.* **2024**, *37*, 1339–1359. [\[CrossRef\]](#)
6. Inverarity, C.; Tarrant, D.; Forrest, E.; Greenwood, P.P. Towards Benchmarking Data Literacy. In Proceedings of the WWW'22: The ACM Web Conference 2022, Lyon, France, 25–29 April 2022; pp. 408–416. [\[CrossRef\]](#)
7. Ozyurt, O.; Gurcan, F.F.; Dalveren, G.G.M.; Derawi, M. Career in Cloud Computing: Exploratory Analysis of In-Demand Competency Areas and Skill Sets. *Appl. Sci.* **2022**, *12*, 9787. [\[CrossRef\]](#)
8. Lacerda, G.; Nogueira, M. Navigating the AI Revolution: Tools and Skills Transforming Marketing Practices. *Int. J. Mark. Commun. New Media* **2024**, *15*, 55–71. [\[CrossRef\]](#)
9. Lau, W.W.F.; Yuen, A.H.K. Internet ethics of adolescents: Understanding demographic differences. *Comput. Educ.* **2014**, *72*, 378–385. [\[CrossRef\]](#)
10. Sari, D.I.; Rejekiningsih, T.; Muchtarom, M. Students' Digital Ethics Profile in the Era of Disruption: An Overview from the Internet Use at Risk in Surakarta City, Indonesia. *Int. J. Interact. Mob. Technol.* **2020**, *14*, 82–94. [\[CrossRef\]](#)
11. Ortega-Barón, J.; González-Cabrera, J.; Machimbarrena, J.M.; Montiel, I. Safety.Net: A pilot study on a multi-risk internet prevention program. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4249. [\[CrossRef\]](#)
12. Kaess, M.; Klar, J.; Kindler, J.; Parzer, P.; Brunner, R.; Carli, V.; Sarchiapone, M.; Hoven, C.W.; Apter, A.; Balazs, J.; et al. Excessive and pathological Internet use—Risk-behavior or psychopathology? *Addict. Behav.* **2021**, *123*, 107045. [\[CrossRef\]](#)

13. Qian, B.; Huang, M.; Xu, M.; Hong, Y. Internet Use and Quality of Life: The Multiple Mediating Effects of Risk Perception and Internet Addiction. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1795. [CrossRef]
14. Machimbarrena, J.M.; Calvete, E.; Fernández-González, L.; Álvarez-Bardón, A.; Álvarez-Fernández, L.; González-Cabrera, J. Internet risks: An overview of victimization in cyberbullying, cyber dating abuse, sexting, online grooming and problematic internet use. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2471. [CrossRef]
15. Sindu, P. Digital transformation in higher education: Advantages and challenges in 2023. In *The Impact of Digitalization in a Changing Educational Environment*; IGI Global: Hershey, PA, USA, 2023; pp. 59–6918. [CrossRef]
16. Tercova, N.; Smahel, D. Digital Skills' Role in Intended and Unintended Exposure to Harmful Online Content Among European Adolescents. *Media Commun.* **2025**, *13*, 8963. [CrossRef]
17. Porat, E.; Blau, I.; Barak, A. Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Comput. Educ.* **2018**, *126*, 23–36. [CrossRef]
18. Ferreira, E.; Silva, M.J. Young people's digital competences: Does gender matter? In Proceedings of the 2023 International Symposium on Computers in Education (SIIE), Setúbal, Portugal, 16–18 November 2023; pp. 1–6. [CrossRef]
19. Calderón, D.; Sanmartín Ortí, A.; Kardelis, S.K. Self-confidence and digital proficiency: Determinants of digital skills perceptions among young people in Spain. *First Monday* **2022**, *27*. [CrossRef]
20. Eynon, R.; Geniets, A. The digital skills paradox: How do digitally excluded youth develop skills to use the internet? *Learn. Media Technol.* **2015**, *41*, 463–479. [CrossRef]
21. ElSawy, A. The Problematic Internet Use and Its Impact on Young People's Online Moral Disengagement. In *Interdisciplinary Approaches for Educators' and Learners' Well-Being*; ElSawy, A., Olowoselu, R., Eds.; Springer: Cham, Switzerland, 2024. [CrossRef]
