Bridging values: The inclusion of young generations in computing

Anna Szlavi[1], Serena Versino[2], Irene Zanardi[3], Karolina Bolesta[4], Letizia Jaccheri[5]

[1] Norwegian University of Science and Technology
[2] University of Pisa
[3] Università della Svizzera Italiana
[4] Warsaw School of Economics
[5] Norwegian University of Science and Technology
anna.szlavi@ntnu.no
serena.versino@phd.unipi.it
irene.zanardi@usi.ch
kboles@sgh.waw.pl
letizia.jaccheri@ntnu.no

Abstract. There is a constantly growing need for skilled professionals in the computing field, which poses challenges for finding the right people for the job. According to the 2022 Digital Economy and Society Index, 55% of companies have problems filling their tech positions. At the same time, the computing sector is going through a diversity crisis, as the majority of its players are Global Northern, heterosexual, white, able-bodied men. Technology permeates our lives, so a lack of diversity in the tech industry, especially when designing software, can lead to bias and exclusionary user experiences. As a consequence, we need to attract young people - for instance, Generation Z (GenZ), born between the mid-1990s and the 2010s - to computing majors. Moreover, there is a need for actions with a retention plan and a strategy to guide a more diverse group toward leadership roles both in academia and industry. Even though the awareness about Diversity, Equity, and Inclusion (DEI) is continually being raised, interventions that focus on inclusiveness are still necessary. With the present paper, we aim to contribute to a better alignment of how to design interventions for including younger people in computing. According to research, GenZ cares about social values and a meaningful contribution to society, that is, DEI, as part of their work. In this paper we are presenting an intervention project, designed to increase DEI in computing, as part of which we collected testimonials by stakeholders working in computing. As a quality check, we performed content analysis after the completion of the project, to investigate to what extent the experiences listed by CS professionals and the interests of GenZ align with one another. Applying multiple methods of cross-checking, we confirmed the presence of social aspects in the lived experiences of CS professionals. Findings show that professionals in the field recognize computing's social embeddedness, which aligns with younger students' values and expectations and confirms that computing is a valid choice to achieve their goals of making a positive change in society. This study is part of a larger effort proposed and realized by EUGAIN, a Horizon Europe-sponsored COST Action, whose purpose is to create a European network that enhances gender balance and diversity in the field of computing.

Keywords: gender balance · DEI · computing · intervention · generation Z

1 Introduction

Technology has become so widespread nowadays that it penetrates people's daily lives. This, on the one hand, results in a constantly growing need for skilled professionals in the computing field [1]. According to the 2022 Digital Economy and Society Index, 55% of companies have problems filling their tech positions [2]. On the other hand, due to technology's ubiquitous role, it is essential to create software which is useful and usable for a wide range of people. This requires diverse development teams so team members can check each other's blind spots and avoid bias [3]. Studies show that diversity also leads to greater creativity and success [4]. Despite this, computing is a homogeneous field: the workforce predominantly constitutes white, cis-gendered, heterosexual, Global Northern, young, middle-class men [5]. Women constitute only 19% of the computing job market within Europe [6].

Even though the awareness about Diversity, Equity, and Inclusion (DEI) in computing is continually being raised, interventions that focus on inclusiveness are still necessary. In order to tackle the growing demand for computing professionals and to achieve DEI, we need to reach young people - for instance, GenZ, born between the mid-1990s and the 2010s [7] - to join computing majors.

In this paper we are presenting an intervention project, proposed and realized by EU-GAIN [8], a Horizon Europe-sponsored COST Action, whose purpose is to create a European network that enhances gender balance, and broadly DEI, in the field of computing. Within the frames of this project, we have designed an initiative that focuses specifically on the better involvement of young people in computing, through short interviews with role models, posted on social media. The project proved to be one of the most successful interventions of the Action, achieving promising results in terms of total views. Therefore, the authors decided to build a content analysis-focused research study on it, in order to understand which key features of this novel intervention intended for GenZ have been successful.

In this study, we are analyzing the testimonials collected for the intervention project, which were produced by stakeholders working in computing, in order to identify the key features describing their experiences in the field. We extracted 31 keywords and grouped them in five macro themes: *Computing*, *Academia*, *Skills*, *Society* and *Gender*. Then, we checked the robustness of our findings with different methods, and confirmed them. The propose of our examination, which marries data mining and feature extraction methods, is to check the content of the testimonials, more precisely to see how much the perception of computing and the interest of GenZ are consistent with each other. Our goal is to contribute to a better understanding of how to design interventions that target younger generations, with the purpose of achieving DEI in computing.

