ISSN 2007-9737

A Comprehensive Synthetic Dataset of Simulated RWH User Daily Activities and Preferences

Yilin Li*, Junke Xu, Alireza Dehghani

University College Dublin,
Centre for Applied Data Analytics, Dublin,
Ireland

{yilin.li, junke.xu, alireza.dehghani}@ucd.ie

Abstract. The global work environment has experienced a transformation in recent years, which has been greatly accelerated by technological advances and the growing importance placed on the benefits of remote working, such as reduced emissions, time savings and improved mental health. This shift has contributed to the growing popularity of Remote Work Hubs (RWHs)/Coworking Space, which combine traditional office infrastructures with the flexibility required for modern remote work, catering to a diverse group of people such as entrepreneurs, freelancers and remote workers. This study presents a pioneering approach to generating synthetic datasets using Large Language Models (LLMs) via APIs to bridge the gap of accessible user data. Leveraging the ability of Large Language Models to generate contextually rich and diverse data, we simulate the nuanced activities and decision-making processes of RWH users. This synthetic dataset provides foundational insights for coworking space design, management, and policy support through extensive market research and personal development. Through a well-designed methodology, including persona generation and diary entry synthesis, we provide a comprehensive picture of the daily activities, workplace decisions, and commuting preferences of shared workspace users based on real-world data sources and advanced model configurations.

Keywords. Remote working hubs, synthetic dataset generation, large language models, user behavior simulation.

1 Introduction

The work environment has changed radically in recent years, with remote working becoming a key element of this new pattern.

Advances in technology have accelerated the shift to remote working, and there is a growing recognition of the benefits of remote working, including reducing emissions, saving personal time and improving mental health [8, 11, 6].

At the same time, Remote Working Hubs (RWHs) are becoming increasingly popular, facilitating remote working by providing spaces that combine traditional office structures with the flexibility required for remote working.

The distribution of these hubs across the country, as shown in Fig. 1, demonstrates the increasing popularity and strategic development of RWHs. The importance of remote working hubs is highlighted by their ability to support various work styles and preferences, catering for entrepreneurs, freelancers and remote workers [8, 11].

The demand for coworking spaces has increased rapidly, reflecting a shift in work habits and a growing awareness of the benefits of such environments. This trend is evidenced by data showing a significant increase in the number of coworking spaces globally, from about 160 in 2008 to nearly 19,000 in 2018 [8].

However, the growth of these spaces is not without its challenges, the key issues such as the impact of the coronavirus pandemic, attracting new members and adapting to an uncertain business environment [3].

These insights highlight the importance of a comprehensive understanding of user behaviour and preferences to ensure their continued growth and relevance [23].

Fig. 1. Decarbonizing Ireland's National Hub Network, the current state of the network and the distribution of hubs (2023, Source: connectedhubs.ie/)

RWHs have the potential to decarbonise and reduceof decarbonising and reducing energy consumption and carbon footprint [17, 19]. Using AI/ML, we can simulate complex patterns, predict future trends and develop strategies based on the Sustainable Development Goals (SDGs) [16].

For example, the ML researchers used deep learning algorithms like artificial neural networks (ANN) to evaluate data from smart cities to discover new insights [2, 14]. However, the effectiveness of these techniques depends on the availability and quality of data.

Modelling and analysing these impacts is a complex task, but is hampered by the lack of accessible real-world user data. This data is critical to understanding and enhancing user behaviours in RWHs, promoting energy-efficient operations, and developing sustainable practices.

The main challenge they face is the lack of accessible real-world data on users. This gap significantly impacts the ability to understand and optimise RWHs user interactions and behaviours.

As a result, there is an urgent need for synthetic data that can accurately model the complex behaviours and usage patterns of RWH users, Synthetic data emerges as a good alternative, offering flexibility in research models, enabling detailed exploration of user behaviour to reduce the limitations of manual data collection [7], thus providing a valuable resource for research and development in this area.

In practice, it enables stakeholders to identify and analyse the energy consumption patterns and carbon footprint of RWH use, and facilitates the development of strategies to effectively achieve decarbonisation. In the quest to generate such synthetic data, the accessibility of Large Language Models (LLMs) through APIs has revolutionised the generation of synthetic data and labels, significantly reducing the need for specialised expertise and streamlining the research process [7, 5]. LLMs are capable of generating rich, contextually diverse data that closely correlates with patterns of any context, including human behaviour.

ISSN 2007-9737

1. The Freelancers

This persona represents self-employed individuals who value 24/7 access, quiet spaces for concentration, and networking opportunities to grow their businesses. The working hours can vary greatly among freelancers depending on their work style, workload, and the nature of their work. Some might work more during the weekends or evenings, while others stick to a traditional Monday-to-Friday schedule.

1.1 Independent contractors

These are professionals who work on their own and might have multiple clients at a time. They typically work on a project-by-project basis, offering their specialised services. These freelancers need a quiet space where they can focus on their work without distractions. Since they might meet with clients, they might also appreciate the availability of meeting rooms in the co-working space. They value 24/7 access because their work hours can be irregular depending on the project deadlines and client demands. Networking opportunities are also important to them for finding new clients and collaborators.

