MemoriQA: A Question-Answering Lifelog Dataset

Quang-Linh Tran
ADAPT Centre, School of Computing,
Dublin City University
Dublin, Ireland
linh.tran3@mail.dcu.ie

Binh Nguyen
University of Science,
Vietnam National University
Ho Chi Minh City, Vietnam
ngtbinh@hcmus.edu.vn

Gareth J. F. Jones
ADAPT Centre, School of Computing,
Dublin City University
Dublin, Ireland
gareth.jones@dcu.ie

Cathal Gurrin
ADAPT Centre, School of Computing,
Dublin City University
Dublin, Ireland
cathal.gurrin@dcu.ie

ABSTRACT

Lifelogging can be referred to as the process of passively collecting data on an individual's daily life. Lifelog data provides a large amount of information which can be used to understand the lifelogger's lifestyle and preferences. This data can also support the lifeloggers in saving their memories and important moments. Question-answering (QA) is a common task in natural language processing (NLP) and can be extended to multi-modal such as the visual question-answering task. QA for lifelog data can be described as the task of answering questions about a lifelogger's past using lifelog data, which can significantly help lifeloggers understand their life by asking questions about their lifelog. QA for lifelogs can also provide useful insights into lifelogger's life for those exploring their lifelog. This paper presents the MemoriQA lifelog dataset designed to explore the question-answering task for lifelogs. This dataset provides 61-day lifelog images and other lifelog data such as internet activity, health metrics, music listening history and GPS. A comprehensive annotation process is performed to create the description as well as question-answer pairs. We propose some methods to address the QA in lifelog problem in this paper.

CCS CONCEPTS

• **Information systems → Question answering**.

KEYWORDS

Personal Lifelog Archive, Question Answering, Lifelog Dataset

ACM Reference Format:
Quang-Linh Tran, Binh Nguyen, Gareth J. F. Jones, and Cathal Gurrin. 2024. MemoriQA: A Question-Answering Lifelog Dataset. In *The 1st ACM Workshop on AI-Powered Q&A Systems for Multimedia (AIQAM '24), June 10, 2024, Phuket, Thailand*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3643479.3662050

AIQAM '24, June 10, 2024, Phuket, Thailand

ACM ISBN 979-8-4007-0547-2/24/06
https://doi.org/10.1145/3643479.3662050

1 INTRODUCTION

Lifelogging can be defined as the process of passively collecting digital data from an individual's daily life [12]. Various types of daily life data can be logged, ranging from visual data such as images and/or videos from wearable cameras, context data such as location, time, and weather, biometric data from smartwatches, and activity on the internet. These data can be used in a wide range of applications from health care [3, 5] to memory support [15] and lifestyle enhancement [16]. Biometrics and visual data can help monitor health indicators and food that a lifelogger consumes [25]. This data can promote a healthier diet or alert individuals to potentially dangerous patterns. Visual and contextual lifelog data can be utilized to retrieve past events [22, 23]. The retrieval system aids lifeloggers to find, browse and explore events stored within their lifelog. This can be used, for example, to support the memory of the lifelogger [2, 19]. Lifelog data provides a significant amount of information about lifeloggers' lives, such as their eating habits and online activities. One type of useful exploration of a lifelogs is use of a question-and-answer (QA) format. In QA, lifeloggers pose questions to the QA system, which then receives answers [20]. The use of QA in lifelog exploration is interesting because it helps to understand more about the lifeloggers' lifestyle and can be used to enhance it. To support research into lifelog QA, we have constructed the MemoriQA dataset.