[1] "The word computing refers to a goal-oriented activity requiring, benefiting from, or associated with the creation and use of computers. [...] computing includes a variety of interpretations such as designing and constructing hardware and software systems for a wide range of purposes: processing, structuring, and managing various kinds of information; problem solving by finding solutions to problems or by proving a solution does not exist; making computer systems behave intelligently; creating and using communications and entertainment media; and finding and gathering information relevant to any particular purpose." [1]

2 Background

2.1 Inclusion in Computing

Gender balance is one of the main challenges of the 21stcentury. According to the UN, one of the Sustainable Development Goals (UN SDG5) is to achieve gender equality [9]. Women are at a disadvantage in a number of areas compared to men. For example, girls are forced to leave education earlier than boys [10], growing up they are less likely to get well-paid jobs [11] or be in leadership roles [10], and they are more often victims of domestic violence [12]. When it comes to the digital world, women are underrepresented in both technical and decision-making roles, which leads to software that excludes the female perspective and reinforces binary gender hierarchies [13]. Being a complex and deeply entrenched problem, gender equality requires concerted efforts in multiple areas of life, including technology.

There is an increasing awareness of the need for inclusion in computing. In fact, tech companies often join the DEI conversation because they realized it is beneficial for their image as an employer. As a consequence, there are a number of initiatives whose purpose is to channel a more diverse crowd into computing education, as well as actions that focus on retention and strategies that guide a more diverse group toward leadership roles in both academia and industry. Role models, networking, and mentoring are among the key strategies to involve non-stereotypical players into the field. In the following, we will present one example for each of the above approaches to increase DEI through the gender perspective.

Firstly, Girls' Day is an international program and campaign, running annually from the early 2000s, with the aim of promoting STEM fields to girls in primary and secondary education mainly through role models (see: in Germany, in Hungary, and in Australia for example[2]). Every year a high number of universities, research institutes, and corporations in the STEM field offer to host and organize programs for girls within the frames of this one-day event. The goal is to give girls a first-hand experience about studying and/or working in the field, not just regarding the environment and the types of activities the different jobs involve, but also through meeting relatable role models in the field [14]. Studies show that role models, especially non-stereotypical role models, have a crucial impact on making computing accessible for women [15–18].

Secondly, to raise awareness about bias within higher education, Girl Project Ada[3] aims to contribute to the effort of helping more women to graduate in the computing field. Ada, operating at the Norwegian University of Science and Technology, organizes social and career events for women, so it puts a lot of attention to networking as its tool. Networking and the visibility of female peers have shown to enhance retention [19].

Thirdly, Women STEM UP[4], a European collaboration project with five partner institutions, aims to increase DEI in STEM – and mainly computing – education through a mentorship scheme. By acknowledging the positive effect of female mentors on the success of female students in computing [20], the initiative wishes to facilitate retention and contribute to a growing awareness of DEI in computing education through its mentorship and inspiration academy.

---

[2] Germany:`https://www.girls-day.de/ueber-den-girls-day/was-ist-der-girls-day2/english`; Hungary: `https://lanyoknapja.hu`; Australia: `https://www.gdostem.com.au`

[3] see link `https://www.ntnu.edu/ada`

[4] see link: `https://www.women-stem-up.eu`

2.2 EUGAIN

To achieve DEI goals in computing, in 2020 Horizon Europe sponsored EUGAIN[5], a 4-year COST Action, whose goal is to improve gender balance in the field of computing through creating and strengthening a multi-cultural European community of academics. By doing so, it aims to use the tools of role models, networking, and mentoring, as described in the previous section, to enhance DEI. Based primarily in academia, the initiative has five working groups, focusing on multiple aspects of DEI in computing. EUGAIN wishes not only to map and invigorate the different areas of computing education, it also wishes to connect the findings to the industry, as well as to design interventions for inclusion. Starting in November 2020, EUGAIN has over 150 members from 39 European countries as of November 2022, at the half-time of the project.

Inclusiveness necessitates the acknowledgement that gender is only one dimension of an individual's identity, and as such, exclusion and bias cannot be eradicated with a single focus on binary gender. As the Theory of Intersectionality points out, the overlapping dimensions of identity, such as gender, sexuality, socio-economic status, ethnicity, religion, nationality, dis/ability, and education, affect one's experiences as a whole [21]. The combination of these factors can account for the individual's unique disadvantages and challenges; thus, inclusiveness can only be achieved with a shift in approach.

In fact, the need for more workforce and more DEI in computing underscores the urgency of finding better ways to attract young people to the computing field. GenZ, also called as digital natives [22], has a pronouncedly different view on the digital world, as well as on social values, than previous generations [23]. Due to their early exposure to the internet and digital technologies, they are much more aware of social issues and global problems than other generations [24], appreciating and expecting the values of DEI from their work environments [25]. They cherish career paths that allow them to realize themselves and meaningfully contribute to social issues according to their values [26].

In line with this, the Action puts great emphasis on supporting younger generations within the field. Early career academics receive support through grants, joint projects, networking, and mentorship within the Action. This way, EUGAIN aims to serve as an intervention that helps the retention of young people in computing and contribute to a more diverse leadership within academia and industry. In the summer of 2022 a new role was created within the Core Group to better fit with this specific goal. Even if it is not a tradition in COST projects, this Action decided to appoint a Young Researcher and Innovator Coordinator, in order to more efficiently manage the group of about 50 PhDs and postdocs within the community called Young Researchers and Innovators (YRIs). Thanks to the dedicated efforts, the Action now has an active early career group, even with people from GenZ.