22. Müller, K.W.; Scherer, L. Excessive Use Patterns and Internet Use Disorders: Effects on Psychosocial and Cognitive Development in Adolescence. *Prax. Kinderpsychol. Kinderpsychiatr.* **2022**, *71*, 345–361. [CrossRef]
23. von Schoultz, D.J.; Thomson, K.-L.; Van Niekerk, J. Internet Self-regulation in Higher Education: A Metacognitive Approach to Internet Addiction. In IFIP Advances in Information and Communication Technology, 593 IFIPAICT, Proceedings of the 14th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2020 Mytilene, Lesbos, Greece, 8–10 July 2020; pp. 186–207. [CrossRef]
24. Korte, M. The impact of the digital revolution on human brain and behavior: Where do we stand? *Dialogues Clin. Neurosci.* **2022**, *22*, 101–111. [CrossRef]
25. de Barros, E.C. Understanding the influence of digital technology on human cognitive functions: A narrative review. *IBRO Neurosci. Rep.* **2024**, *17*, 415–422. [CrossRef]
26. Bejaković, P.; Mrnjavac, Ž. The importance of digital literacy on the labour market. *Empl. Relat.* **2020**, *42*, 921–932. [CrossRef]
27. Kim, K.T. The Mediating Role of Core Competencies in the Relationship between Digital Literacy and Perceived Employability among Korean College Students: Difference by Employment Support Program Participation. *Univers. J. Educ. Res.* **2020**, *8*, 2520–2535. [CrossRef]
28. Suarta, I.M.; Suwintana, I.K. The new framework of employability skills for digital business. *J. Phys. Conf. Ser.* **2021**, *1833*, 012034. [CrossRef]
29. Niyazova, A.Y.; Chistyakov, A.A.; Volosova, N.Y.; Krokhina, J.A.; Sokolova, N.L.; Chirkina, S.E. Evaluation of pre-service teachers' digital skills and ICT competencies in context of the demands of the 21st century. *Online J. Commun. Media Technol.* **2023**, *13*, e202337. [CrossRef]
30. Kovács, I.; Keresztes, É.R. Young Employees' Perceptions about Employability Skills for E-Commerce. *Economies* **2022**, *10*, 309. [CrossRef]
31. Tee, P.K.; Song, B.L.; Ho, M.K.; Wong, L.C.; Lim, K.Y. Bridging the gaps in digital skills: Employer insights on digital skill demands, micro-credentials, and graduate employability. *J. Infrastruct. Policy Dev.* **2024**, *8*, 7313. [CrossRef]
32. Promma, W.; Imjai, N.; Usman, B.; Aujirapongpan, S. The influence of AI literacy on complex problem-solving skills through systematic thinking skills and intuition thinking skills: An empirical study in Thai gen Z accounting students. *Comput. Educ. Artif. Intell.* **2025**, *8*, 100382. [CrossRef]
33. Brodny, J.; Tutak, M. Stakeholder interactions and ethical imperatives in big data and AI development. *J. Open Innov. Technol. Mark. Complex.* **2025**, *11*, 100491. [CrossRef]
34. Kalfeli, P.; Angeli, C. The Intersection of AI, Ethics, and Journalism: Greek Journalists' and Academics' Perspectives. *Societies* **2025**, *15*, 22. [CrossRef]
35. Yagoda, M. Airline Held Liable for Its Chatbot Giving Passenger Bad Advice—What This Means for Travellers. 2024. Available online: <https://bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know> (accessed on 17 October 2024).
36. Chen, H.; Magramo, K. Finance Worker Pays Out \$25 Million After Video Call with Deepfake 'Chief Financial Officer'. CNN 2024. Available online: <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html> (accessed on 17 October 2024).

