EDEN: Towards a computational framework to align incentives in healthy aging

Wasu Mekniran[1,2] [a] and Tobias Kowatsch[3,4] [b]
[1]Centre for Digital Health Interventions (CDHI), Department of Management, Technology, and Economics, ETH Zurich, Switzerland, [2]CDHI, Institute of Technology Management, University of St. Gallen, Switzerland; [3]Institute for Implementation Science in Health Care, University of Zurich; [4]School of Medicine, University of St. Gallen, Switzerland
wmekniran@ethz.ch, tobias.kowatsch@uzh.ch

Keywords: Healthy Aging; Incentive; Natural Language Processing; Retrieval-Augmented Generation; Large-Language Models; Network Analysis

Abstract: Incentive misalignment among healthcare stakeholders poses significant barriers to promoting healthy aging, hindering efforts to mitigate the burden of long-term care. Despite extensive research in public health, incentive gaps persist, as static implementation guidelines often fail to accommodate dynamic and conflicting incentives. This study introduces and evaluates EDEN (eden.ethz.ch), a computational framework designed to dynamically map stakeholder incentives using a Retrieval-Augmented Generation pipeline. A comparative study using a health insurer use case evaluates alternative incentive analyses; qualitative content analysis, large language models, and EDEN. The evaluation assesses their ability to identify and address incentive gaps. Preliminary findings demonstrate the EDEN's ability to map incentives and highlight misalignment compared to alternative approaches. These findings demonstrate how EDEN can offer evidence-based strategies for key healthcare stakeholders, such as health insurers, based on retrieval features to align incentives in healthy aging.

1 INTRODUCTION

The healthy aging initiatives face significant challenges in fostering collaboration among diverse stakeholders, including health insurers, government agencies, healthcare providers, and local communities (WHO, 2020). These stakeholders often operate with distinct and sometimes conflicting goals, leading to fragmented efforts that undermine the scalability and effectiveness of long-term initiatives (Mekniran et al., 2024; Fried et al., 2022). For instance, health insurers' focus on short-term cost containment often conflicts with healthcare providers' long-term goal of improving population health outcomes. Such misaligned incentives, coupled with limited mechanisms for coordination and macro-system integration, impede the development of cohesive and sustainable healthcare strategies.

Despite the recognized importance of collaboration and alignment, significant gaps remain in understanding how stakeholder incentives align and integrate within the ecosystem (Berwick et al., 2008). Current efforts, such as static best-practice checklists, often fail to account for the dynamic and evolving nature of stakeholder priorities. For health insurers, these gaps are particularly evident in their regulated care provision (Mekniran, Kramer, et al., 2024). Furthermore, the lack of systematic tools to identify value propositions, analyze alignment patterns, and generate actionable insights leaves unexplored potential collaborations and unaddressed incentive gaps.

This workshop paper introduces EDEN, a computational framework that augments the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework (Chapman, 2000) by integrating artificial intelligence (AI) techniques, including Natural Language Processing (NLP), Large Language Models (LLMs) along with Retrieval-Augmented Generation (RAG). EDEN adapts CRISP-DM to the dynamic complexities of healthcare ecosystems, making it capable of systematically mapping stakeholder incentives, identifying misalignments, and uncovering actionable collaboration opportunities. Through the analysis of unstructured data, such as value propositions from health insurers and healthcare providers, EDEN dynamically generates user-specific evidence-based recommendations. To this

[a] https://orcid.org/0000-0001-5184-0438
[b] https://orcid.org/0000-0001-5939-4145

end, the study addresses the following research questions:

1. How effectively does EDEN identify thematic clusters in value propositions?
2. How effectively does EDEN generate recommendations for collaboration?

Next, we present the theoretical foundation of EDEN, outline its approach, discuss the results of comparative analysis, and conclude with implications for the proposed tool and its potential development.

2 RELATED WORK

2.1 Collaboration for healthy aging

Promoting healthy aging relies on multi-stakeholder collaboration, as highlighted by frameworks such as the World Health Organization's Decade of Healthy Ageing, which prioritizes fostering age-positive attitudes, empowering older adults in supportive communities, delivering person-centered care, and ensuring access to long-term care (WHO, 2020). The Global Roadmap for Healthy Longevity advocates for inclusive social infrastructure to combat ageism, improve digital access, and integrate health and lifelong learning to support equitable care and active participation for older adults (Dzau & Jenkins, 2019; Fried et al., 2022; National Academy of Medicine, 2022; Wong et al., 2023). Similarly, Cox and Faragher emphasize prioritizing aging biology, fostering interdisciplinary research networks, and scaling aging initiatives through targeted funding and awareness campaigns (Cox & Faragher, 2022).