1.1.1 Persona 1: The Creative Professional

This persona might be a graphic designer, copywriter, photographer, or video producer. They often work on various projects, switching between clients and industries, so they need a space that fosters creativity and inspiration. A modern interior design with access to natural light, aesthetically pleasing communal spaces, and a vibrant, dynamic atmosphere could be particularly appealing to them. High-quality tech equipment, such as colour-accurate monitors for designers or a quiet room for voice recording for podcasters, might be valuable. Additionally, the ability to network with other creatives for potential collaborations or knowledge sharing can be beneficial.

Liam - Freelance Graphic Designer:

Liam has over five years of experience as an independent graphic designer. He primarily works with long-term clients, typically startups and small businesses, helping them establish and maintain their visual identity. His projects often span weeks or months, allowing him to delve deep into a client's branding strategy. He appreciates a workspace that offers a professional and inspiring environment, with modern interior design and lots of natural light. A stable, high-speed internet connection, access to high-quality printing services, and a workstation capable of running resource-intensive design software are all critical to his work. As he spends significant time in the workspace, comfortable furniture and break-out areas are important to him.

Fig. 2. Example of the persona

This capability is particularly important for modelling the activities and decision-making processes of RWH users, providing insights into everyday life, preferences and interactions in collaborative working environments. This paper aims to bridge the gap in accessible user data for RWHs by introducing a synthetic dataset generated using LLM. The dataset aims to model the complex and intricate behaviours and preferences of RWH users, providing a foundation for enhanced coworking space design, management and policy support.

2 Related Work

Human Activities Synthetic Dataset generation: The generation of synthetic datasets for human activities has become increasingly pivotal in the development and refinement of text classification models across various domains, from online toxicity detection to spam filtering [24, 10].

Due to the resource-intensive nature of data collection and curation processes, the reliance on high-quality training data presents a significant challenge. Researchers have begun exploring language models and LLMs' potential for generating synthetic data tailored to specific tasks [1], augmenting training data in scenarios where real-world data is rare [13].

Large language models: Studies have also delved into the feasibility of creating entire synthetic datasets from scratch to support zero-shot learning. However, the effectiveness of LLM-generated synthetic data in achieving comparable model performance to real-world, annotated data remains a subject of debate [13].

Creating "Tiny Stories" and sampling of knowledge graph triplets for text generation with GPT-3 exemplify efforts to produce high-fidelity synthetic datasets [12]. However, the distinctiveness from real human data underscores the ongoing challenge of bridging the gap between synthetic and authentic human data. The development of conversational agents through synthetic dialogue dataset generation using LLM agents further illustrates the potential of synthetic data in creating realistic simulations of human interactions [20, 26].

Zero-shot Synthetic Data Generation: The emergence of zero-shot learning capabilities within LLMs has significantly advanced the field of synthetic data generation [25]. For example, DreamLLM's ability to generate multimodal content highlights the versatility of zero-shot learning in creating complex documents combining text and image content [4]. The exploration of LLMs as zero-shot planners points to using pre-trained language models to decompose high-level tasks into actionable steps, demonstrating the ability to translate abstract concepts into concrete executable actions [9].

Similarly, the study on synthetic data generation for clinical text mining revealed the effectiveness of LLMs in enhancing text mining tasks in the healthcare domain [21], offering insights into the practical applications and limitations of LLMs in sensitive domains. ZeroGen introduces an efficient approach to zero-point learning through dataset generation, showing the impact of LLM in creating synthetic datasets that are useful for training smaller models for various NLP tasks [25]. The exploitation of zero-shot synthetic data generation is a key development in the use of LLMs to simulate user activities and interactions, especially in domains like Remote Work Hubs (RWH) where data scarcity poses a challenge.

3 Methodology

3.1 The User Marketing Research

Due to the limited accessibility to real-world data on RWHs, there's a compelling need to generate synthetic data to simulate users' activities and usage patterns in RWHs. To achieve this, an in-depth understanding of RWH user segmentations and personas is essential.

This research provides critical insights into user preferences, demographics, and trends within co-working spaces, serving as a foundation for creating realistic synthetic personas and scenarios. Utilising the findings from the research on the preferences for co-working space characteristics [23], our synthetic data generation process prioritises user preferences such as accessibility, atmosphere/interior, and layout.

These characteristics are crucial for simulating realistic user decisions and behaviours in RWHs. Information from the demographics of coworking space users and the co-working statistics [18, 15, 22] enriches our understanding of the RWH user base, including age distribution, professional backgrounds, and gender representation.

This ensures our synthetic personas accurately reflect the diversity within co-working spaces. Insights from the benefit of co-working statistics [15] and the research of the coworking spaces overview [8] highlight the primary motivations for using co-working spaces and the growth trends

ISSN 2007-9737

Day 1 (August 01, 2023):
Dear Diary,

August 1st, 2023.

Waking up to a diffused light peeking through the mostly cloudy skies, today seemed to program another day of calm focus.