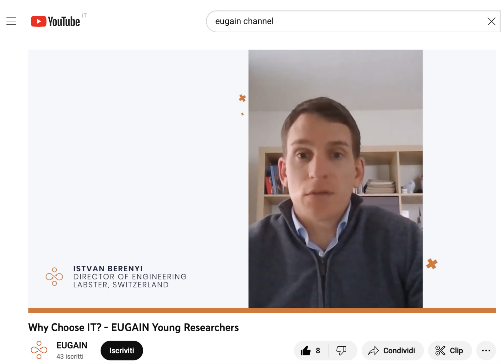
2.3 Project GenZ

One of the recent projects carried out by early career researchers within EUGAIN[6] has the aim to design an intervention that attracts and includes young people in the field. Acknowledging the importance of role models for thriving in computing [16–18], this group of young researchers wanted to reach out to GenZ by creating short visual content: interviews about primarily young role models and mentor figures. Over 30 testimonials were collected from diverse stakeholders from the field, both from industry and academia, all over Europe.

[5] The fifth author is the Chair of the present Action.
[6] The first four authors consisted of this team.

We designed a library of brief videos, shortening each of the testimonials, in order to share them on social media, such as YouTube[7] and TikTok[8], the main channels for GenZ [27]. By this, the project had the intention of boosting both the visibility and the desirability of the field for younger people, portraying it as an inclusive field.

The videos were posted on EUGAIN's YouTube channel between December 2022 and February 2023. Parallel to this, a TikTok account was also created, so that the videos could be posted there too between January 2023 and February 2023. Fig. 1 shows the examples which were created for the YouTube channel (i.e., horizontal longer videos) and the TikTok social platform (i.e., vertical shorter videos). The latter example can be used for the YouTube channel in the "shorts" section as well.



(a) YouTube: longer video

(b) TikTok: shorter video

Fig. 1: Video examples

An example of a shorter video for the TikTok channel or YouTube Shorts is shown in Fig. 1b. One of the videos posted on YouTube's Shorts section, featuring specifically GenZ stakeholders (master students of CS), reached over 300 views over a weekend. Within three days of being posted, our first TikTok video had about 255 views (i.e., it collected about 85 views on average per day). According to statistics, in 2022 young women between the ages of 18 and 24 made up 24% of TikTok's global audience (i.e., three billion downloads globally), while male users in that same age group make up roughly 18% of the app's users [28]. These statistics shed light on the possible visibility we may reach by using this social network.

Due to the popularity of the intervention, the authors saw an opportunity in analyzing the testimonials and their potential to motivate young generations in choosing computing. With the present paper, we aim to contribute to a better alignment of how to design interventions for including younger people in computing.

[7] see link `https://www.youtube.com/@eugain7063`

[8] see account @eugaincost

3 Methodology

According to research, GenZ is concerned with social values and making a significant contribution to society [29], specifically DEI, as part of their profession. In order to identify and classify the main characteristics that best describe experiences in the sector, we created a data collection of testimonies gathered by the YRIs of EUGAIN during the Project GenZ. In this study, we want to investigate to what extent the recognized themes and interests of GenZ align with one another. For the analysis of transcripts, we applied text mining approaches that enables to analyze textual data for the extraction of meaningful information and patterns present in natural language text originally organized as unstructured data [30].

3.1 Data collection process

During the Project GenZ, YRIs of EUGAIN collected 31 testimonials between October and December 2022. Stakeholders were asked to described their professional path and what role DEI play in their perception of it in brief, 1-2 minute videos. The participants work in the computing field, both in academia and industry, and they cover a variety of positions such as professors, researchers, students, managers, UX professionals, and so on. The data collection includes diverse profiles in gender, job position, country, and specialization. We asked them to share valuable information that high school students can use to make an informed decision when choosing their educational path. To collect a good variety of answers, we provided them a list of five questions, explaining the primary motivations for entering and staying in the field, the significant experiences, the work environment, the advantages, and the long-term objectives.

3.2 Data processing

We analyzed their texts to learn what professionals think is important to share. We wanted to see what the recurring topics were by extracting keywords, and if the keywords were social-related, to see if professional perceptions of computing aligned with the values sought by GenZ. To accomplish this, we used term frequency count and feature extraction after cleaning and pre-processing the texts.

Data cleaning and lexical analysis. From the video recordings, we generated the transcripts[9] and manually fixed any misspelling we discovered. We calculated the length of each transcript as the count of characters in the strings. They are 741 characters length on average (Mdn=625). We normalized the distribution of transcript lengths by number of transcripts, and the bimodal distribution revealed two peaks, the highest ranged between 250 and 750 characters. Then, in the lexical analysis, we pre-processed the data. We standardized the corpus appearance, reduced data sparseness and applied data cleaning approaches (e.g., converting the text into lower case characters, removing punctuation, unnecessary symbols and stop words[10]). In the Tokenization process[11], the lexical analyzer reads the source program's stream of characters from left to right, chop them into tokens and retrieve the terms used

[9] Python library SpeechRecognition 3.9.0

[10] Stop words are frequent terms that don't offer meaningful information for the task [30].

