Proto-OKN Theme 1: Creating A Cross-Domain Knowledge Graph To Integrate Health And Justice For Rural Resilience

Overview

Enhancing the resilience of our nation's rural communities to existing and upcoming crises is of pressing importance, as these areas are vital for preserving essential resources such as air quality, water supplies, food production, and supply chains - a realization underscored by the CoVID-19 pandemic. Following our preliminary work, we have pinpointed collaborators and mapped out the capabilities required for rural resilience, particularly in the context of public health and environmental crises. Central to these requirements of resources and community collaborations is the ability to collect and analyze data related to health outcomes and climate changes, along with social determinants of health and justice within rural locales.

Therefore, we have put together a specialized team that's dedicated to overcoming difficulties in data gathering and analysis. The method involves the construction of an extensive, interdisciplinary knowledge graph. This graph is designed to merge, portray, and interconnect previously separate health and justice data sets. It's a powerful resource designed to aid researchers, practitioners, and educators in improving their understanding of risk environments in rural locations and strengthening their resilience. The project's goal is to make use of existing geo-enrichment services and initiatives like the NSF-funded KnowWhereGraph. The team is devoted to synchronizing our efforts with these programs, as well as other Proto-OKN themes, in order to enhance our scientific studies of rural resilience to public health and environmental crises.

Intellectual Merit

We envision this research will advance our knowledge in computer & Information science and social & behavioral science. **Contributions to Computer Science:** (1) Creating a multidisciplinary knowledge graph to integrate diverse datasets with consideration of human, social, and organizational factors signifies a progressive and innovative approach to data management and analysis. and (2) advancing data amalgamation methods to address inherent challenges in data management, such as heterogeneity, sparsity, and privacy, through the implementation of transfer and generative learning mechanisms across a variety of datasets, with an emphasis on social and cultural context. **Contributions to Social Science:** (1) Establishing a novel approach to Community-Based Participatory Research (CBPR) that forms the foundation of our project, paving the way toward impactful and sustainable outcomes. (2) Forging a deeply integrated workflow that synergizes CBPR with developing the knowledge graph to actively engage domain knowledge experts and a broad spectrum of constituents.

Broader Impacts

A comprehensive and accessible knowledge graph will act as a navigational tool, enabling effective resource allocation, policy development, and partnership establishment within rural communities. It will notably empower underrepresented groups, providing them a platform to assert their perspectives and influence decisions of substantial significance. To ensure the widespread dissemination of the knowledge products derived from this project, our outreach initiatives will target a diverse group of stakeholders, including interdisciplinary researchers, educators, students, school administrators, and industry partners across multiple practice domains. This project will generate a robust collection of learning science methodologies, computational models, design frameworks, software artifacts, and empirical data, thus providing a rich resource for the interdisciplinary educational research community to further advance teaching and learning in knowledge graphs.

TABLE OF CONTENTS

For font size and page formatting specifications, see PAPPG section II.B.2.

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

Project Description

1. Integrative Sociotechnical Research And Development

1.1 Significance Of An Integrated, Community-Centered Approach

Our nation's rural communities, central to preserving vital resources like air quality, water supplies, food production, and supply chain, often bear the brunt of public health and environmental crises [1–4]. The recent pandemic crisis underscored the need to proactively address pressing concerns through initiatives that leverage extensive data generated across all life aspects, thereby strengthening these communities' resilience to ongoing and future crises [5–7]. Simultaneously, it is crucial to recognize that health and justice are pivotal for fostering resilience in rural areas, particularly in communities with the lowest socioeconomic status [8, 9]. These communities, often marred by high poverty and uninsured rates, limited education, and healthcare access, are critically underrepresented in data and related workflow [8, 10, 11]. This under-representation manifests as challenges in data availability, quality, standardization, integration, limited sample size, privacy, confidentiality, and bias [12–15]. Hence, a novel approach, such as open knowledge graph [16], would be necessary that amalgamates diverse datasets, providing analysts with comprehensive contextual information. This methodology, paired with technology enabling densely integrated, cross-domain data, can cultivate reliable, mutually beneficial, and cooperative partnerships to boost the resilience of rural areas to both present and forthcoming public health and environmental crises [17, 18]. This perspective has prompted the NSF to advocate for "harnessing the vast amounts of data generated in every sphere of life and transforming them into useful, actionable information and knowledge".

This project is geared towards addressing these challenges by creating a comprehensive, cross-domain knowledge graph. This graph will combine, represent, and interconnect formerly separated health and justice datasets, providing a powerful tool to support researchers, practitioners, and educators in their efforts to deepen their understanding of risk environments in rural regions and fortify their resilience. The project plans to harness existing geo-enrichment services and agencies like the NSF-funded KnowWhereGraph [19] and the Resilient Rural America Project for Extreme Weather [20]. It intends to align with these programs and other research-related themes to augment our scientific endeavors. Additionally, the project aims to broaden its dataset collection by incor-