In my morning silence, I start with a mug of freshly brewed coffee as I peruse my emails. It's fascinating what happens virtually overnight in this industry; updates, requests, feedback – all pouring in from around the globe. I take some time to review these, noting down the critical tasks and strategizing the optimal sequence to tackle them.

Having identified the priorities, I then switch on my home office workspace. Bright screens, a silent keyboard and a comfortable chair – this sanctuary, despite being at home, brings me into a mindset that's solely professional and immensely focused.

The morning is dedicated mostly to the development tasks. The quiet ambience aids my concentration and helps me become more receptive to the problems at hand. My experience over the years greatly supports my ability to come up with creative solutions faster. The satisfaction of seeing lines of code transform into a functional, responsive and user-friendly website, never gets old.

In the afternoon, I venture out to a nearby Remote Working Hub (RWH). It's a slight change of pace, the buzz and energy here are subtly contagious. It helps me stay updated on new technological trends and also allows me to network with other professionals. The RWH is technologically well-equipped, making it an excellent spot for testing my developed websites on different devices and platforms.

I reserve my late-afternoon for client communication. This involves sending over drafts, checking for any revisions, discussing feedback and coordinating the final steps for project completion. Communication is key in freelancing, and this extra step ensures my product meets client requirements to the maximum.

Returning home, the dusky light filtering through the clouds is met with a quiet satisfaction of a productive day. Developing is strenuous, but the sheer joy of experiencing a concept metamorphosing into a final product is unparalleled, making the whole process worth it.

Before turning in, I spend a few minutes preparing for tomorrow. It primes my mind and allows me a smooth start the following day.

Wrapping up another day in the life of a Freelance Web Developer, I hit the lights. There's a lot of beauty in the silence of the night. It's a perfect palette cleanser before the chaos of coding resumes again tomorrow.

Cheers,
Amara

Fig. 3. Example of synthetic diary

in this domain. These insights are integrated into our synthetic scenarios to model environments that cater to these motivations and reflect current trends. The research on the Irish social, urban, and environmental benefits of coworking spaces [22] informs us about the operational challenges and business trends affecting co-working spaces, which are considered when modelling synthetic data scenarios that include potential obstacles RWH users might face.

By integrating these market research insights into our synthetic data generation methodology, we ensure the produced data is not only realistic but also aligned with current trends and user preferences in the co-working space domain.

This effort resulted in the creation of detailed personas representing a wide array of users found in co-working spaces, reflecting varied professional backgrounds, work styles, and preferences.

Our personas fall into nine main categories, each with sub-categories that further refine their profiles. These include freelancers, remote employees, hybrid workers, digital nomads, small and medium-sized enterprises (SMEs), entrepreneurs, employees of large firms, students and Interns and non-profit/NGOs.

3.2 Dataset

In our effort to enhance the realism of our data, we undertook an extensive collection of real-world information. Data about bus stops in Dublin was acquired from the ArcGIS Hub[1]. In contrast, information regarding the LUAS tram stops and DART train stops within the Dublin area was obtained through the Google Places API[2].

[1] hub.arcgis.com
[2] developers.google.com/maps/documentation/places/web-service/overview

doi: 10.13053/CyS-28-4-5298

ISSN 2007-9737

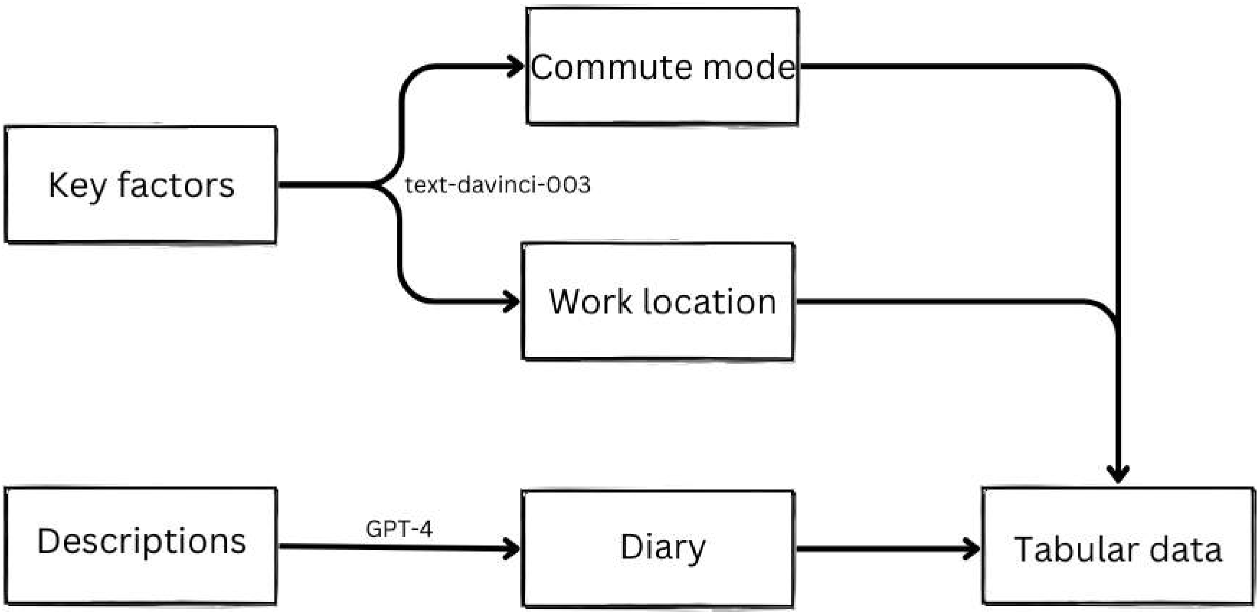

Fig. 4. The process of tabular data generation

This data collection was crucial as public transportation options significantly influence individuals' choices regarding work locations and commuting methods. To ensure the accuracy of distance calculations, data cleaning was conducted, which involved the removal of entries with missing or null values for coordinates.