Although these frameworks highlight the importance of collaboration, they often rely on static principles that do not account for dynamic stakeholder incentives. Altpeter et al. emphasize the participation of non-traditional partners, strengthening healthcare linkages, and adapting local programs to better serve vulnerable populations (Altpeter et al., 2014). Bonnes et al. advocate for longevity clinics that integrate public health with research collaborations, offering early detection and lifestyle interventions to improve clinical outcomes (Bonnes et al., 2024).

2.2 Iterative analytics process

EDEN, an acronym for Emerging Business Models in Digital Health for Healthy Longevity, adopts the CRISP-DM framework to address the complexities of stakeholder analysis in healthcare. CRISP-DM, recognized for its iterative and cyclical structure, is widely adopted for developing data-driven solutions due to its flexibility and domain-agnostic design, making it particularly suited to dynamic and complex healthcare systems such as healthy aging ecosystems (Larose, 2015). EDEN enhances the CRISP-DM framework by integrating NLP using the Natural Language Tool Kit (NLTK) package (Bird et al., 2009), LLMs, and RAG. These techniques enable EDEN to derive actionable insights by modelling and contextualizing stakeholder value propositions. The adapted phases of CRISP-DM within EDEN are shown in Figure 1.

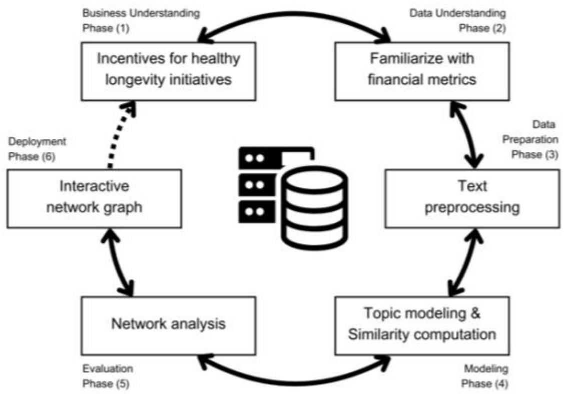

Figure 1: EDEN's iterative analytics process, extends the CRISP-DM framework (Chapman, 2000) by integrating NLP, LLMs, and RAG, making CRISP-DM more suited to evolving priorities of healthcare stakeholders.

(1) Business Understanding phase identifies and analyzes stakeholder incentives in preventive care through qualitative methods, including systematic literature reviews, market research, and expert interviews. These inputs establish a foundational understanding of stakeholder goals and challenges (Giger et al., 2024; Mekniran, Giger, et al., 2024; Mekniran, Kramer, et al., 2024; Mekniran & Kowatsch, 2023).

(2) Data Understanding phase explores structured and unstructured datasets, including organizational metrics from financial database, to identify value propositions and stakeholder priorities in a healthy aging ecosystem (Mekniran, Giger, et al., 2024).

(3) Data Preparation phase involves text preprocessing steps such as tokenization, lemmatization, and vectorization, enabling thematic and relational analyses of stakeholder information (Bird et al., 2009).

(4) The Modeling phase applies topic modeling to uncover thematic clusters and cosine similarity to measure alignment. LLMs contextualize stakeholder propositions, with RAG enhancing organizational

data integration and network analysis to uncover evidence-based collaboration patterns (Silge & Robinson, 2017).

(5) Evaluation phase assesses the model's accuracy and utility using metrics such as similarity scores and network centrality, ensuring robust representation of stakeholder relationships.

(6) Deployment phase presents computational results through interactive network graphs, highlighting collaboration opportunities and enabling users to input value propositions for real-time alignment assessments. Cloud computing ensures scalability and real-time accessibility, supporting larger datasets and complex stakeholder ecosystems.

3 METHODS

This study uses a comparative analysis to assess the effectiveness of EDEN in generating actionable recommendations within the healthy aging ecosystem (Sharma & Kaur, 2017; Verma et al., 2016). Using value propositions from user input and from previous longevity landscape study (Mekniran, Giger, et al., 2024), our comparative analysis compares (A) qualitative content analysis conducted by an author W.M. based on existing guidelines to code organization segment, target customers, and products and services (Fried et al., 2022; Hsieh & Shannon, 2005; WHO, 2020), (B) standalone LLM using GPT-4o model (OpenAI et al., 2024), and (C) proposed EDEN's CRISP-DM methodology, which integrates LLMs GPT-3.5-turbo with NLP techniques such as topic modeling and similarity computation, see Figure 2. The study examines the contextual relevance, granularity, and actionability of the insights generated, providing an initial requirement of EDEN's capability in addressing incentive misalignments.