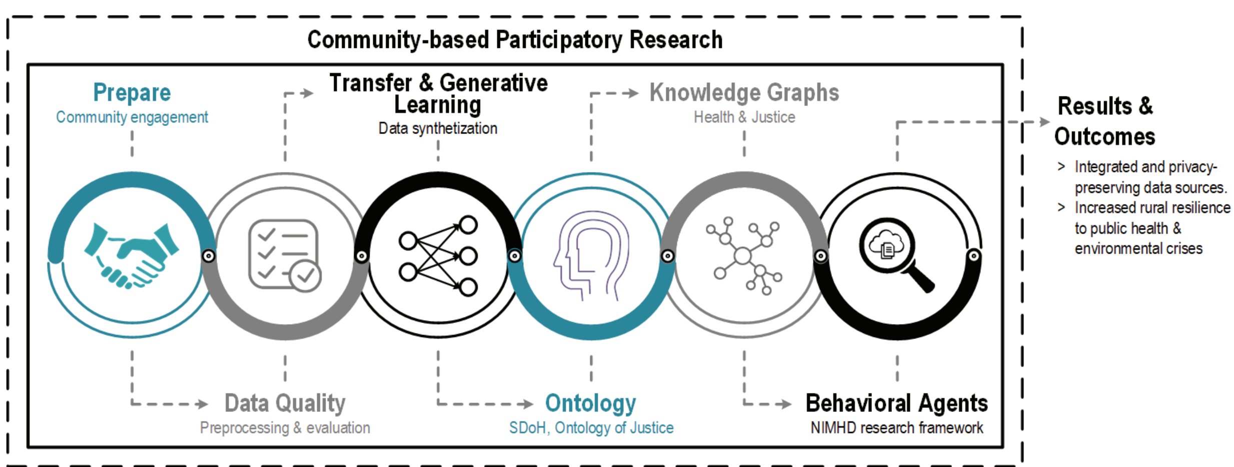

Figure 1: Conceptual representation of the proposed integrated, community-centered approach.

porating the CDC's Social Determinants of Health (SDoH) and PLACES data, electronic health records (EHR) from Medicare and Medicaid, data from the National Institute on Minority Health and Health Disparities' Science Collaborative for Health Disparities and Artificial Intelligence Bias Reduction (ScHARe), along with justice datasets from the Inter-University Consortium for Political and Social Research (ICPSR).

As depicted in Figure 1, the data science workflow commences with a preparatory phase, which encompasses data preprocessing, self-evaluation, team formation, and community interaction. This stage constitutes a pivotal step for human-centered design methods, known as community-based participatory research [21], that lie at the heart of this project and encourages the involvement of local community partners and end users. Following the formation of the team and community engagement, the subsequent stage involves data quality evaluation concerning aspects like sparsity, availability, cost-effectiveness, standardization, limited sample size, privacy, confidentiality, and bias. Drawn from our empirical observations and preliminary work [22], a majority of the existing datasets tend to be sparse for rural areas, and there are data gaps due to various reasons, such as affordability [12–15]. Consequently, the ensuing step is to employ transfer learning and generative techniques to create synthetic data, thereby enhancing data quality while ensuring privacy [23, 24]. Upon augmenting the data, we will formulate an ontology related to public health and environmental crises to generate sub-graphs that bolster the integrated level of the knowledge graph. Primarily, the anticipated knowledge graph will be employed to depict the health and justice aspects of determinants, as illustrated in Figure 2. At the concluding stage, multi-scale data synthesis models, such as behavioral agents, will be developed, utilizing the knowledge graph and the generative techniques, to comply with the guidelines of the National Institute on Minority Health and Health Disparities (NIMHD) Minority Health and Health Disparities Research Framework [25], which is increasingly being embraced by numerous end users.

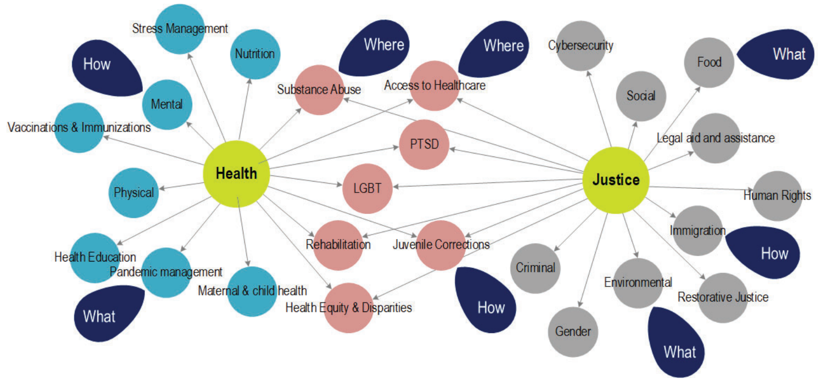

Figure 2: An example of triples from the proposed knowledge graph about substance abuse, impacted areas, and related health and justice aspects.

Figure 2 illustrates a selection of triples from the anticipated knowledge graph, focusing on rural communities impacted by the substance abuse crisis [26], along with health, justice aspects, and services concerning substance abuse. To ensure seamless integration with existing knowledge graphs, particularly the NSF-funded KnowWhereGraph, we will utilize the identical grid system known as the "S2 Grid System" or Discrete Global Grid [27], for the portrayal of spatial regions. This system applies a hierarchical grid to the earth's surface, where each grid cell at one level consists of four subordinate cells at a higher spatial resolution. In comparison to the KnowWhereGraph, our data will be served at a slightly superior resolution, specifically at least at the S2 level 16 (roughly 1200 square meters per cell) for the United States. The rationale behind this approach is our necessity to gather more detailed information on the socioeconomic, health, and justice circumstances of communities and their residents, including aspects like housing and transportation [22]. Employing the S2 grid system, we will create a design pattern that simplifies the process of understanding how features and regions could interact across the hierarchy. Furthermore, we plan to implement several open standards, including the WHO's Conceptual Framework for Action on the Social Determinants of Health, the CDC's Social Vulnerability Index (SVI), the NIMHD Minority Health and Health Disparities Research Framework, and the US Department of Justice's (DOJ) National Information Exchange Model.