This process was essential for accurately representing the locations of LUAS tram stops and DART train stops in Dublin. Furthermore, weather data for a period of 30 days in August 2023 was sourced from the Irish Meteorological Service, Met Éireann[3].

Integrating this weather information directly into our code plays a pivotal role in the decision-making processes related to work location choices and commute methods, as weather conditions can significantly affect commuting preferences.

Data on Remote Working Hubs (RWH) and their facilities within the Dublin area was found on the National Hub Network[4].

[3]www.met.ie/climate/available-data/monthly-data

[4]wdc-ie.maps.arcgis.com/apps/webappviewer/index.html?id=469e9517d9864e95beb64a3ab06e9f9d

This dataset provides detailed information on various hubs, including their facilities and locations. To ensure a comprehensive listing of available hubs for inclusion in the synthetic diaries, the dataset was processed to remove any facilities lacking essential information.

Information about the borders of Dublin was sourced from Zenodo[5]. This information is vital for generating random coordinates of persons within the Dublin area to represent their residences accurately. Having the borders of the entire Dublin area ensures that the characters are situated within the Dublin region.

3.3 Model Selection

In our efforts to synthesise RWH user behaviour, selecting an appropriate LLM is crucial. After a thorough evaluation, GPT-4[6] was selected for its unparalleled language understanding and generation capabilities, which are critical to crafting detailed, authentic narratives of RWH user experiences.

[5]zenodo.org/

[6]platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

Unlike alternatives like Generative Adversarial Networks (GAN) and earlier iterations of GPT, GPT-4 excels at generating context-rich simulations that accurately reflect the nuanced daily activities and preferences of users in coworking spaces. Its advanced context awareness and scalability facilitate the creation of extensive and versatile datasets that capture the complexity of RWH interactions.

To enhance GPT-4's narrative capabilities, we integrated "text-davinci-003"[7] to leverage its superior decision-making and reasoning capabilities, especially in evaluating complex scenarios such as work locations and commuting mode choices.

This dual-model approach enables comprehensive simulation of RWH user behaviour, combining the detailed narrative generation of GPT-4 with the analytical precision of "text-davinci-003". Together, they form a powerful framework for generating detailed and adaptable synthetic data, ensuring our datasets remain relevant and reflect changing coworking trends and user behaviour.

This systematic selection and integration of GPT-4 and "text-davinci-003" underscores our commitment to leveraging cutting-edge artificial intelligence to gain insights into RWH user dynamics. The resulting comprehensive data set helps inform the design and management of more responsive and user-centred coworking spaces, improving our understanding of remote work environments.

3.4 LLM Configuration

Our methodology leverages GPT-4's capabilities to generate synthetic data encompassing diary entries, work location decisions, and commute mode selections, aiming to authentically represent the daily routines of RWHs users. This multifaceted approach is detailed as follows:

Diary Entry Generation: Utilising GPT-4, we crafted prompts incorporating individual profiles and weather conditions to simulate diary entries that narrate a user's entire day, focusing on RWH interactions. The model was configured with a 600 token limit to balance detail with conciseness.

Work Location Decision: A custom algorithm determines whether the user works from home or a hub, considering the weather, personal preferences, and proximity to potential hubs. This decision process employs a temperature setting of 0.7 to introduce variability, reflecting the unpredictability of human choices.

Commute Mode Decision: For hub workers, the optimal commute mode is selected based on distance, weather, and available transport options, using a structured prompt with a temperature setting of 0.5 for logical consistency. This step adds realism by incorporating daily commute logistics into the synthetic narratives.

Both decision-making functions utilise the "text-davinci-003" engine, chosen for its ability to generate nuanced and contextually appropriate responses. Parameters were carefully calibrated to ensure outputs that are both relevant and reflective of real-world decision-making processes.

4 Experiment

4.1 Person Generation

In the initial phase of our study, we engaged in persona generation, leveraging the insights from our preceding marketing research. This involved a systematic consultation with ChatGPT to delineate nine distinct categories of remote worker occupations, namely:

Freelancers, Remote Employees, Hybrid Workers, Digital Nomads, Small and Medium-sized Enterprises (SMEs), Entrepreneurs, Employees of Large Firms, Students and Interns, and Non-Profit/NGO Workers. Subsequently, we further segmented these categories where applicable; for instance, 'Freelancers' were subdivided into 'Independent Contractors' and 'Gig Workers'.