37. Arayankalam, J.; Krishnan, S. Relating foreign disinformation through social media, domestic online media fractionalization, government's control over cyberspace, and social media-induced offline violence: Insights from the agenda-building theoretical perspective. *Technol. Forecast. Soc. Change* **2021**, *166*, 120661. [CrossRef]
38. Kalaichelvi, T.; Mane, S.B.; Dhanalakshmi, K.M.; Prasad, S. The detection of phishing attempts in communications systems. *Int. J. Electron. Secur. Digit. Forensics (IJESDF)* **2023**, *15*, 2023. [CrossRef]
39. Bartoletti, M.; Lande, S.; Loddo, A.; Pompianu, L.; Serusi, S. Cryptocurrency Scams: Analysis and Perspectives. *IEEE Access* **2021**, *9*, 148353–148373. [CrossRef]
40. Ciaramella, G.; Iadarola, G.; Martinelli, F.; Mercaldo, F.; Santone, A. Explainable Ransomware Detection with Deep Learning Techniques. *J. Comput. Virol. Hacking Tech.* **2024**, *20*, 317–330. [CrossRef]
41. Tambe Ebot, A. Advance fee fraud scammers' criminal expertise and deceptive strategies: A qualitative case study. *Inf. Comput. Secur.* **2023**, *31*, 478–503. [CrossRef]
42. European Commission. Ethics Guidelines for Trustworthy AI. 2019. Available online: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed on 20 January 2025).
43. Chauhan, S.; Sharma, D.; Jindal, P. Data ethics in digital age: Practices and strategies for responsible data handling. In *Banktech 4.0: The Next Wave of Transformative Banking*; John Wiley & Sons: Hoboken, NJ, USA, 2024; pp. 63–82.
44. Alshahrani, N.S.; Alotaibi, F.S.; Alyoubi, K.H.; Ramzan, M.S. Cybercrime in the AI Era: Definitions, Classification, Severity Assessment and the Role of AI in Combating Threats. *J. Comput. Sci.* **2025**, *21*, 665–684. [CrossRef]
45. Livingstone, S.; Helsper, E.J. Taking risks when communicating on the Internet: The role of offline social-psychological factors in young people's vulnerability to online risks. *Inf. Commun. Soc.* **2007**, *10*, 619–644. [CrossRef]
46. Kirschner, P.A.; Karpinski, A.C. Facebook® and academic performance. *Comput. Hum. Behav.* **2010**, *26*, 1237–1245. [CrossRef]
47. OECD. *Skills for the Digital Transition: Assessing Recent Trends Using Big Data*; OECD Publishing: Paris, France, 2022. [CrossRef]
48. Hargittai, E. Digital na (t) ives? Variation in internet skills and uses among members of the "net generation". *Sociol. Inq.* **2010**, *80*, 92–113. [CrossRef]
49. Helsper, E.J.; van Deursen, A.J.A.M.; Eynon, R. *Measuring Types of Internet Use: From Digital Skills to Tangible Outcomes Project Report*; University of Oxford: Oxford, UK, 2016.
50. Lissitsa, S. Generations X, Y, Z: The effects of personal and positional inequalities on critical thinking digital skills. *Online Inf. Rev.* **2025**, *49*, 35–54. [CrossRef]
51. Filandri, M.; Nazio, T.; O'Reilly, J. Youth transitions and job quality: How long should they wait and what difference does the family make? In *Youth Labor in Transition. Inequalities, Mobility, and Policies in Europe*; Oxford University Press: Oxford, UK, 2019; pp. 271–293.
52. Redmond, P.; McFadden, C. Young people not in employment, education or training (NEET): Concepts, consequences and policy approaches. *Econ. Soc. Rev.* **2023**, *54*, 285–327.
53. Eurostat. Youth NEET Rate by Age and Sex. 2024. Available online: https://ec.europa.eu/eurostat/databrowser/view/edat_lfse_20/default/table?lang=en (accessed on 5 November 2024).