1.2 Social Aspects Of The Proposed Research

The project will establish a new instance within the scope of community-based participatory research (CBPR) to engage a broad range of partners and end-users. It will actively engage with aware and affected communities to ensure the inclusivity of data, users, and involved communities. Moreover, the CBPR will work with these communities to protect individuals' privacy, civil rights, and liberties [21, 28], thus facilitating the creation of an ethically sound and responsible Proto-OKN. As illustrated in Figure 3, compared to other CBPR instances, our proposed CBPR leverage the knowledge graphs and the computational agents to support social-economic research. The initial phase of the CBPR involves identifying regional populations that haven't yet been reached by community-based efforts. Following this, we plan to gather insights and viewpoints from community members, utilizing methods like in-depth interviews, focus group discussions, and online surveys. The amalgamation of the knowledge graph and CBPR will take place during the data collection and analysis stage, prioritizing health and justice facets such as housing, population density, availability of insurance, and healthcare accessibility.

For data collection and curation in the CBPR process, we'll employ a mixed-method approach that encompasses recognized methods in socio-economic research, such as statistical testing [29], the general linear model [30], and factor analysis [31], along with machine learning techniques like classification, clustering, and time-series simulation. Furthermore, we'll develop a blend of informal and formal strategies. The informal approach will offer a short-term response to community changes in a flexible manner, whereas the formal strategies are meant to guarantee the sustainability and scalability of our outreach efforts. Figure 4 depicts the planned

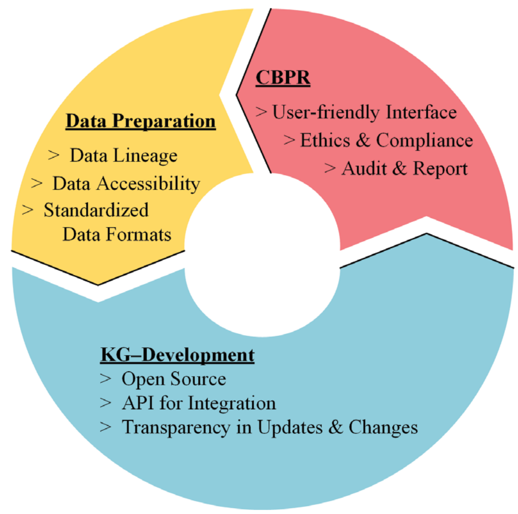

Figure 4: Openness and Transparency during Workflow of data, knowledge graph, and CBPR.

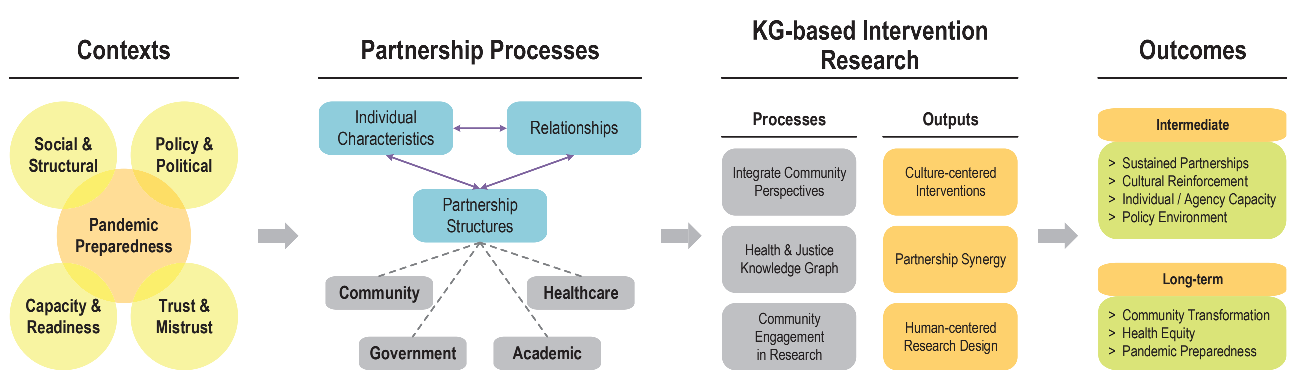

Figure 3: An example of a community-based participatory research framework for a use case of examining pandemic preparedness.

life cycle of the project development process, including strategic measures to prevent the risk of cultivating a closed, proprietary system that may only cater to a select group. It also highlights the endeavor to enhance transparency. During the data preparation phase, we will employ standardized data formats [32, 33], ensuring that data can be effortlessly accessed, comprehended, and utilized by all system users. We will also put into action a data governance plan [34] to safeguard data availability, usability, integrity, and security, while also confirming the data's accessibility to all pertinent stakeholders. Further, we will maintain a detailed record of data lineage [35], tracing its origin, destinations, and alterations throughout its lifecycle. In the knowledge graph development phase, we will create open-source software, leveraging sharing platforms such as GitHub and The Open Science Framework [36]. We will also provide APIs to allow seamless data integration from external systems. Whenever updates or modifications occur within the system, we will disseminate clear and comprehensive release notes detailing what has been altered, the reasons behind the changes, and how these changes affect the users. Finally, the CBPR will aid in refining the system interface to ensure user-friendliness, compliance with ethical standards, and adherence to regulatory norms. Regular online workshops will also be conducted to carry out system audits, the reports of which will be shared with users and stakeholders.

1.3 Technological Dimensions

Use Cases and Capability Requirments. Our team has collaborated closely with end users, comprising entities like the National Association of Social Workers, the Department of Public Health, Emergency Medical Services, the Opioid Overdose and Addiction Council, and the Department of Veterans Affairs. We have partnered with them to pinpoint specific scenarios that can enhance rural resilience, particularly in relation to public health and environmental crises. The focus areas within public health crises have been mental health emergencies [37], substance abuse issues [38], disease pandemics [39], and epidemics, as well as challenges in vaccine uptake [40]. In the context of environmental crises, we've concentrated on climate change [41], drought conditions, and flood events [42].