Research into QA for lifelogging has recently attracted interest from several researchers [20, 21]. In this task, a question regarding lifelog data serves as the input and a QA system generates an answer based on information from the lifelog dataset. For example, when the lifelogger asks the system the question "How long did I have dinner at the ABC restaurant in May 2023?", the QA system finds the event in the lifelog and generates the answer based on the information it locates. QA in lifelogging encompasses visual QA tasks, which involve interpreting visual data from specific moments, as well as answering questions based on context data and other sources [21]. However, the question "How long did I spend at the ABC restaurant for dinner in May 2023?" necessitates calculating both contextual and visual data to derive the answer, thus making it distinct from a visual QA task. QA in lifelogging differs from traditional QA tasks in several ways. It involves a retrieval process to find relevant lifelog moments before generating the answer. In addition, the QA in lifelogging task requires computation for aggregated questions and several data sources to find a correct answer, so it can be considered a more challenging task than traditional QA. For example, a topic in the QA task in LSC'23 [10] is "Which airline did I fly with most often in 2019?" involving metadata and visual information to find which airline the lifeloggers flew in 2019 before aggregating the candidate answers and get the most often travelled airline.

There are several challenges that the QA in lifelogging imposes. Firstly, due to the nature of lifelogs, there are a lot of life moments that can be unique, but there are also many recurring moments. Incorporating diverse data sources is crucial for enriching the contextual understanding of lifelog data and ensuring accurate responses. Secondly, given the vast amount of lifelog data, an accurate retrieval component is necessary for the QA model to find relevant moments that address the question. Thirdly, integrating various modalities of lifelog data, such as images/videos, time-series data, and textual descriptions, poses a challenge. To tackle these challenges, we develop a comprehensive dataset for lifelogging, aiming to construct an efficient and accurate QA model tailored to lifelog data.

In this paper, we introduce the MemoriQA dataset, which is a lifelog dataset specializing for the QA task. This dataset comprises 61 days lifelog data, from 10th December 2023 to 23rd February 2024 by a lifelogger in their daily life. The data includes lifelog images captured by an AutoGrapher wearable camera, GPS data from the Geo Tracker application[1] on mobile phones, health data from Fitbit[2], My Google Activity data[3], and music listening data from Spotify[4]. In addition, we also provide the annotation on the lifelog data with narrative descriptions and the question-answers. These data are expected to be the training data for building a QA system that can answer the question in lifelog domain.

2 RELATED WORK

Although lifelog applications have been researched for many years, from the early proposal for a personal information system (Memex) by Vannevar Bush in 1945 [6], to the MyLifeBits project by Gordon Bell in 2001 [9], the QA task in lifelog has only been explored in recent years. The Lifelog Search Challenge (LSC) [10] has organized a task of question answering in lifelog data since LSC'22 in 2022 [13], but in this challenge, the answers were lifelog moments rather than textual answers. In the following challenge LSC'23 [10], the QA task became a standard task with textual answers required. MyEachtra [22] achieved the top 1 position in the QA task in LSC'23 with a FrozenBiLM video question-answering (QA) model [24] to answer the question. The system adapts the model to lifelog question-answering by using a prompt "[CLS] Question: <Question>? Answer: [MASK]."

The Workshop on AI-Powered Q&A Systems for Multimedia (AIQAM) [18] is for the first time organized at the International Conference on Multimedia Retrieval 2024 Conference (ICMR 2024). This workshop draws significant attention from researchers about QA systems on multimedia, and QA on lifelog is also an application of multimedia. Several QA datasets have been constructed recently. Tran et al [21] built a QA lifelog dataset of 85 days of lifelog data from the LSC20 dataset [11]. They used the lifelog images and metadata to construct the dataset. The dataset comprises a total of 15,065 QA pairs, which are primarily formatted as multiple-choice and yes-no questions. Both rule-based and neural network approaches were employed to generate the question-answer pairs, followed by a review process to eliminate redundant or duplicated entries. Despite the considerable challenge posed by this dataset, as evidenced by the human gold-standard baseline accuracy of 0.8417 for yes/no questions and 0.8625 for multiple-choice questions, there are still limitations in this dataset in the question format. The multiple-choice and yes/no questions have limited use cases in real world application, and the automatically generated wrong answers can at times be misleading. In contrast to the use of multiple-choice and yes/no questions, we opted for open-ended question-answering in the MemoriQA dataset to enhance its realism for practical application.