54. Montag, C.; Wegmann, E.; Sariyska, R.; Demetrovics, Z.; Brand, M. How to overcome taxonomical problems in the study of Internet use disorders and what to do with "smartphone addiction"? *J. Behav. Addict.* **2021**, *9*, 908–914. [CrossRef]
55. Twenge, J.M.; Campbell, W.K. Media use is linked to lower psychological well-being: Evidence from three datasets. *Psychiatr. Q.* **2019**, *90*, 311–331. [CrossRef] [PubMed]
56. Anderson, M.; Jiang, J. Teens, social media & technology 2018. *Pew Res. Cent.* **2018**, *31*, 1673–1689.
57. European Commission. *Digital Competence Framework for Citizens, DigComp Framework—European Commission*; European Commission: Brussels, Belgium, 2018.
58. Eurostat. Single Integrated Metadata Structure (SIMS) Guidelines, Version 2.0. 2015. Available online: <https://ec.europa.eu/eurostat/web/metadata/reference-metadata-reporting-standards> (accessed on 23 October 2024).
59. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 7th ed.; Pearson Education: London, UK, 2010.
60. Razali, N.M.; Wah, Y.B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Villiers, and Anderson-Darling tests. *J. Stat. Model. Anal.* **2011**, *2*, 21–33.
61. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 6th ed.; Sage Publications Limited: Thousand Oaks, CA, USA, 2024.
62. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 7th ed.; Pearson: New York, NY, USA, 2019.
63. Pallant, J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*; Routledge: Oxfordshire, UK, 2020.
64. Bryman, A.; Cramer, D. *Quantitative Data Analysis with IBM SPSS 17, 18 & 19: A Guide for Social Scientists*; Routledge: Oxfordshire, UK, 2012.
65. The Analysis Factor. *Factor Analysis: A Short Introduction, Part 2-Rotations*. Available online: <https://www.theanalysisfactor.com/rotations-factor-analysis/> (accessed on 20 January 2025).
66. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 8th ed.; Cengage Learning: Boston, MA, USA, 2019.
67. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* **1974**, *39*, 31–36. [CrossRef]

68. Auerswald, M.; Moshagen, M. How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychol. Methods* **2019**, *24*, 468. [CrossRef]
69. Jolliffe, I.T. *Principal Component Analysis for Special Types of Data*; Springer: New York, NY, USA, 2002; pp. 338–372.
70. Subramanian, S.V.; Jones, K.; Duncan, C. *Multilevel Methods for Public Health Research*; Neighborhoods and Health; Oxford University Press: New York, NY, USA, 2003; pp. 65–111.
71. Diez Roux, A.V. Investigating neighborhood and area effects on health. *Am. J. Public Health* **2001**, *91*, 1783–1789. [CrossRef]
72. Hosmer Jr, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
73. Menard, S.W. *Logistic Regression: From Introductory to Advanced Concepts and Applications*; Sage: Newcastle upon Tyne, UK, 2010.
74. Hilbe, J.M. *Practical Guide to Logistic Regression*. CRC Press/Taylor & Francis Group: Boca Raton, FL, USA, 2016.
75. Shipe, M.E.; Deppen, S.A.; Farjah, F.; Grogan, E.L. Developing prediction models for clinical use using logistic regression: An overview. *J. Thorac. Dis.* **2019**, *11* (Suppl. S4), S574. [CrossRef]
76. PySAL. Exploratory Spatial Data Analysis—Esda v2.7.0 Manual. Available online: <https://pysal.org/esda/> (accessed on 7 February 2025).
77. Kass, G.V. An exploratory technique for investigating large quantities of categorical data. *Appl. Stat.* **1980**, *29*, 119–127. [CrossRef]
78. Rokach, L.; Maimon, O. Data mining with decision trees: Theory and applications. *World Sci.* **2008**, *69*, 264. [CrossRef]
79. Van Deursen, A.J.; Helsper, E.J. The third-level digital divide: Who benefits most from being online? In *Communication and Information Technologies Annual*; Emerald Group Publishing Limited: Leeds, UK, 2015; Volume 10, pp. 29–52.