After consolidating the capabilities required for each use case to bolster rural resilience in public health and environmental crises, we've categorized them into two primary groups [1, 4, 8]. The first category pertains to resources, while the second is associated with community engagement and education. To illustrate, in the context of public health crises, elements like healthcare infrastructure, medical personnel, telehealth facilities and internet capacity, emergency readiness, and mental health services fall under the resource-related category. This area requires consistent financial backing for advancement. On the other hand, facets such as education and public awareness, community participation, and policy advocacy fall under the community engagement and education category. This area necessitates the forging of robust partnerships across various sectors, including health, justice, and social services, and at different governmental tiers, as well as cooperation with non-profit organizations, academic institutions, and the private sector.

Underpinning these requirements of resources and community collaborations is the ability to gather and interpret data pertaining to health outcomes and climate alterations, accessibility to care and emergency services, as well as social determinants of health and justice within rural areas. This capability can serve as a compass, guiding the effective and efficient deployment of these resources, the shaping of policies, and the forging of partnerships [43, 44]. Although the ability to gather and analyze data is crucial, rural communities and their governance structures continuously grapple with obstacles and limitations [12–15]. Firstly, the current state of data collection in these communities is quite scant due to cost-related factors and affordability issues. Secondly, the existing data silos are segregated across various domains and cultural perspectives, preventing a comprehensive view. Thirdly, there is a dearth of efficient data governance plans and practices.

Consequently, a majority of the data are managed centrally, leading to scalability issues and delays in updates when attempts are made to include more diverse data. Lastly, there are minimal efforts to integrate this disparate data, which hampers the possibility of conducting detailed follow-up inquiries using the enriched data.

Data Sources and Their Utility. Building resilient rural communities is deeply intertwined with public health and the principles of social, economic, and environmental justice. Rural re-

Health Datasets

Dataset Name/Theme	Source Agency	Key Attributes
SDoH	CDC	Socioeconomic status
		Education level
		Employment status and job security
		Food insecurity
		Housing stability
		Access to healthcare
		Neighborhood, housing, transportation, and physical environment
		Social support and community inclusivity
PLACES/500 cities	CDC	Health outcomes (e.g., prevalence of various health conditions such as heart disease, diabetes, chronic obstructive pulmonary disease)
		Prevention measures (e.g., rates of health screenings, vaccinations)
		Unhealthy behaviors (e.g., smoking, lack of physical activity, poor diet)
		Clinical care factors (e.g., access to and quality of healthcare, health insurance status)
		Physical environment factors (e.g., air and water quality, housing conditions)
ScHARe	NIMHD	Economic Stability
		Education Access and Quality
		Health Care Access and Quality
		Neighborhood and Built Environment
		Social and Community Context
		Health Behaviors
		Diseases and Conditions
		Clinical Data/EHRs
The National Health Care Surveys (NAMCS, NHAMCS, and NHCS)	CDC	Demographics
		Health Status and Conditions
		Health Care Services
		Provider Characteristics
		Medications
		Patient Experiences and Outcomes
Behavioral Risk Factor Surveillance System (BRFSS)	CDC	Demographics
		Health Status
		Healthcare Access/Coverage
		Chronic Health Conditions
		Health Behaviors
		Injury
		Preventive Practices
Medicare/Medicaid	ADPH	Beneficiary Variables
		Claim Variables
		Provider Variables
		Plan Variables
		Prescription Drug Variables

Justice Datasets

Dataset Name/Theme	Source Agency	Key Attributes
National Incident-Based Reporting System (NIBRS)	ICPSR	Incident Variables
		Offense Variables
		Victim Variables
		Offender Variables
		Arrestee Variables
		Property Variables
National Judicial Reporting Program (NJRP)	ICPSR	Demographics
		Offense Characteristics
		Legal Variables
		Sentencing Information
		Criminal History
		Case Processing Information
National Crime Victimization Survey (NCVS)	ICPSR	Household Variables
		Person Variables
		Victimization Variables
		Incident Variables
		Crimes of Interest
Survey of Inmates in State and Federal Correctional Facilities	ICPSR	Demographics
		Educational and Economic Background
		Physical and Mental Health
		Offense and Sentence
		Incarceration Experience
		Family Background
National Survey on Drug Use and Health	SAMHSA	Demographics
		Substance Use
		Substance Use Disorders
		Mental Health
		Healthcare Utilization
		health status, disability, HIV risk behavior, pregnancy, and health insurance coverage

Environmental Datasets

Dataset Name/Theme	Source Agency	Key Attributes
FAOSTAT	FAO	Crops, Livestock, Fisheries and Aquaculture, Land Use, Trade, Prices, Food Balance, Emissions, Population
Global Historical Climatology Network	NOAA	Temperature, Precipitation, Wind, Pressure, Humidity, Social Radiation, Weather Phenomena, Climate Indices
National Agricultural Statistics Service (NASS)	USDA	Farm Economics, Agricultural Census, Farm Practices, Rural Demographics, Environmental and Natural Resources, Agribusiness
CCAFS-Climate	CCAFS	Climate, Socioeconomic, Land Use and Land Cover, Emissions
Transportation	DOT	Road Type, Road Length, Road Sign
Land Use	NASA	MODIS Land Cover Type/Dynamics
Water Availability	USGS	Streamflow data, groundwater data, and water quality data
National Survey of American Life (NSAL)	ICPSR	Cultural, mental health, and physical health outcomes
The Panel Study of Income Dynamics (PSID)	ICPSR	economic, health, and social behavior of individuals and families

Place-Centric Datasets

Place-Centric Dataset	Defining Authority	Spatial Coverage
S2 Cells	Google	Lvl 16 (US)
Building Footprints	Microsoft	Buildings (US)
ZIP	ZCTA	US
Census Metropolitan Area	US Census	US
Climate Division	NOAA	US
US Federal Judicial District	DOJ, ESRI	US
Drought Zone	NDMC, USDA. NOAA	US
Geographic Name Information System	USGS	US
FIPS Codes	NRCS	US
National Weather Zones	NOAA	US
Global Administrative Regions	University of Berkeley…	Global