Another QA lifelog dataset is TimelineQA [20]. TimelineQA generates textual lifelogs of imaginary people, spanning a wide range of time and activities, from significant life events such as high school graduation to routine activities like going for a run. They created synthetic episodes for each life experience of the imaginary person and also the question-answers for these episodes. Questions were categorized into atomic queries, complex multi-hop queries, complex aggregate queries, and temporal queries. Additionally, they explored various methods to address the QA challenge in lifelog data, including extractive QA [14], retrieval-augmented QA [17], and table QA [4]. The performance of these QA models is promising with good accuracy. However, due to the synthetic process of generating lifelog data, the data is not realistic and cannot reflect a person's daily activity. Moreover, the data of lifelog is various in modalities, from visual to textual, but the synthetic data in this dataset is mainly text, which makes it more like a textual open-ended QA. In addition, the dataset contains only a limited number of episodes per day, whereas real lifelog data tends to be more diverse and complex. While this research provides a valuable baseline, real-world datasets may present additional challenges for the QA task.

3 DATASET

This section provides detailed information about the MemoriQA dataset. The process of constructing the dataset follows Dang-Nguyen's process of disclosing a lifelog dataset [7]. Figure 1 illustrates the process of constructing the MemoriQA dataset. The blue box shows the data collecting stages, in which lifelog data from different sources are collected. The data goes through a time alignment process to synchronize the datetime in different data sources. Then the lifelog data is annotated manually and semi-automatically by LLM to generate the descriptions and question-answers, which are depicted in the green box. All annotated data is double-checked to ensure the data quality. The final result is the completed MemoriQA dataset. The dataset is available upon request, please get in touch with the corresponding authors to request the dataset.

[1]https://geo-tracker.org/
[2]https://www.fitbit.com/
[3]https://myactivity.google.com/
[4]https://spotify.com/