80. UNESCO. *Youth, Digital Skills and the Future of Work: Bridging the Digital Divide*; UNESCO: Paris, France, 2022; Available online: <https://unesdoc.unesco.org/> (accessed on 9 October 2024).
81. European Commission. *Digital Economy and Society Index (DESI) 2021*; European Commission: Brussels, Belgium, 2021; Available online: <https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2021> (accessed on 19 October 2024).
82. Hastie, T.; Tibshirani, R.; Friedman, J.; Franklin, J. The elements of statistical learning: Data mining, inference and prediction. *Math. Intell.* **2005**, *27*, 83–85. [CrossRef]
83. Baltagi, B.H. *Econometric Analysis of Panel Data*, 6th ed.; Springer: Berlin/Heidelberg, Germany, 2021.
84. Twenge, J.M.; Haidt, J.; Joiner, T.E.; Campbell, W.K. Underestimating digital media harm. *Nat. Hum. Behav.* **2020**, *4*, 346–348. [CrossRef] [PubMed]
85. Vuorikari, R.; Jerzak, N.; Karpinski, Z.; Pokropek, A.; Tudek, J. *Measuring Digital Skills Across the EU: Digital Skills Indicator 2.0*; Joint Research Centre Publications Office of the European Union: Brussels, Belgium, 2022; pp. 4–26.
86. Shi, D.; Lee, T.; Terry, R.A. Revisiting the model size effect in structural equation modeling. *Struct. Equ. Model. A Multidiscip. J.* **2018**, *25*, 21–40. [CrossRef]
87. Rosseel, Y. Small sample solutions for structural equation modeling. In *Small Sample Size Solutions*; Routledge: Oxfordshire, UK, 2020; pp. 226–238.
88. Agresti, A. *An Introduction to Categorical Data Analysis*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2018.
89. ESRI. ArcGIS Esri’s Enterprise Geospatial Platform. 2012. Available online: <https://www.esri.com/en-us/arcgis/geospatial-platform/overview> (accessed on 5 November 2024).
90. Milanović, M.; Stamenković, M. CHAID decision tree: Methodological frame and application. *Econ. Themes* **2016**, *54*, 563–586. [CrossRef]
91. Mienye, I.D.; Jere, N. A Survey of Decision Trees: Concepts, Algorithms, and Applications. *IEEE Access* **2024**, *12*, 86716–86727. [CrossRef]
92. Heinlein, P.; Hartleben, P. *The Book of IMAP: Building a Mail Server with Courier and Cyrus*; No Starch Press: San Francisco, CA, USA, 2008. (In English)
93. Checkland, P. *Systems Thinking, Systems Practice*; Wiley: Chichester, UK, 1981.
94. Hynes, W.; Lees, M.; Müller, J.M. *Systemic Thinking for Policy Making*; OECD: Paris, France, 2020.
95. Leischow, S.J.; Milstein, B. Systems thinking and modeling for public health practice. *Am. J. Public Health* **2006**, *96*, 403–405. [CrossRef] [PubMed]
96. Paina, L.; Peters, D.H. Understanding pathways for scaling up health services through the lens of complex adaptive systems. *Health Policy Plan.* **2012**, *27*, 365–373. [CrossRef]
97. Meadows, D.H. *Thinking in Systems: A Primer*; Chelsea Green Publishing: White River Junction, VT, USA, 2008.
98. Hsiao, C. *Analysis of Panel Data*; Cambridge University Press: Cambridge, UK, 2022.
99. Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*; MIT Press: Cambridge, MA, USA, 2010.