[11] The Tokenization is the process of dividing the input text into collections of characters that have a collective meaning (called tokens) [30].

in the corpus which then, have been tagged using the Part-of-speech (POS) technique[12]. At this stage, the root form of the word called lemma[13], is altered by the presence of prefixes or suffixes that define its grammatical purpose. Therefore, in order to reduce the dimensionality of the description of documents within the collection (i.e., the vocabulary), we applied the Lemmatization approach[14]. Once we removed the inflectional morphology, we built the vocabulary with lemmas [30]. We analyzed the frequencies of the nouns [31] in the vocabulary and thus, we identified the primary keywords of the domain. As follows, we grouped them into five themes. *Computing* theme contains keywords that are concerned with technical aspects of computing field, such as the development of cutting-edge technologies. *Academia* theme groups keywords related to the educational background and research activities in the field. *Skills* theme refers to the use of problem-solving techniques, the pooling of ideas and the creativity aimed at facing technological challenges. *Society* theme includes keywords about the main driver that led participants to get into the field, that is the desire to make an impact on the global community and the enhancement of the people's quality of life. Finally, *Gender* theme concerns women and includes narratives about the gender-based barriers perceived by stakeholders in choosing a career in computing.

Feature extraction. Separately, we adopted different natural language processing (NLP) approaches for both the processing of data and the clustering analysis, in order to compare the results with the frequency count. In this procedure, we transformed input sentences by using Count Vectorizer[15] and we represented each document in a numeric matrix using the Bag-of-words technique[16]. Therefore, we created a vector representation per document and a weight has been assigned to each lemma based on its frequency in the document. We adopted the TF-IDF[17] as weighting factor. In the last stage, we applied the Principal Component Analysis (PCA)[18] for reducing the dimensionality of data and noise in the K-means cluster analysis [32]. The optimal number of clusters (K) has been defined by calculating and comparing the Silhouette Index, the Calinski-Harabasz Index and the Elbow method [33], the latter provided the optimal K. We calculated the Sum Square Error (SSE)[19] of each cluster and in correspondence of the fifth one, the curve formed an angle ("elbow") where the distortion value had the largest decrease (at 0.43 score) reporting the optimal K equal to 5. Finally, in order to confirm our results, we used the keywords search to find relationships between the five themes and the K-means clusters.

[12] Part-of-speech tagging (POS) determines the part of speech tag, e.g. noun, verb, adjective, etc. for each term [30].

[13] A lemma or lexeme or canonical form is a single dictionary entry with a single meaning [30]

[14] Lemmatization methods map verb forms to the infinite tense and nouns to the singular form [30].

[15] Scikit learn Python package

[16] In the Bag-of-words (BOW) representation, text (such as a sentence or a document) is treated as an unordered collection of words by using a fixed length sparse vector containing word occurrence counts [30].

[17] TF-IDF (Term Frequency–Inverse Document Frequency) is a numerical statistic which reveals that a word is how important to a document in a collection [30].

[18] The Principal Component Analysis (PCA) is used to project the original feature space onto a lower dimensional subspace including a subset of variables from a larger set, based on which original variables have the highest correlations with what is called principal components [32].

[19] The Sum Square Error is the sum of the average Euclidean Distance of each point against the centroid [33].

4 Findings

In this section we summarize results from the frequency and feature extraction analysis. At first, we examined the vocabulary of the whole corpus, we identified its key features by frequency count and grouped them in themes by manually evaluating examples from the corpus. Apart, we applied different techniques for data clustering (i.e., PCA and K-means). Finally, we analyzed the vocabulary of each cluster and the keywords search allowed us to confirm the frequency count results.

Frequency analysis. After the corpus normalization, we built a word cloud[20] from the vocabulary in order to visualize the most frequent terms. As shown in Fig. 2, where the frequency of each lemma is displayed with font size and color, the most recurring words were informatics, computer, and person. We sorted the frequencies in decreasing order and established a threshold a priori (N=5): lemmas that occurred more than five times have been considered as representative, otherwise they have been excluded.