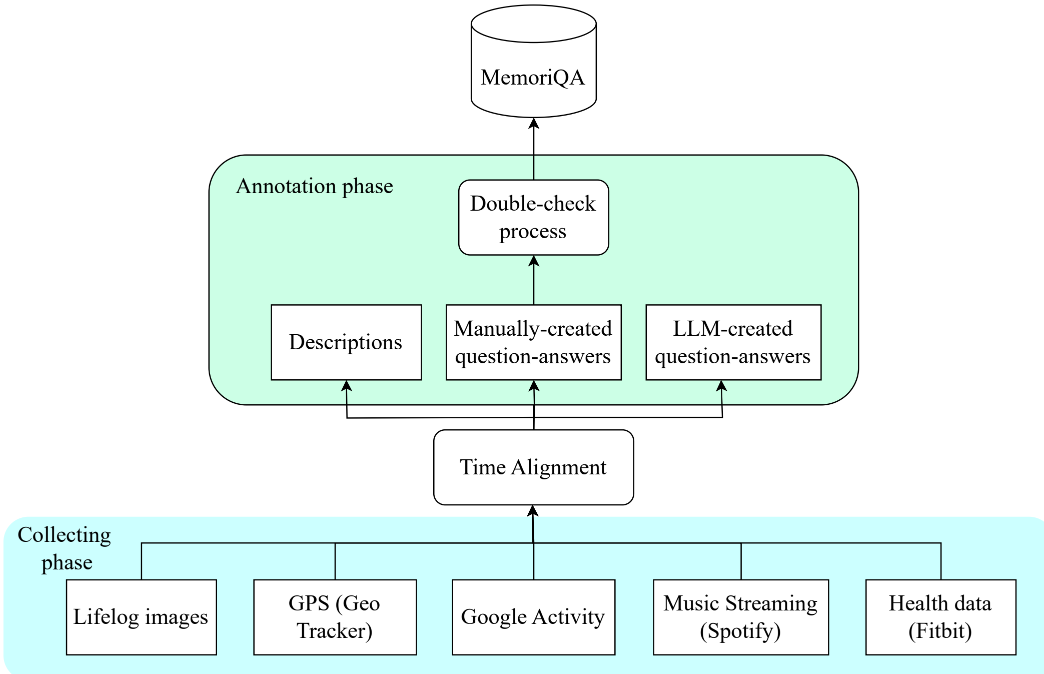

Figure 1: Dataset construction process

3.1 Data Collecting

We collect lifelog data from five different data sources, namely: images, GPS, Google Activity, music streaming, and health data. A lifelogger wore an Autographer camera in front of the chest for 61 days to collect the lifelog images. This lifelogger performed their normal daily activities, without impact from the lifelogging device. The camera was set to a medium capture rate, allowing it to capture approximately 2 to 3 images per minute. The camera could continuously capture images for up to 4 hours before requiring recharging. The resolution of images is 2592×1936 pixels. The camera collected a total of 84,249 images with approximately 10 hours per day. The lifelogger then filtered out images they do not wish to include in the dataset. Furthermore, we discarded low-quality images, including those that are completely black or highly blurred. The number of remaining images is 57,021 images. We further anonymized the images by utilizing the Retina Face detection model [8] to blur faces and manually inspect each image to remove any other identifiable objects (items that may identify individuals such as name tags). The lifelog images are named according to the time of capture and stored on a local hard disk for future use. Examples of lifelog images from the dataset are shown in Figures 2.

The GPS information is collected through the Geo Tracker application, which is installed in the data gatherer's mobile phone. This app tracks and saves the latitude and longitude of the phone every second. The exported data in the format of GPX is processed and converted to a CSV file for easy processing. The CSV file of GPS data has three columns, namely time, latitude, and longitude to present the location for each time.

Google provides a large number of applications on mobile phones applications and websites. These applications range from office applications such as Google Drive, and Google Docs to daily used applications such as Youtube, and Maps, and Google stores every action that users do in their applications. This data can provide many insights into what the users do on the internet. Google also supports exporting this data through the MyActivity website [5]. We collected the Google Activity data of the lifelogger in the period of collecting data from December 2023 to February 2024. The activity data of over 32 Google applications is aggregated into a single file with a standard format. The CSV file that saves the Google activity data contains 3 columns with datetime, application, and action and 36,787 data points.

Spotify is a big music streaming application and the lifelogger used this application to listen to music in the collecting data period. Spotify supports exporting the data with JSON files and we convert and standardize the data to a CSV file with 4 attributes, namely endTime, artistName, trackName and msPlayed. There is a total of 1599 tracks that had been streamed in the collecting period.

Biometric / Health data is collected from a Fitbit Versa 2 smartwatch. The lifelogger wore this device for 24 hours in 7 days to collect biometric data. We export the Fitbit data from Google and get the following biometrics: Active Zone Minutes, heart rate variability, sleep score, oxygen saturation (SPO2), stress score and wrist temperature.

[5]https://myactivity.google.com/

Figure 2: Lifelog images captured from the AutoGrapher camera

Month	# Day	# Images	# Descriptions	# QA pairs	AVG question length	AVG answer length
Dec 2023	22	19995	886	1468	10.39	4.62
Jan 2024	27	26254	737	1515	9.97	5.35
Feb 2024	12	10772	302	661	9.52	5.20
Total	61	57021	1925	3644	9.96	5.05

Table 1: Some statistics on the MemoriQA dataset

3.2 Data Annotation

After collecting all the lifelog data and performing the time alignment, we annotate the description of lifelog data for each day. Firstly, we segment the lifelog images into events, which are sequences of images that indicate a single activity. An annotator manually creates a description for each event in the view of the first person. For example, the description "20231012_120391, I was having lunch with rice and fish" indicates the event of having lunch at the given time. The dataset contains a total of 1,925 descriptions, averaging approximately 31.55 descriptions per day.

Once the description annotation process is complete, we proceed to annotate question-answer pairs. Initially, the annotator manually creates approximately 20 question-answers per day based on the description. There are 9 categories with many question templates for the annotator to create question-answer based on that. They also classify the created question-answers into 1 of 9 categories and 1 of 4 question types, referencing the time of the event. There are 1,256 question-answers created by this method.