100. Berigel, M.; Boztaş, G.D.; Rocca, A.; Neagu, G. A model for predicting determinants factors for NEETs rates: Support for the decision-makers. *Socio-Econ. Plan. Sci.* **2023**, *87*, 101605. [CrossRef]
101. Iorfino, F.; Oliveira, R.; Cripps, S.; Marchant, R.; Varidel, M.; Capon, W.; Crouse, J.J.; Prodan, A.; Scott, E.M.; Hickie, I.B.; et al. A prognostic model for predicting functional impairment in youth mental health services. *Eur. Psychiatry* **2024**, *67*, e87. [CrossRef]

102. Choudhary, H.; Bansal, N. Addressing digital divide through digital literacy training programs: A systematic literature review. *Digit. Educ. Rev.* **2022**, *41*, 224–248. [CrossRef]
103. Helsper, E.J. The social relativity of digital exclusion: Applying relative deprivation theory to digital inequalities. *Commun. Theory* **2017**, *27*, 223–242. [CrossRef]
104. Eurostat. *Labour Force Survey (LFS)—Regional NEET Rates and Digital Indicators by NUTS 2/3 Regions*. 2024. Available online: <https://ec.europa.eu/eurostat> (accessed on 20 October 2024).
105. Anselin, L. Spatial econometrics. In *Handbook of Spatial Analysis in the Social Sciences*; Newcastle University: Newcastle upon Tyne, UK, 2022; pp. 101–122.
106. Fotheringham, A.S.; Brunson, C.; Charlton, M. Geographically weighted regression. *Sage Handb. Spat. Anal.* **2009**, *1*, 243–254.
107. LeSage, J.; Pace, R.K. *Introduction to Spatial Econometrics*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2009.
108. OECD. *Empowering Youth in a Digital Era; Empowering Young Children in the Digital Age* | European School Education Platform; OECD Publishing: Paris, France, 2023.
109. van Dijk, J. *The Digital Divide*; Polity Press: Cambridge, UK, 2020.
110. Livingstone, S.; Mascheroni, G.; Staksrud, E. European research on children’s internet use: Assessing the past and anticipating the future. *New Media Soc.* **2018**, *20*, 1103–1122. [CrossRef]
111. Duke, B.; Grigorescu, A.; Lincaru, C.; Ciuca, V.; Dragomir, M.S. No One Left Behind: Enabling Digital Transformation Of Human European Workers. *Transform. Bus. Econ.* **2024**, *23*, 664.
112. Binsaeed, R.H.; Yousaf, Z.; Grigorescu, A.; Samoila, A.; Chitescu, R.I.; Nassani, A.A. Knowledge Sharing Key Issue for Digital Technology and Artificial Intelligence Adoption. *Systems* **2023**, *11*, 316. [CrossRef]
113. Faggian, A.; Gemmiti, R.; Jaquet, T.; Santini, J. Regional economic resilience: The experience of the Italian local labor systems. *Ann. Reg. Sci.* **2018**, *60*, 393–410. [CrossRef]
114. Eurofound. *Youth Integration in the EU: Navigating Digitalisation and Labour Shortages—Background Paper*; Eurofound: Dublin, Ireland, 2024.
115. Magidson, J. The CHAID Approach to Segmentation Modeling: Chi-squared Automatic. *Adv. Mark. Res.* **1994**, *118*.
116. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 2, pp. 403–413.
117. IBM Corp. IBM SPSS Decision Trees 29 Documentation. 2022. Available online: <https://www.ibm.com/docs/en/spss-statistics/29.0.0> (accessed on 20 October 2024).
118. Aroldi, P.; Colombo, F. Media, Generations, and the Platform Society. In *Human Aspects of IT for the Aged Population. Healthy and Active Aging*; HCII 2020. Lecture Notes in Computer Science; Gao, Q., Zhou, J., Eds.; Springer: Cham, Switzerland, 2020; Volume 12208. [CrossRef]
119. European Commission. *2030 Digital Compass: The European way for the Digital Decade*; European Commission: Brussels, Belgium, 2025; Available online: <https://transition-pathways.europa.eu/policy/2030-digital-compass-european-way-digital-decade> (accessed on 4 October 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.