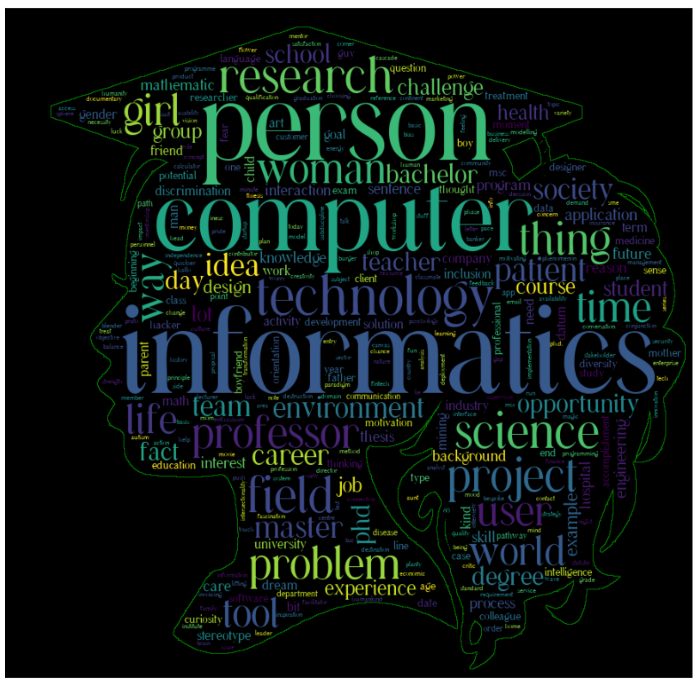


Fig. 2: Vocabulary

Therefore, we analyzed 45% of the 729 total occurrences of lemmas present in the vocabulary and 13% of the 326 total lemmas. From the frequency analysis we detected 31 keywords. After that, we conducted a corpus search and categorized them into five major themes. Tab. 1 reports some examples extracted from the corpus. Tab. 2 shows the frequency distribution of keywords, and the percentage distribution of occurrences and lemmas over the threshold by theme.

[20] In text analytics, word clouds provide an overview by distilling text down to words that appear with highest frequency [34].

Table 1: Themes, keywords and examples from the corpus.

Theme	**Keywords**	**Examples**
Computing	informatics, computer, technology, science, field, project, user, tool, career, job	*"I work in health informatics and mainly I am focused on a subject that is called process mining that is a subdiscipline of data mining and our goal is finding workflow activity from data....", "..very powerful tool that can be used to create user-friendly.."*
Academia	research, professor, master, bachelor, degree, teacher, phd, school, student, course	*"..my work is doing research in the area of artificial intelligence..", "My professors helped me to find the right PhD program.."*
Skills	problem, idea, team, challenge, group	*"I really enjoy working in IT because you can really solve complex problems..", "At the same time, we can work together sharing ideas.."*
Society	person, world, life, society	*"How to develop solutions that improve the life of people..", "I wanted to make a change in the world.."*
Gender	woman, girl	*"In high school I was in a class of 15 girls with non-scientific orientation..", "I believe girls/women are discouraged from studying Informatics at a very early age.."*

Table 2: Frequency distribution by theme

Theme	**Keywords**	**Occurrences**	**Lemmas**
Computing	10	33%	24%
Academia	10	20%	24%
Skills	5	9%	12%
Society	4	12%	10%
Gender	2	5%	5%
not relevant	(-)	21%	25%
Total	**31**	**100%**	**100%**

As we can see, *Computing* includes 10 keywords representing about 33% of total occurrences and 24% of total lemmas. This theme is the most representative in terms of occurrences. *Academia* groups 10 keywords representing about 20% of total occurrences and 24% of total lemmas. *Skills* and *Society* show similar frequencies, 9% and 12% of total occurrences and 12% and 10% of total lemmas respectively. Overall, *Gender* is the less frequent (i.e., 2 keywords representing about 5% of total occurrences and 5% of total lemmas). Finally, *Not relevant* groups conversational terms - that are not meaningful for the task - and cross-theme lemmas such as "opportunity" (e.g., *Computing-Skills* in *"..the opportunity of creating through technology attracted my interest.."*).
Tab. 3(a,b,c,d,e) show the distribution of keywords' occurrences by theme in percentage. According to Tab. 3(a), "informatics" is the most representative over 10 keywords in *Com-*

puting (i.e., 20% of the 118 total occurrences). *Academia* in Tab. 3(b) shows that "research" is the most representative over 10 keywords (i.e., 17% of the 69 total occurrences). Tab. 3(c) refers to *Society* and shows that "person" is the most representative over 4 keywords (i.e., 48% of the 42 total occurrences). However, "person" was used in opposite cases referring to both encouraging or discouraging situations. In *Skills*, Tab. 3(d), "problem" is the most representative keyword over 5 (i.e., 30% of the 33 total occurrences). Finally, in Tab. 3(e), *Gender*'s keywords are equally distributed over 18 total occurrences. Additionally, scrolling through the absolute frequencies of the vocabulary in decreasing order, we built Tab. 3(f) that reports the most frequent keywords by theme, in the following order: "informatics" in *Computing*, "person" in *Society*, "research" in *Academia*, "problem" in *Skills* and "woman" in *Gender*.

Table 3: Percentage distribution of keywords by theme and most frequent keywords

Theme	**Keyword**	**Occurrences**
Computing	informatics	20%
	computer	14%
	technology	13%
	science	10%
	field	10%
	project	9%
	user	8%
	tool	7%
	career	5%
	job	4%
	Total	**100%**

(a) Tot. Occurrences=118

Theme	**Keyword**	**Occurrences**
Academia	research	17%
	professor	14%
	master	10%
	bachelor	9%
	degree	9%
	teacher	9%
	phd	9%
	school	9%
	student	7%
	course	7%
	Total	**100%**

(b) Tot. Occurrences=69

Theme	**Keyword**	**Occurrences**
Society	person	48%
	world	19%
	life	19%
	society	14%
	Total	**100%**

(c) Tot. Occurrences=42

Theme	**Keyword**	**Occurrences**
Skills	problem	30%
	idea	21%
	team	18%
	challenge	15%
	group	15%
	Total	**100%**