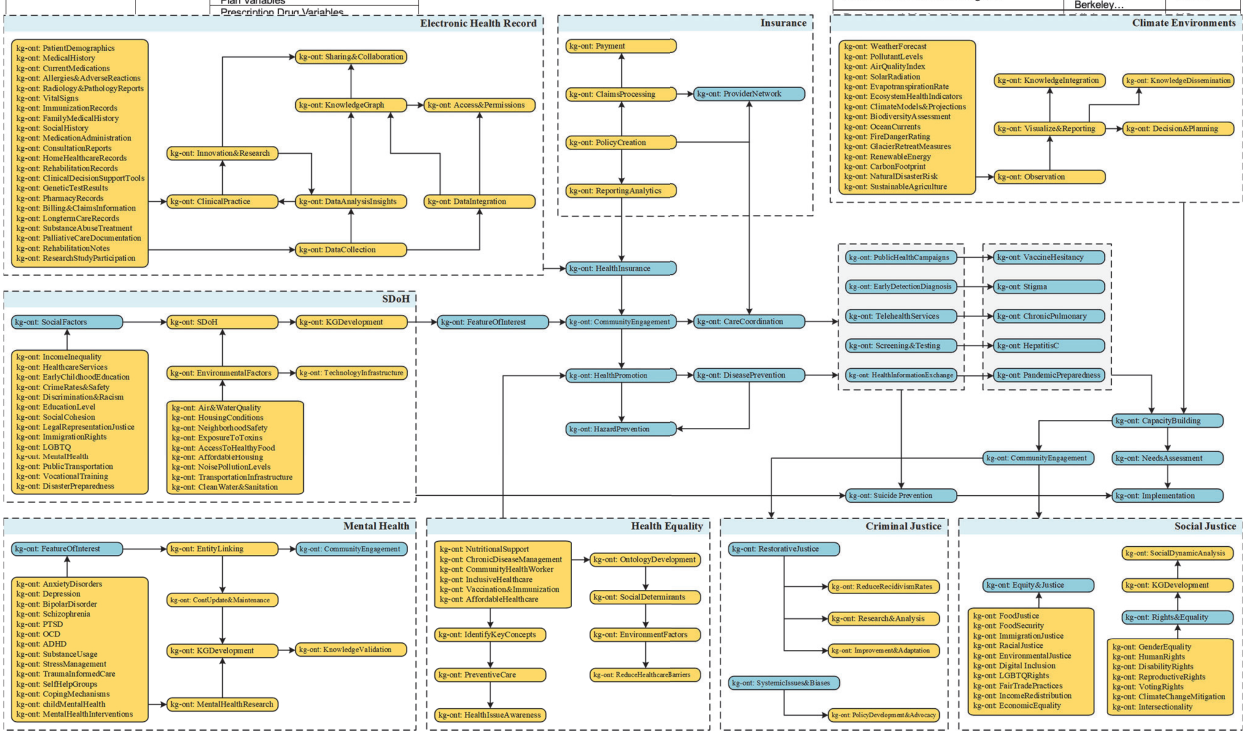

Figure 5: A summary of the raw datasets and their providers, and an example of knowledge graph.

silience, particularly in the face of public health and environmental crises, embodies the capacity of these communities to not only anticipate and prevent potential hazards but also to effectively manage, respond to, and recover from various health and environmental disruptions [5–7]. The foundation of such resilience is built upon a range of health and justice-related pillars. These include the presence of a robust healthcare infrastructure, investment in public health education and training, effective environmental management strategies, preparedness and resources for emergency situations, an ethos of participatory decision-making, and a steadfast commitment to health equity and access to justice [45, 46]. Therefore, the project aims to consolidate and integrate various public datasets that are relevant to the factors outlined, which are displayed in Figure 5, along with an illustrative example of a knowledge graph.

- Health data on equity and mental health disparities, primarily provided by the CDC and NIMHD, will incorporate crucial insurance information from the Alabama Department of Public Health (ADPH). The ADPH, a critical end user of this data, will focus on addressing substantial public health issues like substance abuse [47], vaccine hesitancy [48, 49], and suicide [50], especially in Alabama's underrepresented Black Belt communities.
- Justice-related datasets will explore the intricate interplay between social and criminal justice, health equity, and mental health disparities. The chief provider of this data is the Inter-university Consortium for Political and Social Research (ICPSR). The National Survey on Drug Use and Health will also be incorporated to refine use case scenarios. Individuals entangled in the criminal justice system frequently encounter obstacles to healthcare access, and incarceration can lead to lasting adverse health effects [51]. Furthermore, communities of color have been disproportionately affected by policies associated with the War on Drugs, intensifying existing health disparities [52].
- Datasets related to environmental issues will scrutinize the effects of environmental crises—ranging from climate change, pandemics, land utilization, and water scarcity—on health equity, as well as social and criminal justice. Climate change events, such as heatwaves, floods, and hurricanes, often impose a disproportionate burden on lower-income and ethnic minority communities, thereby magnifying pre-existing health inequalities. Other crises, like pandemics, have the potential to catalyze spikes in unemployment and crime rates, including intimate partner violence, serious assaults, and homicides, due to the disturbances they cause in social structures and services [53].

Data Representation and Quality Assurance. The aforementioned datasets, which explore the intricate intersections of health and justice within the framework of public health and environmental crises in rural communities, encapsulate an extensive and sophisticated array of knowledge encompassing diverse subjects, clusters of concerns, and their interconnections. During the creation of the proposed knowledge graph, the validation of the graph structures will likely play a vital role. Consequently, this project will leverage OWL to define the semantics of the data, as depicted in Figure 6. Additionally, we'll employ SHACL to validate the data against specific shapes or constraints, supplementing the capabilities of OWL [54, 55].