From the manually created descriptions and question-answers, we continue to generate more question-answers by using Large Language Models (LLM). We utilize GPT-4 [1], developed by OpenAI, to generate more question-answers. The prompt used for this task is: "Based on these descriptions <DE> descriptions <DE>, each row indicates a description and the start time and the end time is the start time of the next description. You should provide the correct time. Generate 100 questions and answers along with category and difficulty level, with 20 How long questions and 80 other questions, from the provided descriptions and follow the template: <QA> manually-created question-answers <QA>." Through this automated process, 2,388 question-answers are generated, averaging approximately 39.1 question-answers per day. A double-checking process is conducted to ensure annotation quality. An annotator checks all LLM-created question-answers to modify any wrong information in the answer. All duplicated question-answers are deleted.

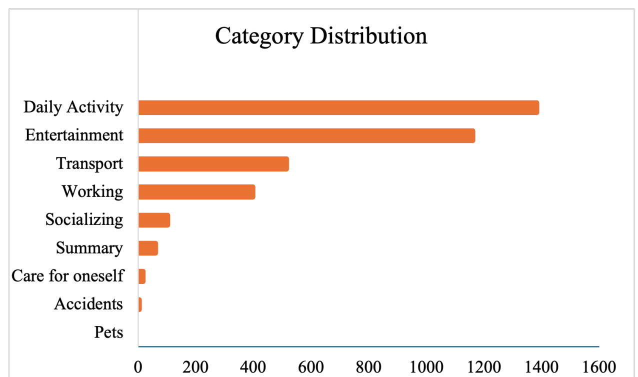

Figure 3: Question Category Distribution

3.3 Dataset Information

After the annotation phase, we have a completed MemoriQA dataset. Table 1 provides some general information about the MemoriQA dataset. There are a total of 61 days from December 2023 to February 2024 in the dataset. There are a total of 57,021 images, with January 2024 having the largest number. The manual annotation process creates 1,925 descriptions. Despite December 2,023 having fewer images than January 2024, it has more descriptions. This is due to the lifelogger engaging in more activities in December, including a trip to Japan. A total of 3,644 question-answers are created, including both manual and LLM-generated. The average question length is 9.96 words, while the average answer length is 5.05 words.

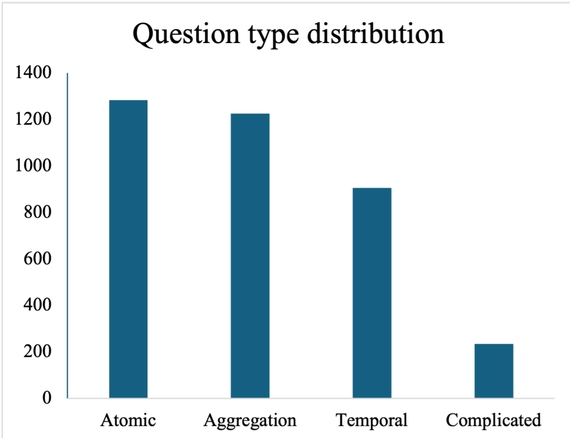

Figure 4: Question Type Distribution

We also do some analysis on the question category and type. There are 9 categories of life activity in the dataset and each question should belong to one of these categories. These categories are selected from proposed categories in previous research [20], including daily activity, entertainment, transportation, working, socializing, care for oneself, accidents, pets and summary. The first eight categories cover nearly all aspects of a normal person's lifestyle, while the last category is for summarising the narratives for the whole day. Figure 3 shows that the Daily Activity category has the highest number of questions, totalling over 1,400. The category Entertainment follows with nearly 1,200 questions. In contrast, categories such as Accidents and Pets have a relatively low number of questions. This is because the lifelogger does not have any pets and rarely encounters any accidents. There are 4 types of questions in the MemoriQA dataset, that inherit from TimelineQA [20]. Figure 4 illustrates the distribution of question types. The atomic and aggregation question types have the highest number of questions, totalling 1,282 and 1,225, respectively. The temporal question type has 904 questions while the complicated type only has 233 questions.