For each subdivision, we then crafted detailed personas, utilising a structured process to ensure each persona accurately reflected the person and their occupation of its respective category. Fig. 2 shows an example of the person we generated.

[7]platform.openai.com/docs/models/gpt-3-5-turbo

ISSN 2007-9737

	Weather	Residential Address Latitude	Residential Address Longitude	Work Location	Work Location Latitude	Work Location Longitude	...	Lunch Location	Mood or Sentiment Attribute	Social Interaction Level	Special Events or Occasions	External Influences	Personal Growth Activities
0	Mostly Cloudy	53.34078	-6.26423	Office Suites Club - Harcourt	53.33602	-6.263639	...	Restaurant	negative	Low	None	local festivals	try new art materials
1	Mostly Cloudy	53.34078	-6.26423	Workhub - Camden Street	53.33537	-6.265490	...	Hub	negative	Medium	None	None	brainstorming sessions with peers
2	Partly Cloudy	53.34078	-6.26423	home	53.34078	-6.264230	...	Home	negative	Medium	None	None	give design guest lectures
3	Mostly Cloudy	53.34078	-6.26423	Workhub - Camden Street	53.33537	-6.265490	...	Hub	positive	Low	None	traffic jams	give design guest lectures
4	Mostly Cloudy	53.34078	-6.26423	home	53.34078	-6.264230	...	Cafe	positive	Low	Friend's Birthday	None	Sketch

Fig. 5. The sample of tabular data

4.2 Diary Generation

To make the data more realistic, we first carried out a detailed analysis of the daily behavioural patterns of the Irish population. This analysis aimed to identify a range of activities that individuals typically undertake during the course of a day.

We then incorporated details of these identified activities, as well as details of previously generated personas, into our prompts. We then analysed these daily activities using the GPT-4 API, ultimately generating a synthetic diary of the user. Fig. 3 shows an example of our generated diary.

This approach can create a comprehensive and realistic dataset that mirrors the daily routines and activities of individuals living in Ireland. By leveraging the advanced capabilities of the GPT-4 API, we could simulate nuanced and varied diary entries that reflect the complexity of human behaviour and activities.

After generating synthetic diaries using the GPT-4 API, our experiments extend to the critical decision-making process of selecting an appropriate remote work hub (RWH) and commuting mode for remote work hub users. To accurately simulate these decisions, we adopted the ”text-davinci-003” model.

The process began with the development of prompts that include key factors influencing work location and commute decisions, such as weather conditions, day of the week, individual preferences, and the availability of transportation options.

These prompts were designed to reflect the realistic considerations individuals in Ireland make when choosing where to work and how to commute on a daily basis.

For example, on days with severe weather conditions, the model might recommend working from a nearby RWH to minimise commute time and exposure, or it might recommend working remotely from home if the individual’s preferences are consistent with the scenario.

Likewise, the choice of commuting mode is influenced by factors such as distance, transport availability and environmental factors, with the model weighing the pros and cons of each option before making recommendations.

This dual-model approach leverages both GPT-4 to generate diary entries and ”text-davinci-003” for decision-making tasks, allowing us to create a dataset that not only captures the daily lives of individuals in Ireland but captures their decision-making process about where they work and how they commute.

The result is a synthetic dataset that provides a holistic view of remote workers’ lifestyles, providing valuable insights into their preferences, behaviours and challenges.

4.3 Data Transformation and Analysis

In the final stage of our research, we transformed the synthetic diary entries and decision-making simulations into a structured tabular dataset.