The implementation of quality assurance for both schema and entity mappings will be carried out via a cycle of validation measures. These will encompass validation of the schema, verification

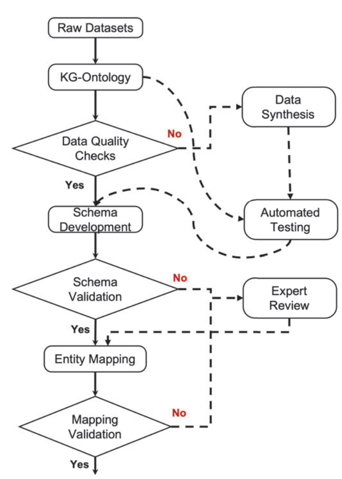

Figure 6: Quality assurance process.

of entity mapping, assessment of data quality, and execution of automated tests, in addition to a review from an interdisciplinary perspective. The iterative process of refinement is graphically presented in Figure 6. In a bid to guarantee open accessibility to both the employed methods and the derived results, comprehensive documentation of all these stages will be conducted, and this record will be made openly available to the public.

Query within the NIMHD Research Framework. The query design for the proposed knowledge graph will adhere to the National Institute on Minority Health and Health Disparities Research Framework [25], which encompasses two dimensions: levels of influence and domains of influence. The levels of influence span from individual persons to societal factors, while the domains of influence include biological, behavioral, physical/built environment, sociocultural environment, and the healthcare system.

As part of the project's scope, the query design will be specifically tailored to the individual, interpersonal, and community levels of influence, aligning with the interests and needs of the end users. This focus ensures that the queries generate valuable insights regarding individuals, their interactions, and the broader community.

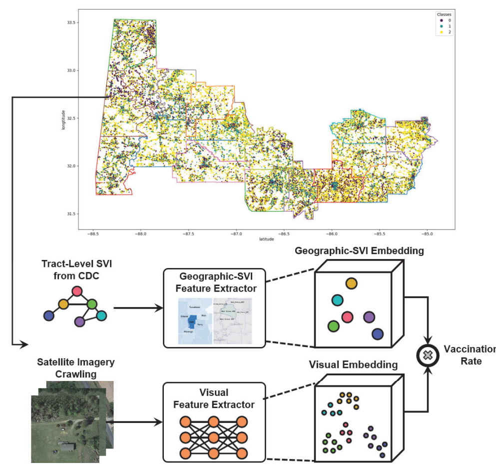

Figure 7: Our previous work integrated CDC's SVI and remote sensing data to understand individual-level vaccine hesitancy in Alabama's Black Belt region.

In the context of vaccine hesitancy, the national alliance of social workers seeks to optimize the allocation of mobile healthcare units to under-served communities [56]. To make well-informed decisions, leveraging the knowledge graph and querying it for individual-level information on vaccine hesitancy in targeted regions becomes invaluable. This data-driven approach enables the identification of areas where hesitancy is more prevalent, facilitating strategic and effective resource allocation, such as mobile healthcare units. As a result, the specific needs of communities are addressed, leading to enhanced access to vaccinations. Our previous work [22], depicted in Figure 7, showcases the development of individual-level information on vaccine hesitancy across the Alabama Black Belt, further emphasizing our commitment to tackling this issue. Another example is substance abuse, which is more prevalent in rural areas with lower socioeconomic status, such as Alabama [47]. To promote public education and training on substance abuse, a telehealth program with incentives may be implemented. For the success of such a program, community-level knowledge graph insights are essential. This would include critical ontologies related to health literacy, access to health infrastructure, social and cultural influences, and the corresponding social and religious networks [47]. The knowledge graph would provide a comprehensive understanding of these factors, empowering the development of targeted interventions and support systems for the affected communities.

User Interfaces and User-centered Design. Designing a user interface for a knowledge graph database that incorporates various geographic scales and data domains necessitates careful attention to data visualization, search capabilities, and user interactions. Our approach involves leveraging maps and geographic visualizations to depict the levels of influence, as outlined in the NIMHD research framework.

The interface will be interactive, allowing users to explore different regions and delve into specific levels and areas of interest. Alongside the geographic data, we will utilize charts, graphs, and infographics to showcase the relationships between entities such as socioeconomics, health, and justice data. To enable efficient searching, we will implement search functionalities using the Simple Protocol and RDF Query Language (SPARQL) as well as harness the capabilities of large language models (LLMs). We will fine-tune the LLM to support natural language capabilities in search, empowering users with intuitive query abilities. Additionally, we will provide tailored guidance to enhance users' proficiency in utilizing advanced search filters. In the pursuit of enhancing creativity and user acceptance, we will employ generative AI tools to design the user interface. This will facilitate the generation of innovative interface components and layouts. Figure 8 showcases an example of a generatively designed web-based interface for the proposed knowledge graph database.

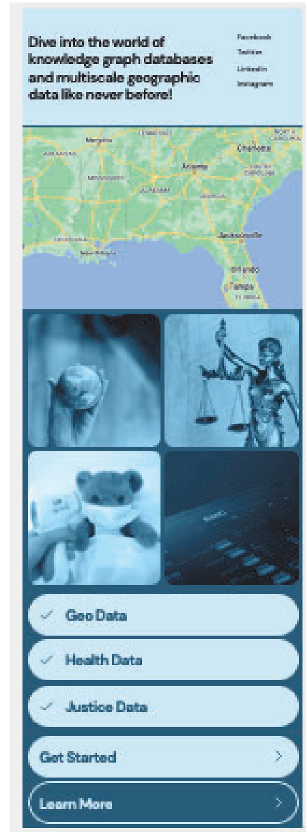

Figure 8: An example of a web-based interface for the proposed knowledge graph.