(d) Tot. Occurrences=33

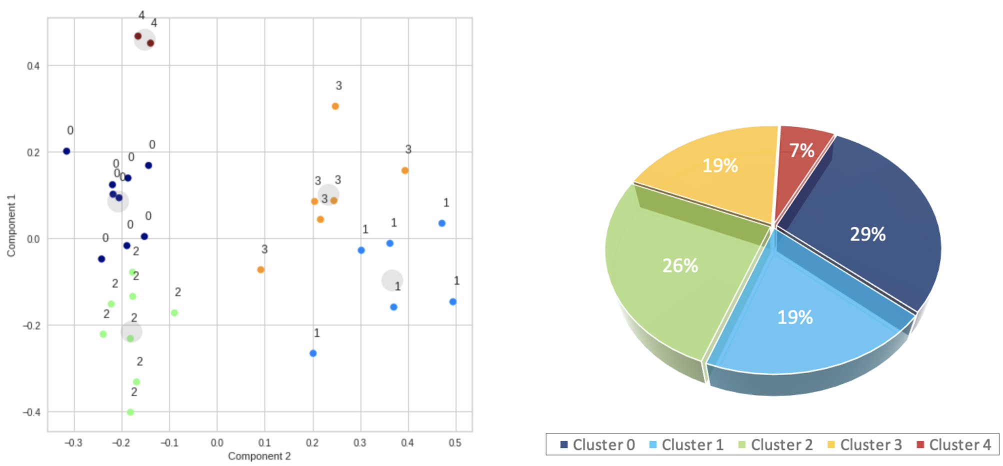
Theme	**Keyword**	**Occurrences**
Gender	woman	50%
	girl	50%
	Total	**100%**

(e) Tot. Occurrences=18

Keyword	**Nr. Occurrences**
informatics	24
person	20
research	12
problem	10
woman	9

(f) Most frequent keywords per theme

Feature Extraction. As shown in Fig. 3a, the y-axis and x-axis represents, respectively, the first and second component we identified with the PCA. The plot shows the results of the K-means clustering; data points are grouped by clusters with different colors and the grey circles are the centers[21]. As we can see, the clustering analysis identified five classes as the analysis of frequencies.



(a) K-means clustering (b) % Distribution of transcripts by cluster

Fig. 3: Clusters (K=5)

As shown by Fig. 3b, the largest clusters are 0 and 2, representing about 55% of the corpus overall. At this stage, we used the same approach as before to assess the frequencies of lemmas by cluster. We built a vocabulary per each cluster, we sorted the frequencies in decreasing order and fixed the threshold a priori (N = 2): lemmas that appeared at least two times were considered indicative of the cluster; otherwise, they were excluded. Tab. 4 shows the percentages of occurrences and lemmas that we have analyzed by cluster, for instance, we analyzed 62% of 260 total occurrences and 31% of 143 total lemmas for Cluster 0.

Table 4: Distribution of occurrences and lemmas by cluster

Cluster	**Tot. Occurrences**	**Occurrences (>N)**	**Tot. Lemmas**	**Lemmas(>N)**
0	260	62%	143	31%
1	105	53%	65	25%
2	253	56%	158	30%
3	124	58%	79	34%
4	42	36%	33	18%

We found a match between the set of 31 keywords and the vocabularies of the clusters for a total of 56 lemmas. Tab. 5 shows the bivariate frequency distribution of matching lemmas

[21] Centers or centroids, correspond to the arithmetic mean of data points assigned to the cluster [30]

by cluster and theme. It reports also the conditional distributions by theme and cluster (% Total).

Table 5: Bivariate frequency distribution of lemmas by theme and cluster (Tot. lemmas=56)

Theme	**Cluster 0**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**%Total**
Computing	16%	5%	11%	5%	2%	39%
Academia	7%	4%	11%	9%	(-)	30%
Skills	7%	(-)	4%	(-)	2%	13%
Society	4%	(-)	4%	4%	2%	13%
Gender	(-)	4%	(-)	2%	(-)	5%
%Total	33%	13%	29%	20%	5%	100%

About 33% of total lemmas matched with the set of keywords in Cluster 0, in particular about 16% with the ones of *Computing* theme. About 29% of total lemmas matched with the set of keywords in Cluster 2, in particular about 11% respectively with the ones of *Computing* and *Academia*. About 20% of total lemmas matched with the set of keywords in Cluster 3, in particular about 9% with the ones of *Academia*. Cluster 1 and 4 showed the lowest matching rates. Overall, the conditional distributions by theme show that about 69% of total lemmas matched with *Computing* and *Academia* (i.e., respectively 39% and 30%).