EDEN integrates NLP with an adapted RAG framework to analyze unstructured stakeholder value propositions and provide actionable recommendations. NLP techniques, such as topic modeling and Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, are employed to identify thematic clusters, extract key incentives, and map similarities among stakeholders (Jelodar et al., 2019; Valdez et al., 2018). These foundational processes enable a structured analysis of the value proposition and collaboration opportunities within the healthy aging ecosystem, supporting strategic decision-making (Frow & Payne, 2011; Murtaza & Ikram, 2010).

The integration of RAG in EDEN addresses key limitations of traditional generative AI systems, such as hallucinations (Béchard & Ayala, 2024; Gao et al., 2023; Huang & Huang, 2024), by dynamically retrieving context-specific data, including thematic clusters and similarity scores derived from network analysis, and incorporating it into the prompt context for LLMs. This bridging of retrieval and generation ensures that output is evidence-based, actionable, and aligned with domain-specific knowledge (Li et al., 2024). RAG further enhances contextualization, scales to dynamic datasets, and adapts to the complex dynamics of the healthy aging ecosystem. Figure 2. illustrates the EDEN workflow and integration of these components.

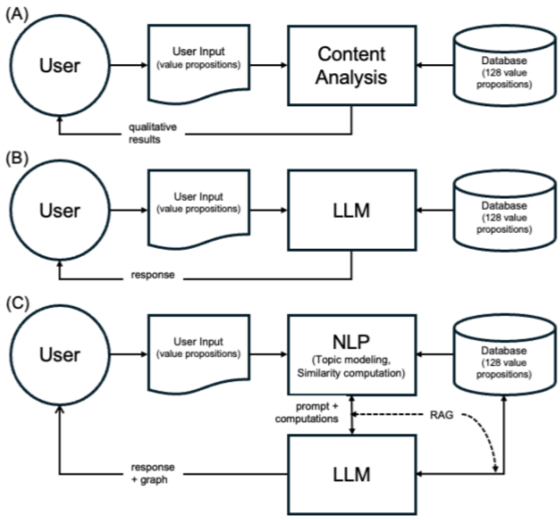

Figure 2: A comparative analysis evaluates (A) content analysis, (B) standalone LLM, and (C) EDEN.

3.1 Data preprocessing

The dataset, comprising stakeholders' value propositions (e.g., diagnostics, wellness, monitoring) and organizational details (e.g., funding amount, location, founded year), was sourced from a previous study (Mekniran, Giger, et al., 2024). Using Python's Pandas package (McKinney, 2010), this dataset was structured into a data frame for further analysis. Preprocessing steps were applied to standardize and clean the textual data, ensuring its suitability for downstream analyses:

1. Stopword Removal: Eliminating commonly used words with minimal semantic value.
2. Lemmatization: Reducing words to their root forms to ensure linguistic consistency.
3. Tokenization: Splitting text into individual words to facilitate granular analysis.

Next, the preprocessed data was vectorized using the NLTK package (Bird et al., 2009), creating numerical representations that could be analyzed using thematic clustering and network analysis. Clean data plays a crucial role in this step, as its high quality ensures that relevant relationships are effectively captured and modeled (Bird et al., 2009), see Code Snippet 1S.

3.2 Topic modeling

To analyze textual data, we employed Term Frequency-Inverse Document Frequency (TF-IDF) to construct a document-term matrix. This numerical representation assigned weights to terms based on their contextual importance, balancing term frequency within individual documents and term rarity across the entire dataset (Bengfort et al., 2018). This approach ensured that the matrix emphasized the most relevant phrases for specific stakeholders, capturing contextually significant terms.

Non-negative Matrix Factorization (NMF) was applied to the document-term matrix using the Scikit-learn package (Pedregosa et al., 2011), following Code Snippet 1S. NMF decomposed the high-dimensional document-term matrix into two non-negative, lower-dimensional matrices, identifying thematic clusters such as "digital innovation" and "community engagement." Its non-negativity constraint ensured that the resulting clusters were concise, interpretable, and suitable for the smaller datasets used in this study, see Code Snippet 1S. This decomposition is mathematically represented as (1):

$$V \approx WH \tag{1}$$

Where:
V: Original document-term matrix
W: Basis matrix representing the thematic clusters
H: Coefficient matrix indicating the document-topic associations.

Using the nmf.fit() function, the algorithm iteratively adjusted W and H to learn patterns within the data, forming a basis for identifying alignment patterns and collaboration opportunities among stakeholders. These thematic clusters served as a foundation to measure stakeholder alignment.