To ensure a user-centered approach, we will adopt participatory design practices throughout the project, actively involving end users in the design process of the user interfaces. Their insights and feedback will be invaluable in refining the interface design and overall user experience. As outlined in Section 5, the Evaluation Plan, we will conduct a user-centered design and usability test to comprehensively assess the usability of the web-based database service. This design process will involve recruiting participants representative of the target user group and evaluating specific usability aspects, such as ease of use, navigation, data entry, and search functionality. By incorporating end users' perspectives and conducting usability tests, we aim to create a highly usable and intuitive web-based database service.

Data Synthesis and Computational Agents via Transfer and Generative Learning. Our previous work focused on vaccine hesitancy and highlighted the challenges we encountered when attempting to collect and integrate health and justice data in rural areas. These challenges arise due to factors such as lower population density, limited resources, and difficulties in data collection. As a result, the availability of data is severely limited, leading to data sparsity. This sparsity poses a significant hurdle in developing accurate and robust knowledge graphs that rely on large and diverse datasets.

In order to overcome the challenges posed by data sparsity in rural areas, the implementation of data synthesis efforts using transfer learning and generative learning techniques becomes crucial. Transfer learning facilitates the utilization of knowledge and patterns acquired from one domain or dataset to enhance performance in another domain or with limited data [57]. By trans-

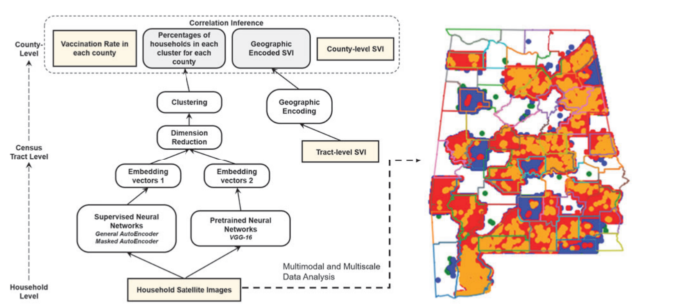

Figure 9: Our preliminary work in transfer learning for CoVID-19 vaccine hesitancy in Alabama.

ferring knowledge from well-resourced or related areas, models can leverage pre-existing information and adapt to the specific characteristics of rural data. Additionally, generative learning techniques [13] such as generative adversarial networks (GANs) [58] or variational autoencoders (VAEs) can be employed.

These methodologies facilitate the production of synthetic data that closely mimic the traits and patterns of actual rural data across varying levels of influence, from the individual to the community [59, 60]. They also provide the capability to introduce behavioral agents, which could be utilized in simulated studies focused on Community-Based Participatory Research (CBPR) interventions [60–62]. By generating synthetic data, it becomes possible to increase the overall dataset size, thereby enabling more reliable model training and evaluation processes. Through the integration of transfer learning and generative techniques, we can mitigate the impact of data sparsity and enhance the effectiveness of knowledge graph development in rural areas. Figure 9 illustrates our previous implementation of transfer learning to infer individual-level information regarding vaccine hesitancy across the state of Alabama.

Enhancing Query Interfaces for Rural End Users using LLMs. Integrating knowledge graphs and LLMs poses significant challenges within the scientific community [63, 64]. Thus, this project will primarily focus on exploring querying methods that leverage the emerging capabilities of LLMs in natural language processing [65, 66]. Notably, the project will specifically address the unique context of rural areas in the United States, particularly the underrepresented communities, such as the Alabama Black Belt region. This region predominantly comprises residents of color, accounting for 80% of the population. Their cultural and social factors have been largely overlooked in the development of LLMs [67, 68]. Additionally, considering their lower public health literacy and education levels and socioeconomic status, it is essential to incorporate natural language interactions that align with their social and cultural languages. This project aims to bridge this gap and facilitate meaningful interactions between underrepresented communities and the knowledge graph.

To collect valuable input from the community, including health workers, social workers, religious leaders, and social network influencers, we will employ the CBPR process described earlier. This approach will enable us to gather their questions, queries, and concerns. Subsequently, we will preprocess this data to ensure compatibility with large language models (LLMs). The preprocessing step will involve encoding entities, relationships, and attributes into embeddings or other representations that the language model can comprehend. Following the preprocessing stage, we will proceed to fine-tune the large language model using the curated data. This process will entail training the model on a combination of general language understanding tasks and specific tasks that utilize the knowledge graph, such as entity linking, relation extraction, or question answering.

The ultimate objective of this step is to develop an interface that empowers end users to interact and query the knowledge graph database using natural languages, even in the presence of cultural and social constraints. By incorporating the insights, questions, and concerns of the community participants and fine-tuning the language model accordingly, we aim to create a user-friendly interface that accommodates diverse linguistic and sociocultural contexts.

2. Partnerships And Engagement

This project has effectively identified and engaged a diverse range of stakeholders spanning government, industry, academia, non-profits, and citizen service groups. This comprehensive network encompasses major organizations such as the Alabama Department of Public Health, Office of EMS, the National Association of Social Workers-Alabama Chapter, and the National Alliance on Mental Illness-Alabama Chapter. We have also secured the backing of Alabama State Senator Mr. Bobby Singleton, the representative for District 24, a crucial part of the Alabama Black Belt region. Further, our engagement strategy prominently features the active involvement of citizen service groups. A key example of this is the Alabama Health & Wellness Education Center, which provides vital services across numerous counties within the Alabama Black Belt area. In the ongoing journey of this project, we remain committed to expanding our network of participating constituents and intend to capitalize on the recent efforts made by the University of Alabama. We will continue to explore new opportunities for collaboration to ensure a holistic and inclusive approach.