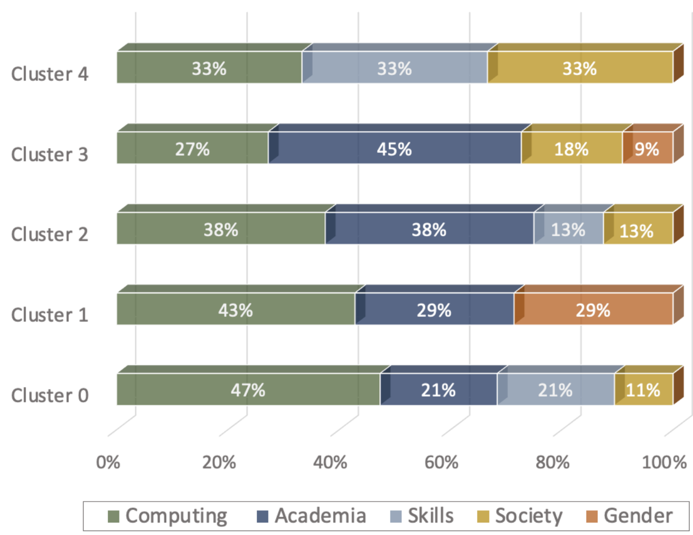


Fig. 4: % Distribution of themes per cluster

Fig. 4 shows the percentage distribution of themes calculated over the total matching lemmas per cluster. *Computing* is dominant in Cluster 0 (47%) and Cluster 1 (43%), and it is present in each cluster. *Academia* is dominant in Cluster 3 (45%). *Computing*, *Skills* and *Society* are equally distributed in the smallest cluster (i.e., Cluster 4). *Gender* is present in

Cluster 1 (29%) and Cluster 3 (9%). In Cluster 2, *Computing* and *Academia* are equally distributed (38% respectively) as *Society* and *Gender* (13% respectively).

5 Discussion

According to Leslie et al. (2021) [29], GenZ cares about social values and a meaningful contribution to society, that is, DEI, as part of their work. In this paper we collected testimonials by stakeholders working in computing in order to extract and group the key features in macro themes when describing experiences in the field. We wanted to check to what extent the identified themes and the interest of GenZ are consistent with each other.

In the analysis of frequencies, 31 keywords were extracted. In the qualitative follow-up, we detected and analyzed samples from the corpus according to keywords. Similarly to the analysis of frequencies, the feature extraction technique produced five classes, and we used the keywords to uncover a correspondence between the themes and the clusters. The clusters turned out to be cross-thematic with some of them having dominant themes.

In line with the analysis of frequencies, *Computing* was significantly represented in all clusters (especially, in Cluster 0, the biggest one). This can be due to the domain of analysis. In addition, the *Academia* theme is present in four out of five clusters, particularly Cluster 3. These two themes tend to co-occur, most likely as a result of how actively scientific research participates in the advancement of modern technology. The *Skills* theme is present in three over five clusters and co-occur with *Society* but not with *Gender*. The co-occurrence with *Society* can be due to the tendency of participants of linking specific *Skills* with the development of technologies that can have an impact on society. The *Society* theme represents the main driver for participants to enter the field, and indeed, it is present in four out of five clusters. *Gender* is the less represented theme over the clusters and it is mainly concentrated in Cluster 1.

The *Society* theme, which is cited in four out of five clusters, and the *Gender* theme, which is mentioned in two out of five clusters, both revealed the presence of social features overall. From the viewpoint of DEI, we can group them into a macro-theme involving social issues and refer to it as the *DEI* theme. We observed that the *DEI* theme includes narratives about women and gender-based barriers in the computing industry as both mentioned in the *Gender* theme. However, the main motivator of *Society*, which is important to potential GenZ students, was perceived as being stronger, encouraging the participants to pursue a career in computing.

6 Conclusion

Nowadays, despite efforts to ensure inclusiveness, only a tiny portion of students select computing. Diversity in education and industry is also far from being realized [35–37], notwithstanding the calls for its enhancement [3, 38]. The misalignment between GenZ and the values associated with computing is one potential factor [29]. In order for them to see computing as a viable career option, it is essential to demonstrate how it can help them realize their goal of having a positive impact on the environment and society through their work.

According to the findings of this study, while GenZ perceives computing as a career that is misaligned with their values, professionals in the field recognize the social embeddedness of their work contributions. As a result, in order to diversify the field by engaging and attracting younger people to computing, we should continue to promote and share the positive impact of the field on society.

In this study, we presented an intervention effort by EUGAIN. We chose to assess the information we were publishing on a social media platform to discover what professionals consider important to share with younger generations, and if the content was naturally connected with their beliefs. In future work, we hope to collect additional testimonials while tackling intersectionality more fully, such as by engaging more non-binary and LGBTQ+ stakeholders to share their experiences and serve as role models. To accomplish DEI in computing, we must promote a varied set of role models who represent an inclusive atmosphere.

7 Acknowledgements

This project was partially supported by COST Action CA19122 - EUGAIN (European Network for Gender Balance in Informatics).