ISSN 2007-9737

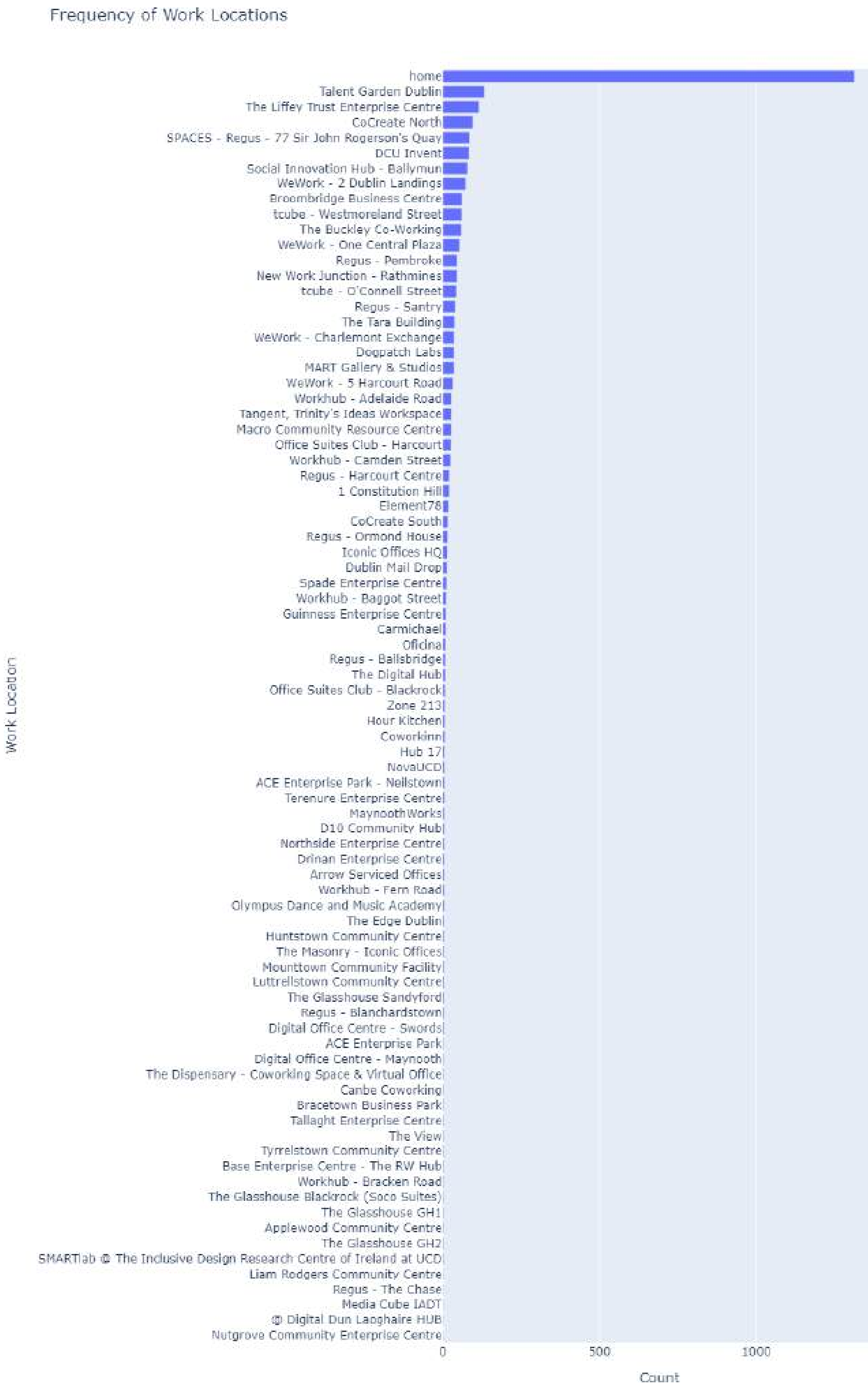

Fig. 6. Analysis of the main work locations of users from the synthetic diary dataset

ISSN 2007-9737

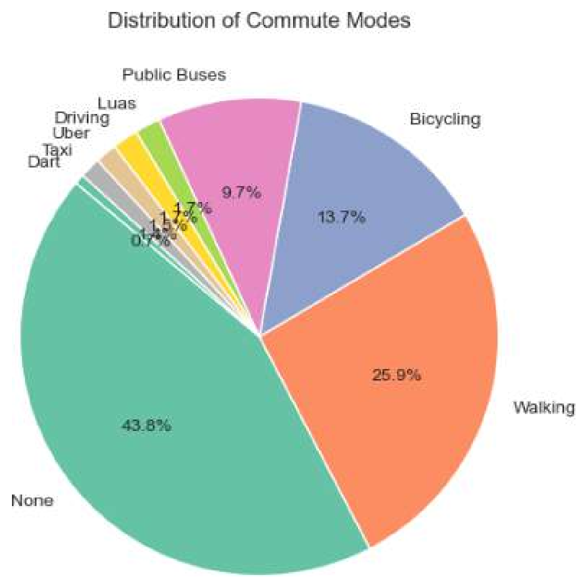

Fig. 7. Analysis of the primary transportation choices of individuals commuting to work

This process involved extracting key information such as work locations, commute modes, weather conditions, and personal sentiments from the narrative and decision outputs. Each piece of data was meticulously catalogued into rows representing daily routines, enabling quantitative analysis of remote work habits and RWH space utilisation.

The structured table provides a comprehensive reflection of an individual's daily life, allowing for an in-depth analysis of the impact of various factors on remote working and RWH decisions. For example, it allows us to study how weather affects the choice of work location, preferences for commuting patterns under different conditions, and the relationship between mood and work productivity or social interaction.

This data-driven approach provides valuable insights into the dynamics of remote work, providing a basis for recommendations for designing adaptable, user-centred coworking spaces. By synthesising rich narrative data into a quantifiable format, our study provides new insights into understanding remote work environments, bridging the gap between qualitative experience and quantitative analysis. Fig 4 shows the whole process of tabular data generation.

5 Results

The synthetic diary dataset, comprising 3000 meticulously detailed entries, offers a view into the daily routines, work locations, and personal sentiments of a diverse population. Spanning across unique occupations and distinct user tags, the dataset encapsulates a wide spectrum of professional and personal experiences, uniformly distributed across all days of the week.

With a predominant weather condition of "Mostly Cloudy", it integrates environmental factors into the daily narratives, enriching the dataset's complexity and relevance to studies on lifestyle and mobility. Geographic specificity is a hallmark of this dataset, with precise latitude and longitude coordinates for both residential and work locations, allowing for in-depth analyses of commute patterns, location preferences, and the spatial dynamics of daily life.