3.3 Network analysis

To assess thematic alignment, cosine similarity was calculated for NMF-transformed topic vectors, quantifying the degree of alignment between stakeholder value prepositions (Bengfort et al., 2018), following Code Snippet 2S. The nodes in the resulting network graph represented stakeholders, while the edges constructed an undirected network connecting stakeholders whose similarity scores exceeded a predefined threshold (e.g., the 50th percentile). Cosine similarity, a widely used metric in text analysis, quantifies thematic alignment by measuring the cosine of the angle between two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ as in (2) (Silge & Robinson, 2017).

$$\text{Cosine similarity}(A, B) = \frac{\sum A_i B_i}{\sqrt{\sum A_i^2} \cdot \sqrt{\sum B_i^2}} \tag{2}$$

3.4 Adapted RAG for recommendations engine

Within EDEN, the RAG framework retrieves curated data and network analysis outputs, which are integrated into LLM prompts that include the user's input value proposition, identified similar organizations, and suggestions for collaboration and strategy refinement (OpenAI et al., 2024). By combining retrieval with thematic analysis and LLM-driven recommendations, EDEN provides a systematic, data-driven approach to address incentive misalignments and foster partnerships, see Code Snippet 2S. Furthermore, cloud-based deployment enhances scalability and accessibility, enabling EDEN to process larger datasets and adapt to more intricate stakeholder ecosystems while supporting real-time, iterative strategy refinement (Khurana, 2014).

4 RESULTS

This section evaluates the effectiveness of EDEN in addressing research questions by comparing its ability to identify thematic clusters (RQ1) and generate actionable collaboration recommendations (RQ2) with alternative approaches: qualitative content analysis and standalone LLM. The evaluation is based on a given health insurer's value proposition; see user input in Code Snippet 3S.

4.1 How effectively does EDEN identify thematic clusters in value propositions? (RQ1)

Thematic clusters were identified using (A) qualitative content analysis, (B) standalone LLM, and (C) EDEN topic modeling. The qualitative content analysis based on existing guidelines (Fried et al., 2022) and manually coding each sentence of the user-provided text. In contrast, the standalone LLM generated thematic clusters based on its contextual interpretation of the user's input. EDEN employed a topic modeling technique, independent of external contextual information, to categorize the user's input systematically.

(A) Content analysis

1. guidelines for the well-being of older adults: 'A companion for the next stage of your life' At our health insurance, we understand that the transition to retirement requires more than just financial preparation.'

2. coordination of care delivery: 'Our digital platform is all about helping you make the new phase of your life healthy and enriching.'

3. evidence-based personalized healthcare: 'Advice from experts Benefit from personalised advice from our health coaches tailored to your needs as a retiree.' 'We offer comprehensive support for a fulfilling life in retirement. We combine content and offers on health topics that are specifically designed to meet the interests of those approaching retirement.'

(B) Standalone LLM

1. holistic support for retirement transition: emphasize the need for preparation and support during the transition to retirement life, addressing physical, mental, and social health needs.

2. health and wellness focus, personalized guidance: enabling a fulfilling and enriching lifestyle by offering tools and resources that cater to retirees' interests and aspirations.

3. expert-driven services: providing expert guidance and advice customized to individual needs for healthy aging.

(C) EDEN

1. care, health, patient, insurance, platform, virtual, support, academic, service, management

2. research, aging, funding, innovation, foster, community, institute, collaboration, program, public

3. skin, data, tool, app, recommendation, acne, therapy, analysis, personalized, artificial

4. cell, cellular, therapy, disease, human, medicine, allogeneic, mogrify, technology, type

5. test, age, blood, dna, epigenetic, based, epiage, kit, insight, methylation

6. healthcare, network, health, within, management, medical, developing, online, provider, behvavioral

7. longevity, fund, company, offer, series, venture, medicine, investment, disease, therapeutic

8. life, science, health, healthy, people, digital, solution, integrating, capability, goal

EDEN's topic modeling generated 8 clusters, each containing 10 topics derived from the user input, demonstrating comparable performance to alternative methods by identifying similar overarching themes.

4.2 How effectively does EDEN generate recommendations for collaboration? (RQ2)

Recommendations were generated respectively to alternative approaches: Manual content analysis still relied on identified guidelines and qualitative coding, while recommendations from standalone LLM were generated based on the model's contextual understanding of the user's input and given dataset (Mekniran, Giger, et al., 2024). At this stage, EDEN coupled LLM with the computed results of network analysis. EDEN for collaboration suggestions is based on network insights built on similarity computations.

(A) Content analysis

The following recommendations reflect value propositions derived from user-provided text, aligned with the principles outlined in the WHO's Decade of Healthy Ageing baseline report (WHO, 2020) and the Global Roadmap for Healthy Longevity (Fried et al., 2022).

1. Collaborate with governmental agencies: establish partnerships to develop practical, measurable indicators that focus on successful retirement, including access and availability of equitable long-term care.

2. Engage multilayered organizations: strengthen care delivery for the aging population that consolidates resources and expertise of retirement planning.