References

1. A. Clear and A. Parrish. Computing Curricula 2020 CC2020, Paradigms for Global Computing Education encompassing undergraduate programs in Computer Engineering, Computer Science, Cybersecurity, Information Systems, Information Technology, Software Engineering with data science. *Association for Computing Machinery (ACM) IEEE Computer Society (IEEE-CS)*, page 23, 12 2020.
2. European Commission. Digital Economy and Society Index (DESI), 2022.
3. Landoni M. Szlávi, A. Human computer interaction-gender in user experience. In *International Conference on Human-Computer Interaction*, pages 132–137. Springer, 2022.
4. Diversity in Tech. The benefits of diversity in tech, 2022.
5. A. Peters et al. Gender-inclusive HCI research and design: A conceptual review. *Found. Trends Hum.-Comput. Interact.*, 13(1):1–69, 2020.
6. European Commission. Women in digital scoreboard, 2021.
7. M. Dimock. Defining generations: Where millennials end and generation z begins. *Pew Research Center*, 17(1):1–7, 2019.
8. COST. Action website, 2023.
9. United Nations. Goal 5: Achieve gender equality and empower all women and girls, 2022.
10. Winthrop R. Sperling, G. B. What works in girls' education: Evidence for the world's best investment. Brooking Institution, 2016.
11. G. Kirkup. ICT as a tool for enhancing women's education opportunities, and new educational and professional opportunities for women in new technologies. United Nations Division for the Advancement of Women (UNDAW), 2002.
12. European Union Agency for Fundamental Rights (FRA). Violence against women: An eu-wide survey, 2014.
13. J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research 81*, page 1–15, 2018.
14. Szlávi A. and P. Bernat. Young women's barriers to choose IT and methods to overcome them – A case study from Hungary. *Teaching Mathematics and Computer Science*, 2020.
15. L. Happe, , et al. Effective measures to foster girls' interest in secondary computer science education. *Educ Inf Technol*, (26):2811–2829, 2021.
16. A. Szlávi. Barriers, Role Models, and Diversity: Women in IT. *Central-European Journal of New Technologies in Research, Education and Practice*, 3(3), Dec. 2021.
17. González-Pérez S., R. Mateos de Cabo, and M. Sáinz. Girls in stem: Is it a female role-model thing? 2020.

18. D. Oyserman, K. Terry, and D. I. Bybee. A possible selves intervention to enhance school involvement. *Journal of adolescence*, 25 3:313–26, 2002.
19. C. Convertino. Nuancing the discourse of underrepresentation: a feminist post-structural analysis of gender inequality in computer science education in the us. *Gender and Education*, 32(5):594–607, 2020.
20. L. A. Lyon and E. Green. Coding boot camps: Enabling women to enter computing professions. *ACM Transactions on Computing Education*, 21(2), 2021.
21. K. Crenshaw. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist policies. *University of Chicago Legal Forum*, 1989(1):139–167, 1989.
22. M. Prensky. Digital natives, digital immigrants part 2: Do they really think differently? *On the horizon*, 2001.
23. M. Fodor and K. Jaeckel. What does it take to have a successful career through the eyes of generation z-based on the results of a primary qualitative research. *International Journal on Lifelong Education and Leadership*, 4(1):1–7, 2018.
24. K. Lanier. 5 Things HR professionals need to know about generation Z: thought leaders share their views on the HR profession and its direction for the future. *Strategic HR review*, 2017.
25. T. Nguyen Ngoc et al. Generation Z job seekers' expectations and their job pursuit intention: Evidence from transition and emerging economy. *International Journal of Engineering Business Management*, 14:18479790221112548, 2022.
26. P. Bohdziewicz. Career anchors of representatives of generation Z: some conclusions for managing the younger generation of employees. *Zarządzanie Zasobami Ludzkimi*, (6 (113)-" Managing Diversity in the Context of International HRM"):57–74, 2016.
27. Media Insight Project. The news consumption habits of 16- to 40-year-olds, 2022.
28. Y. Jiang. Social media platforms and public relations for brand promotion. *BCP Business Management*, 34:769–775, 12 2022.
29. B. Leslie et al. Generation Z perceptions of a positive workplace environment. *Employee Responsibilities and Rights Journal*, 33(3):171–187, 2021.
30. A. Hotho, A. Nürnberger, and G. Paaß. A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, 20(1):19–62, 2005.
31. A. Hulth. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 216–223, 2003.
32. C. Ding and H. Xiaofeng. Cluster structure of k-means clustering via principal component analysis. In Honghua Dai, Ramakrishnan Srikant, and Chengqi Zhang, editors, *Advances in Knowledge Discovery and Data Mining*, pages 414–418, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
33. R. Gustriansyah, N. Suhandi, and A. Fery. Clustering optimization in RFM analysis based on k-means. *Indones. J. Electr. Eng. Comput. Sci*, 18(1):470–477, 2020.
34. F. Heimerl et al. Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii International Conference on System Sciences*, pages 1833–1842, 2014.
35. E. Dias Canedo et al. Barriers faced by women in software development projects. *Information*, 10(10):309, 2019.
36. Statista. Software developer gender distribution worldwide as of 2022, 2022.
37. V. Barr. Gender diversity in computing: are we making any progress?, 2017.
38. A. Serebrenik. *Social software engineering*. Technische Universiteit Eindhoven, November 2022. Intreerede, gehouden op 4 november 2022, aan de Technische Universiteit Eindhoven.