4 PROPOSED RESEARCH

In this section, we propose a research direction that can utilize the MemoriQA dataset to build a QA model in lifelog data. With recent developments in LLMs, the QA task becomes less challenging thanks to the generative capability of LLMs. Methods such as RAG [17] and fine-tuning are employed to integrate external knowledge, such as medical and legal data, to improve the accuracy of LLMs in specific domains. We can apply these methods to the MemoriQA dataset to construct a QA model. A retrieval component retrieves all relevant information related to an incoming question about lifelog in the dataset. The retrieved results and question can be concatenated to a prompt and an LLM can generate the answer from that prompt. The fine-tuning methods can be used to incorporate the MemoriQA data into an LLM. The question-answers and descriptions can be formulated to a prompt. An LLM can be fine-tuned on that prompt to provide a desired answer for each question and description. A comparative experiment can be conducted to measure the pros and cons of both methods in building a QA model from the MemoriQA dataset.

To evaluate the performance of QA models in the MemoriQA dataset. We can use some traditional metrics about QA such as accuracy, and BLUE score or we can evaluate these models' performance using LLM. Many research now use state-of-the-art LLM such as GPT-4 to evaluate the performance of the LLM model. GPT-4 scores several metrics such as relevance score, accuracy score, coverage, coherence, etc. These metrics can help to evaluate the performance of the QA model.

5 CONCLUSIONS

In this paper, we introduce the MemoriQA dataset, which is a lifelog dataset specialized for the question-answering task. We describe the process of collecting data from different sources such as wearable cameras, smartwatches, google activity, music streaming, and GPS tracking. The annotation process to create the description and question-answer is presented. We also propose some methods that can be applied to this dataset to solve the QA task in lifelog data. In the future, we will carry out the experiment to examine the quality of this dataset and construct QA systems that can accurately answer the question in this MemoriQA dataset.

ACKNOWLEDGMENTS

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at Dublin City University [13/RC/2106_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Mélissa C Allé, Liliann Manning, Jevita Potheegadoo, Romain Coutelle, Jean-Marie Danion, and Fabrice Berna. 2017. Wearable cameras are useful tools to investigate and remediate autobiographical memory impairment: A systematic PRISMA review. *Neuropsychology review* 27 (2017), 81–99.

[3] Adrienne H Andrew, Kevin Eustice, and Andy Hickl. 2013. Using location lifelogs to make meaning of food and physical activity behaviors. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 408–411.

[4] Gilbert Badaro, Mohammed Saeed, and Paolo Papotti. 2023. Transformers for tabular data representation: A survey of models and applications. *Transactions of the Association for Computational Linguistics* 11 (2023), 227–249.

[5] Andrea LS Bulungu, Luigi Palla, Jan Priebe, Lora Forsythe, Pamela Katic, Gwen Varley, Bernice D Galinda, Nakimuli Sarah, Joweria Nambooze, Kate Wellard, et al. 2021. Validation of a life-logging wearable camera method and the 24-h diet recall method for assessing maternal and child dietary diversity. *British Journal of Nutrition* 125, 11 (2021), 1299–1309.

[6] Vannevar Bush et al. 1945. As we may think. *The atlantic monthly* 176, 1 (1945), 101–108.

[7] Duc-Tien Dang-Nguyen, Liting Zhou, Rashmi Gupta, Michael Riegler, and Cathal Gurrin. 2017. Building a Disclosed Lifelog Dataset: Challenges, Principles and Processes. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing* (Florence, Italy) *(CBMI '17)*. Association for Computing Machinery, New York, NY, USA, Article 22, 6 pages. https://doi.org/10.1145/3095713.3095736

[8] Jiankang Deng, Jia Guo, Zhou Yuxiang, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. 2019. RetinaFace: Single-stage Dense Face Localisation in the Wild. In *arxiv*.

[9] Jim Gemmell, Gordon Bell, and Roger Lueder. 2006. MyLifeBits: a personal database for everything. *Commun. ACM* 49, 1 (2006), 88–95.