3. Partner with research centers: accelerate integrated care for retirees by using clinical trials to introduce novel interventions that address common age-related challenges.

(B) Standalone LLM

The standalone LLM generated recommendations by synthesizing several actionable collaboration areas tailored to stakeholder groups (e.g., governmental agencies, private platforms, digital health companies); the following are extracted key points from LLM responses that focused on collaboration.

1. Promote Age-Positive Attitudes and Combat Ageism: Foster partnerships between advocacy groups and NGOs to develop campaigns challenging aging stereotypes and promoting inclusive narratives. For example, AGE-WELL can work with age-tech startups to develop assistive technologies for aging populations.

2. Deliver Person-Centered and Preventive Care: Collaborate with healthcare providers and digital health companies to implement tailored health interventions integrated into existing systems. For example, InsideTracker can integrate with clinical research networks (e.g., Max Planck Institute) to validate their solutions.

3. Scale Digital Health Solutions: Partner with research institutions and private platforms to adopt AI-driven tools, wearable technology, and biomarker insights for chronic disease prevention and management. For example, Cedars Sinai to expand their reach into underserved populations.

(C) EDEN

EDEN computed a similarity score between user input "Your Idea," and government agencies was 53%, while the similarity with insurers was 52%, see Figure 3. These scores indicate alignment with thematic priorities such as public health and preventive care, offering quantitative nuanced insights.

1. Co-creating educational resources: Addressing shared challenges in stakeholder education and awareness.

2. Integrating health data from life science platforms: Enabling data-driven personalization of health services.

3. Bundling health and retirement planning services: Combining offerings to enhance scalability and address broader user needs.

5 DISCUSSION

This paper contributes to the Scale-IT-up 2025 workshop by providing a concrete example of how advanced computational tools can address complex systemic challenges in care implementation. Building on literature that emphasizes the importance of multi-stakeholder collaboration in healthy aging initiatives (Cox & Faragher, 2022; National Academy of Medicine, 2022; WHO, 2020), EDEN advances these efforts by introducing a scalable framework capable of dynamically mapping stakeholder incentives beyond alternative approaches. A comparative analysis with manual content analysis and standalone LLM approaches underscores the advantages of EDEN in identifying thematic groups (RQ1) and generating personalized collaboration recommendations (RQ2).

The evaluation demonstrates that EDEN surpasses manual and standalone LLM approaches in both granularity of value proposition analysis and specificity of its similarity. For instance, it could identify overarching themes beyond healthy aging, such as preventive care and personalized health services, and link actionable recommendations to specific stakeholders with numerical ranking. By coupling network analysis with similarity metrics, EDEN highlights high-centrality stakeholders as ecosystem integrators while uncovering hidden collaboration opportunities with isolated nodes. Unlike static frameworks or generic guidelines, EDEN provides actionable strategies, such as co-creating educational resources with government agencies and bundling health and retirement services with health insurers, tailored to stakeholder interrelations and thematic overlaps.

Limitations

Although EDEN demonstrates clear advantages, limitations remain. Reliance on curated datasets and NMF-based topic modeling may restrict scalability and semantic depth. Furthermore, the lack of broader user studies limits the generalizability of its findings. Future iterations should extend calculation to publicly available datasets and integrate a dynamic temporal