Qualitative attributes such as mood or sentiment, social interaction levels, and engagement with nearby amenities add depth to each entry, painting a vivid picture of the individual's day-to-day experience.

The absence of missing values across all columns underscores the dataset's completeness and reliability for robust analyses. The diversity of data types, from categorical descriptors of occupation and weather to numerical measures of distance and geographical coordinates, enables a comprehensive exploration of the interconnections between work life, personal well-being, and environmental context.

This dataset is poised to serve as a foundational resource for investigating the nuanced impacts of occupational diversity, geographical settings, and lifestyle choices on the daily experiences and overall quality of life of individuals across various demographics and professions. Fig. 5 shows the sample of our tabular data.

Finally, we performed statistical analysis and visualisation of the data we generated, as shown in these figures. The Fig. 6 illustrates the main work locations of users from the synthetic diary dataset, showing that working from home is the most popular option.

This preference highlights the continuing trend towards remote working, and the visualisation further suggests that RWHs vary in popularity, a difference that can be attributed to the diversity of facilities and geographic locations they offer. This variation highlights the richness of the data in capturing the multifaceted nature of work environments, from traditional office space to more unconventional settings.

The Figure 7 demonstrates the primary transportation choices of individuals commuting to work, with a significant highlight on "NONE" as the predominant mode, reflecting the prevalent trend of working from home. This preference underscores a broader shift towards remote work, where the need for physical commuting is eliminated.

Furthermore, the chart reveals a notable preference for eco-friendly and health-conscious modes of transportation, such as walking and biking, suggesting that when commuting is necessary, individuals prefer to choose options that are both environmentally sustainable and beneficial for personal health. This distribution also hints at the possibility that most of the respondents live near their workplaces, allowing for such modes of transport to be feasible.

6 Discussion

Modelling the behaviour of RWH users through synthetic datasets generated by LLMs, such as GPT-4, shows a new approach to understanding the coworking space. This approach provides insights into the preferences, daily routines and decision-making processes of RWH users, enriching our understanding of coworking dynamics in a way that traditional data collection methods are unable to due to limitations in accessibility and privacy.

The insights derived from this study will be of great value to urban planners, coworking space operators and policymakers in designing spaces and policies that are more responsive to the needs and behaviours of a diverse workforce. In addition, detailed personas and simulated interactions provide the basis for optimising coworking environments to better support freelancers, remote employees and digital nomads, contributing to a more inclusive and productive teleworking ecosystem. However, this research highlights the need for continuous improvement, particularly in terms of increasing the diversity and inclusiveness of datasets to more accurately reflect differences in global RWHs users.

Future research directions should focus on validating the reliability of these synthetic datasets against real-world data, incorporating dynamic modelling to accommodate global changes such as pandemics, and fostering cross-disciplinary collaborations to enrich datasets with comprehensive insights.

In addition, keeping up with technological advances in the field of artificial intelligence is crucial to realise the full potential of synthetic data for modelling complex human behaviours and preferences in collaborative environments.

7 Conclusion

This study demonstrates the potential of LLMs in generating synthetic datasets for exploring the complex behaviours and preferences of RWH users, filling an important gap in the lack of real-world data.

Through detailed simulations of everyday activities and decision-making processes, this research provides valuable insights into the changing dynamics of the RWHs, highlighting the importance of adaptive and inclusive environments to meet various professional needs.

Our findings demonstrate the feasibility and utility of synthetic data in understanding and enhancing RWHs, and also suggest directions for future research to ref ine these methods for broader applicability, contributing to the development of RWHs.

Acknowledgments

This research has been done in the AI Digital Twin Lighthouse project of CeADAR, Ireland's National Centre for AI. National Hub Networks support also has helped us gain the required information and network about Ireland's RWHs.

References

1. **Anaby-Tavor, A., Carmeli, B., Goldbraich, E., Kantor, A., Kour, G., Shlomov, S., Tepper, N., Zwerdling, N. (2020).** Do not have enough data? Deep learning to the rescue! Proceedings of the 34th AAAI Conference on Artificial Intelligence, pp. 7383–7390. DOI: 10.48550/ARXIV.1911.03118.

2. **Chen, C., Hui, Q., Xie, W., Wan, S., Zhou, Y., Pei, Q. (2021).** Convolutional neural networks for forecasting flood process in internet-of-things enabled smart city. Computer Networks, Vol. 186, pp. 107744. DOI: 10.1016/j.comnet.2020.107744.

3. **DESKMAG (2022).** 2022 coworking space business trends. deskmag.com/%2Fen%2Fcoworking-news%2F2022-results-of-the-coworking-space-business-trends-survey.

4. **Dong, R., Han, C., Peng, Y., Qi, Z., Ge, Z., Yang, J., Zhao, L., Sun, J., Zhou, H., Wei, H., Kong, X., Zhang, X., Ma, K., Yi, L. (2024).** DreamLLM: Synergistic multimodal comprehension and creation. Proceedings of the 12th International Conference on Learning Representations, pp. 1–37. DOI: 10.48550/arXiv.2309.11499.