[10] Cathal Gurrin, Björn Þór Jónsson, Duc Tien Dang Nguyen, Graham Healy, Jakub Lokoc, Liting Zhou, Luca Rossetto, Minh-Triet Tran, Wolfgang Hürst, Werner Bailer, et al. 2023. Introduction to the sixth annual lifelog search challenge, LSC'23. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*. 678–679.

[11] Cathal Gurrin, Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Björn Þór Jónsson, Jakub Lokoč, Wolfgang Hürst, Minh-Triet Tran, and Klaus Schöffmann. 2020. Introduction to the Third Annual Lifelog Search Challenge (LSC'20). In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (Dublin, Ireland) *(ICMR '20)*. Association for Computing Machinery, New York, NY, USA, 584–585. https://doi.org/10.1145/3372278.3388043

[12] Cathal Gurrin, Alan F Smeaton, Aiden R Doherty, et al. 2014. Lifelogging: Personal big data. *Foundations and Trends® in information retrieval* 8, 1 (2014), 1–125.

[13] Cathal Gurrin, Liting Zhou, Graham Healy, Björn Þór Jónsson, Duc-Tien Dang-Nguyen, Jakub Lokoć, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Klaus Schöffmann. 2022. Introduction to the Fifth Annual Lifelog Search Challenge, LSC'22. In *Proceedings of the 2022 International Conference on Multimedia Retrieval*. 685–687.

[14] Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906* (2020).

[15] Basel Kikhia, Josef Hallberg, Johan E Bengtsson, Stefan Savenstedt, and Kare Synnes. 2010. Building digital life stories for memory support. *International journal of Computers in Healthcare* 1, 2 (2010), 161–176.

[16] Gunjan Kumar, Houssem Jerbi, Cathal Gurrin, and Michael P O'Mahony. 2014. Towards activity recommendation from lifelogs. In *Proceedings of the 16th international conference on information integration and web-based applications & services*. 87–96.

[17] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[18] Tai Tan Mai, Quang-Linh Tran, Ly-Duyen Tran, Tu Ninh, Duc-Tien Dang-Nguyen, and Cathal Gurrin. 2024. The First ACM Workshop on AI-Powered Question Answering Systems for Multimedia. *Proceedings of the 2024 International Conference on Multimedia Retrieval (ICMR'24)*. https://doi.org/10.1145/3652583.3658890

[19] Ali Mair, Marie Poirier, and Martin A Conway. 2017. Supporting older and younger adults' memory for recent everyday events: A prospective sampling study using SenseCam. *Consciousness and Cognition* 49 (2017), 190–202.

[20] Wang-Chiew Tan, Jane Dwivedi-Yu, Yuliang Li, Lambert Mathias, Marzieh Saeidi, Jing Nathan Yan, and Alon Y Halevy. 2023. TimelineQA: A benchmark for question answering over timelines. *arXiv preprint arXiv:2306.01069* (2023).

[21] Ly-Duyen Tran, Thanh Cong Ho, Lan Anh Pham, Binh Nguyen, Cathal Gurrin, and Liting Zhou. 2022. LLQA-Lifelog Question Answering Dataset. In *International Conference on Multimedia Modeling*. Springer, 217–228.

[22] Ly Duyen Tran, Binh Nguyen, Liting Zhou, and Cathal Gurrin. 2023. MyEachtra: Event-based interactive lifelog retrieval system for lsc'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge*. 24–29.

[23] Quang-Linh Tran, Ly-Duyen Tran, Binh Nguyen, and Cathal Gurrin. 2023. MemoriEase: An Interactive Lifelog Retrieval System for LSC'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge*. 30–35.

[24] Antoine Yang, Antoine Miech, Josef Sivic, Ivan Laptev, and Cordelia Schmid. 2022. Zero-shot video question answering via frozen bidirectional language models. *Advances in Neural Information Processing Systems* 35 (2022), 124–141.

[25] Qianling Zhou, Di Wang, Cliona Ni Mhurchu, Cathal Gurrin, Jiang Zhou, Yu Cheng, and Haijun Wang. 2019. The use of wearable cameras in assessing children's dietary intake and behaviours in China. *Appetite* 139 (2019), 1–7.