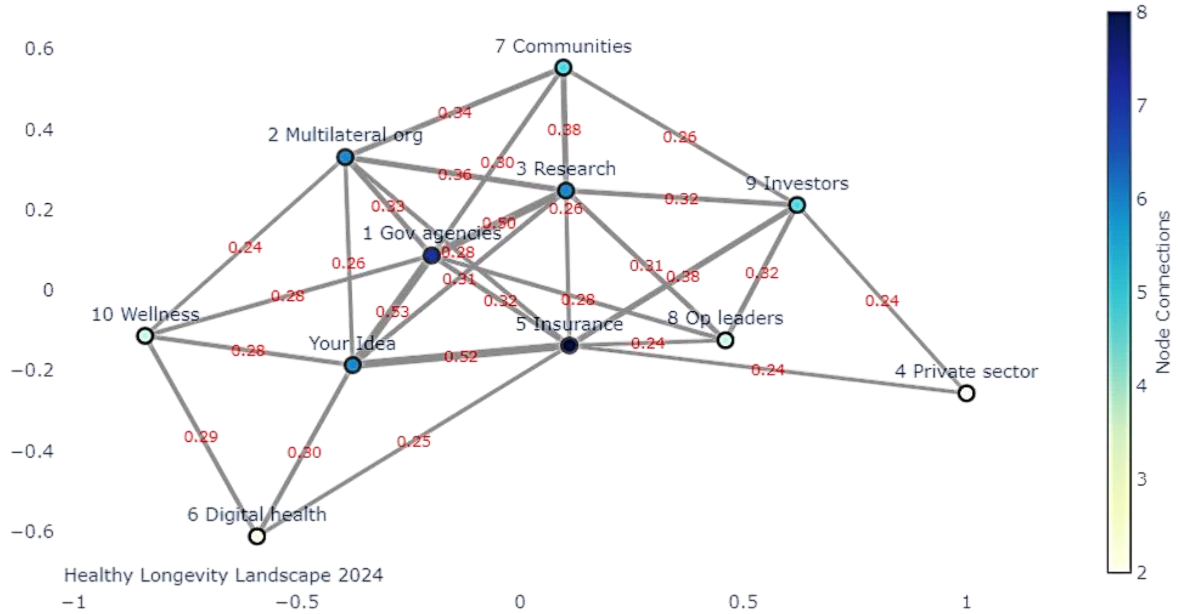

Figure 3: Visualization of stakeholders as a node and collaboration as an edge between nodes

network analysis to improve adaptability and scalability. Testing EDEN performance in various healthcare systems with user studies will also validate its practical utility and usability in varied financial and operational scenarios, such as direct to consumer model, licensing, and subscription model.

6 CONCLUSIONS

This study introduces EDEN as a computational framework to align stakeholder incentives and advance healthy aging strategies. By integrating network analysis and RAG-powered LLMs, EDEN provides granular, user-specific recommendations, surpassing manual and standalone LLM approaches in generating actionable collaboration insights. Future work should focus on scaling EDEN's application to larger datasets, incorporating advanced techniques like contextual embeddings and financial database analysis, and validating its adaptability across diverse healthcare ecosystems.

CONFLICTS OF INTEREST

WM and TK are affiliated with the Centre for Digital Health Interventions (CDHI), a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich; the Department of Management, Technology, and Economics at the Swiss Federal Institute of Technology in Zürich; and the Institute of Technology Management and School of Medicine at the University of St Gallen. CDHI is funded in part by CSS, a Swiss health insurer, MavieNext, an Austrian healthcare provider, and MTIP, an equity firm investing in European Healthtech companies. TK is a co-founder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways, he is no longer a shareholder since 2024. However, neither Pathmate Technologies, MTIP nor MavieNext were involved in this research.

REFERENCES

Altpeter, M., Schneider, E. C., & Whitelaw, N. (2014). Examining Strategies to Build and Sustain Healthy Aging Programming Collaboratives. Health Education & Behavior, 41(1_suppl), 27S-33S. https://doi.org/10.1177/1090198114537065

Béchard, P., & Ayala, O. M. (2024). Reducing hallucination in structured outputs via Retrieval-Augmented Generation. https://doi.org/10.48550/ARXIV.2404.08189

Bengfort, B., Bilbro, R., Ojeda, T., & Bengfort, B. (with Bilbro, R., & Ojeda, T.). (2018). Applied text analysis with python: Enabling language-aware data products with machine learning (First edition). O'Reilly.

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The Triple Aim: Care, Health, And Cost. Health Affairs, 27(3), 759–769. https://doi.org/10.1377/hlthaff.27.3.759

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly.

Bonnes, S. L. R., Strauss, T., Palmer, A. K., Hurt, R. T., Island, L., Goshen, A., Wang, L. Y. T., Kirkland, J. L., Bischof, E., & Maier, A. B. (2024). Establishing healthy longevity clinics in publicly funded hospitals. GeroScience. https://doi.org/10.1007/s11357-024-01132-0

Chapman, P. (2000). CRISP-DM 1.0: Step-by-step data mining guide. https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman/54bad20bbc7938991bf34f86dde0babfbd2d5a72

Cox, L. S., & Faragher, R. G. A. (2022). Linking interdisciplinary and multiscale approaches to improve healthspan—A new UK model for collaborative research networks in ageing biology and clinical translation. The Lancet Healthy Longevity, 3(5), e318–e320. https://doi.org/10.1016/S2666-7568(22)00095-2

Fried, L. P., Wong, J. E.-L., & Dzau, V. (2022). A global roadmap to seize the opportunities of healthy longevity. Nature Aging, 1–4. https://doi.org/10.1038/s43587-022-00332-7

Frow, P., & Payne, A. (2011). A stakeholder perspective of the value proposition concept. *European Journal of Marketing*, *45*(1/2), 223–240. https://doi.org/10.1108/03090561111095676

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey (Version 5). arXiv. https://doi.org/10.48550/ARXIV.2312.10997

Giger, O.-F., Pfitzer, E., Mekniran, W., Gebhardt, H., Fleisch, E., Jovanova, M., & Kowatsch, T. (2024). Collaboration and Innovation Patterns in Diabetes Ecosystems. https://doi.org/10.1101/2024.04.25.24306351