5. **Gilardi, F., Alizadeh, M., Kubli, M. (2023).** ChatGPT outperforms crowd workers for text-annotation tasks. Proceedings of the National Academy of Sciences, Vol. 120, No. 30. DOI: 10.1073/pnas.2305016120.

6. **Giménez-Nadal, J. I., Molina, J. A., Velilla, J. (2019).** Work time and well-being for workers at home: Evidence from the American time use survey. International Journal of Manpower, Vol. 41, No. 2, pp. 184–206. DOI: 10.1108/ijm-04-2018-0134.

7. **Horton, J. (2023).** Large language models as simulated economic agents: What can we learn from Homo Silicus? National Bureau of Economic Research. DOI: 10.3386/w31122.

8. **Howell, T. (2022).** Coworking spaces: An overview and research agenda. Research Policy, Vol. 51, No. 2, pp. 104447. DOI: 10.1016/j.respol.2021.104447.

9. **Huang, W., Abbeel, P., Pathak, D., Mordatch, I. (2022).** Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. International Conference on Machine Learning, pp. 1–3. DOI: 10.48550/ARXIV.2201.07207.

10. **Jindal, N., Liu, B. (2007).** Review spam detection. Proceedings of the 16th International Conference on World Wide Web, pp. 1189–1190. DOI: 10.1145/1242572.1242759.

11. **Kelly, J. A., Kelleher, L., Guo, Y., Deegan, C., Larsen, B., Shukla, S., Collins, A. (2022).** Assessing preference and potential for working from anywhere: A spatial index for Ireland. Environmental and Sustainability Indicators, Vol. 15, pp. 100190. DOI: 10.1016/j.indic.2022.100190.

12. **Li, Y., Eldan, R. (2023).** TinyStories: How small can language models be and still speak coherent English? DOI: 10.48550/ARXIV.2305.07759.

13. **Li, Z., Zhu, H., Lu, Z., Yin, M. (2023).** Synthetic data generation with large language models for text classification: Potential and limitations. Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 10443–10461. DOI: 10.18653/v1/2023.emnlp-main.647.

14. **Liu, L., Zhang, Y. (2021).** Smart environment design planning for smart city based on deep learning. Sustainable Energy Technologies and Assessments, Vol. 47, pp. 101425. DOI: 10.1016/j.seta.2021.101425.

15. **McCain, A. (2023).** 33 Captivating coworking statistics [2023]: Facts and trends you need to know. www.zippia.com/advice/coworking-statistics/.

16. **Nosratabadi, S., Mosavi, A., Keivani, R., Ardabili, S. F., Aram, F. (2019).** State of the art survey of deep learning and machine learning models for smart cities and urban

sustainability. International Conference on Global Research and Education, pp. 228–238. DOI: 10.31219/osf.io/gmuzk.

17. **Ohnmacht, T., Z'Rotz, J., Dang, L. (2020).** Relationships between coworking spaces and CO2 emissions in work-related commuting: First empirical insights for the case of Switzerland with regard to urban-rural differences. Environmental Research Communications, Vol. 2, No. 12, pp. 125004. DOI: 10.1088/2515-7620/abd33e.

18. **ONES Software (2024).** People-centric hybrid working platform: ONES (Global). ones.software/hybrid-workspace.

19. **Oswald, K., Zhao, X. (2020).** What is a sustainable coworking space?. Sustainability, Vol. 12, No. 24, pp. 10547. DOI: 10.3390/su122410547.

20. **Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., Bernstein, M. S. (2023).** Generative agents: Interactive simulacra of human behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1–22. DOI: 10.48550/ARXIV.2304.03442.

21. **Tang, R., Han, X., Jiang, X., Hu, X. (2023).** Does synthetic data generation of LLMs help clinical text mining?. DOI: 10.48550/ARXIV.2303.04360.

22. **Wall, S., Crowe, P. R. (2023).** Identifying the social, urban, and environmental co-benefits of coworking spaces in Irish towns. Sustainability, Vol. 16, No. 1, pp. 175. DOI: 10.3390/su16010175.

23. **Weijs-Perrée, M., van-de-Koevering, J., Appel-Meulenbroek, R., Arentze, T. (2018).** Analysing user preferences for co-working space characteristics. Building Research and Information, Vol. 47, No. 5, pp. 534–548. DOI: 10.1080/09613218.2018.1463750.

24. **Wiegand, M., Ruppenhofer, J., Kleinbauer, T. (2019).** Detection of abusive language: The problem of biased datasets. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, pp. 602–608. DOI: 10.18653/v1/n19-1060.

25. **Ye, J., Gao, J., Li, Q., Xu, H., Feng, J., Wu, Z., Yu, T., Kong, L. (2022).** ZeroGen: Efficient zero-shot learning via dataset generation. DOI: 10.48550/arXiv.2202.07922.

26. **Yuan, A., Coenen, A., Reif, E., Ippolito, D. (2022).** Wordcraft: Story writing with large language models. Proceedings of the 27th International Conference on Intelligent User Interfaces, pp. 841–852. DOI: 10.1145/3490099.3511105.

Article received on 16/04/2024; accepted on 20/06/2024.
**Corresponding author is Yilin Li.*