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. Qualitative Health Research, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huang, Y., & Huang, J. (2024). A Survey on Retrieval-Augmented Text Generation for Large Language Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2404.10981

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. Multimedia Tools and Applications, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Khurana, A. (2014). Bringing Big Data Systems to the Cloud. IEEE Cloud Computing, 1(3), 72–75. https://doi.org/10.1109/MCC.2014.47

Larose, D. T. (2015). *Data Mining and Predictive Analytics* (1st ed). John Wiley & Sons, Incorporated.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Scipy. https://doi.org/10.25080/Majora-92bf1922-00a

Mekniran, W., Giger, O.-F., Fleisch, E., Kowatsch, T., & Jovanova, M. (2024). The Longevity Landscape: Value Creation for Healthy Aging. https://doi.org/10.1101/2024.05.28.24308017

Mekniran, W., & Kowatsch, T. (2023). Scalable Business Models in Digital Healthy Longevity: Lessons from Top-Funded Digital Health Companies in 2022. 609–615. https://doi.org/10.5220/0011778400003414

Mekniran, W., Kramer, J.-N., & Kowatsch, T. (2024). Reimagining Preventive Care and Digital Health: A Paradigm Shift in a Health Insurance's Role. 852–858. https://doi.org/10.5220/0012400300003657

National Academy of Medicine. (2022). Global Roadmap for Healthy Longevity (p. 26144). National Academies Press. https://doi.org/10.17226/26144

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). GPT-4 Technical Report (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON.

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach (First edition). O'Reilly.

Sharma, A., & Kaur, B. (2017). A Research Review on Comparative Analysis of Data Mining Tools, Techniques and Parameters. *International Journal of Advanced Research in Computer Science*, *8*(7), 523–529. https://doi.org/10.26483/ijarcs.v8i7.4255

Valdez, D., Pickett, A. C., & Goodson, P. (2018). Topic Modeling: Latent Semantic Analysis for the Social Sciences. Social Science Quarterly, 99(5), 1665–1679. https://doi.org/10.1111/ssqu.12528

Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

WHO. (2020). Decade of healthy ageing: Baseline report. World Health Organization. https://apps.who.int/iris/handle/10665/338677.

APPENDIX

Code Snippet 1S: Text preprocessing and topic modeling

```
from nltk.corpus import stopwords
from nltk.stem.wordnet import
WordNetLemmatizer
from nltk.tokenize import word_tokenize
```

```
def preprocess(text):
    tokens =
word_tokenize(text.lower())
    tokens =
[WordNetLemmatizer().lemmatize(token)
for token in tokens if token.isalpha()
and token not in
set(stopwords.words('english'))]
    return ' '.join(tokens)

# Topic modeling with TF-IDF and
training NMF
from sklearn.feature_extraction.text
import TfidfVectorizer
from sklearn.decomposition import NMF

# Vectorize the text data
doc_term_matrix =
TfidfVectorizer().fit_transform(df['Wha
t'].apply(preprocess))

# Apply NMF for topic modeling
nmf = NMF(n_components=8,
random_state=42).fit(doc_term_matrix)
```

Code Snippet 2S: Similarity computation and recommendations engine.

```
import networkx as nx
import numpy as np
from sklearn.metrics.pairwise import
cosine_similarity

# Calculate cosine similarity
org_topic_matrix =
nmf.transform(doc_term_matrix)
cosine_sim =
cosine_similarity(org_topic_matrix)

# Build the graph
G = nx.Graph()
for i, seg1 in
enumerate(df['Organization']):
    for j, seg2 in
enumerate(df['Organization']):
        if i < j and cosine_sim[i][j] >
np.percentile(cosine_sim, 50):
            G.add_edge(seg1, seg2,
weight=cosine_sim[i][j])

# Recommendations engine with RAG
import openai

# Define the prompt
collaboration_prompt = (
"Based on the identified similar
organizations and their value
propositions, provide succinct
suggestions in bullet points on how to
design a value proposition and
collaborate with these organizations. "
"Consider the following aspects: 1. Key
elements to include in the value
proposition. 2. Potential areas of
collaboration. 3. Benefits of
collaboration for both parties."
)

response =
openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
    {"role": "system", "content": "You
are an AI that helps identify how to
collaborate with other stakeholders to
enable healthy longevity." },
    {"role": "user", "content":
collaboration_prompt + "Your value
proposition now: "+ user_input + "Your
identified similar organizations: "+
similar_scores }
    ],
)
```

Code Snippet 3S. User input prompt (151 words)

```
"A companion for the next stage of your
life
At our health insurance, we understand
that the transition to retirement
requires more than just financial
preparation. Our digital platform is
all about helping you to make the new
phase of your life healthy and
enriching.
Our service in a nutshell
Holistic preparation and health
We offer comprehensive support for a
fulfilling life in retirement. We
combine content and offers on health
topics that are specifically designed
to meet the interests of those
approaching retirement.
Advice from experts
Benefit from personalised advice from
our health coaches tailored to your
needs as a retiree.
Pilot phase and availability
Your health partner at any age
As your healthcare partner, we go
beyond the role of a health insurer and
is actively committed to healthy
ageing. Our service is an expression of
this commitment, with the aim of
supporting you in all phases of life."
```

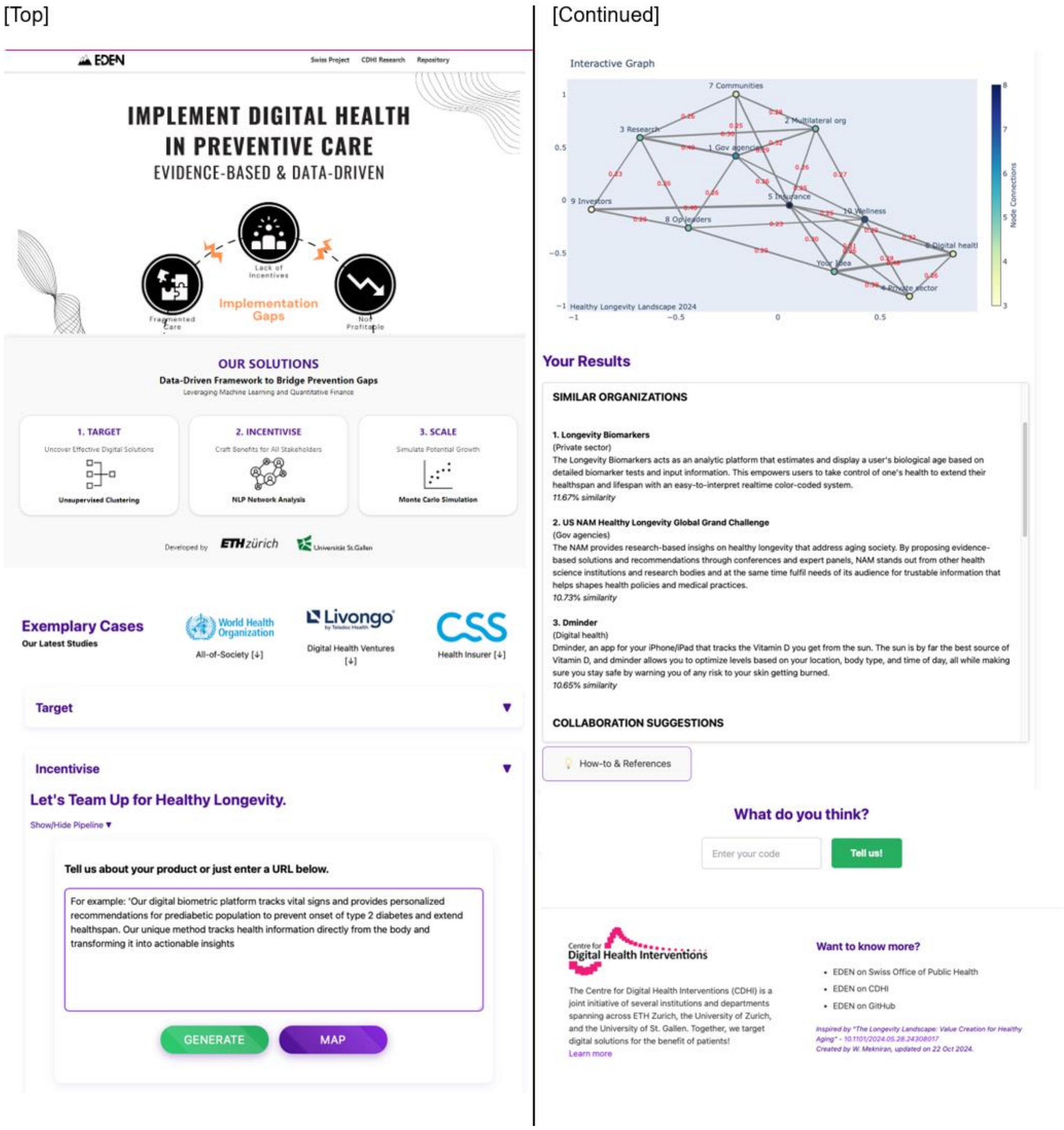

Figure 4S: EDEN's user interface (eden.ethz.ch), users can input value propositions within 'incentivise' module to generate network map and suggestion for collaboration. 'target' and 'scale' modules are under development to further EDEN's computational capabilities, accessed on 12 December 2024.