Uniqueness of this Project and Community Engagement Under the expert guidance of Principal Investigators (PIs) Gong, Lee, and Geyer, this project emerges as one of the first initiatives designed to develop a database primed for artificial intelligence (AI) applications for Rural Resilience within Alabama State. Their collaborative work is directed towards leveraging AI technology to enhance data analysis and decision-making capabilities for the betterment of Alabama. PI Gong's progressive research vision has been particularly noteworthy. Recognizing the immense potential and transformative power of his work, benefactors recently granted him a generous donation of $2 million from a local resident. These funds will be dedicated towards the establishment of an Artificial Intelligence Center led by the University of Alabama. The center promises to become a hub of innovation, fostering advanced research and development in AI.

The AI Center is strategically positioned to act as a catalyst, further extending PI Gong's ability to involve and interact with a diverse range of stakeholders and constituents. This platform will enable an increase in participation from government agencies, industry representatives, academia, non-profits, and citizen groups, thereby fostering a truly collaborative approach. Moreover, the center will significantly contribute to the real-world use cases that this project targets. With AI technology's potential to solve complex problems and generate insights, the center is expected to enhance the effectiveness and impact of the targeted use cases.

Consultation and Collaboration Leveraging the resources of the in-development AI center, our project will commence with a kick-off meeting. This introductory gathering will serve to acquaint all stakeholders with one another, along with establishing a clear understanding of the project's objectives and timeline. We will maintain regular weekly meetings as a platform for all participants to share progress updates, tackle challenges in a collaborative manner, brainstorm solutions, and make necessary adjustments to the project plan. In-depth exploration of key themes, such as the government's decision-making process in relation to public health and justice, will be facilitated through round-table discussions and specialized consultation sessions. These forums aim to foster a deeper understanding and encourage open dialogue on pivotal aspects of the project. Moreover, to ensure that stakeholders have the opportunity to contribute their thoughts and feedback beyond the confines of formal meetings and workshops, we will introduce various feedback mechanisms. These will include options like surveys and suggestion boxes, encouraging a constant flow of communication and promoting an inclusive, participatory environment.

Data Integration and Validation In our endeavor, we will collaborate closely with data providers such as federal agencies (including CDC, NASA, NOAA), data repositories (like ICPSR), and regional agencies (including ADPH). Our collaboration will aim to facilitate the data science workflow, which encompasses activities like data inventory, data curation, schema mapping, entity resolution, data integration and quality assessment, and data reconciliation. It's essential to acknowledge that the transfer and generative learning processes necessitate legal frameworks provided by government agencies. Additionally, they rely on the technical infrastructure and industry-specific insights offered by our industry partners. Non-profit organizations and citizen groups play a key role by providing access to ground-truth data and understanding of the local social and cultural context. They also help evaluate the practical applicability of the knowledge graph in real-world scenarios.

Governance and Oversight We will form a data governance board with diverse stakeholder representatives for oversight, identified via the Community-Based Participatory Research process. This board will establish policies and procedures for key areas like data management, privacy, security, and ethics, in compliance with regulations and best practices. The University's Institutional Review Board will enforce these policies through regular compliance checks. We'll also establish mechanisms for conflict resolution and consistently monitor project progress. Transparent reporting to the board and stakeholders on milestones, budget, compliance, and potential risks will keep all parties informed.

Training and Capacity Building In the Alabama State context, training and capacity-building are vital for this project, fostering partnerships and community engagement. We'll create accessible training resources like manuals, guides, and tutorials using knowledge graphs and other resulting knowledge and materials during the R&D processes. Regular training sessions will be conducted in diverse formats according to stakeholder preferences. We'll also enhance the capabilities of entities like ADPH, NASW, and NAMI, aiding them to improve their data infrastructure and create their own training programs. This effort will be reinforced by the AI center currently in development, aiming to empower participants and ensure successful project execution.

3. Collaboration Plan

As shown in Figure 10, our project team understands the immense value of collaboration, particularly when working with large-scale knowledge graph systems. Therefore, we express our sincere eagerness to collaborate with existing initiatives that have demonstrated significant strides in related areas. These include the NIH-supported Knowledge Network efforts, the NIJ's projects focusing on nonfatal firearm injuries and mining criminal justice insights from sources like the National Archive of Criminal Justice Data (NACJD). We are also keen to collaborate with NOAA's environmental data projects related to climate and the Blue Economy, and with the USGS's efforts around topographic data integration and adherence to FAIR principles. Furthermore, we look forward to potential collaborations with the Department of Transportation on Transportation Equity, as well as projects that focus on Climate Change, Disaster Management, and Energy Systems. Lastly, we value the importance of accurate Health Communications and are excited about potential partnerships in this area. We believe these collaborations will allow us to build a richer, more comprehensive knowledge graph, which can better serve our collective goals and the communities we aim to support.

Drawing on the resources provided by the AI center currently in development, we are planning to facilitate interdisciplinary meetings to tackle project challenges, brainstorm solutions, and identify potential avenues for collaboration. We are also eager to partake in joint research initiatives that resonate with the goals of several projects, with the objective of sharing resources, broadening our expertise base, and contributing to high-impact publications. We enthusiastically look forward to contributing to the development and usage of collaborative tools and platforms. This includes cloud-based platforms that allow for efficient data storage, sharing, and querying, project management tools, and

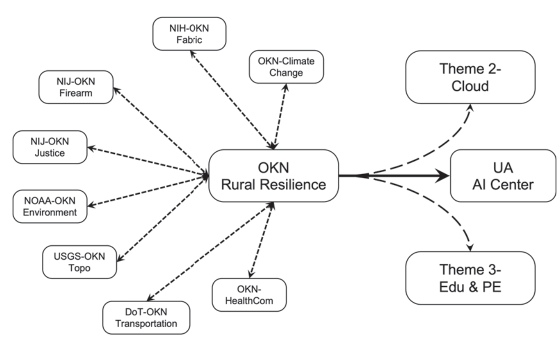

Figure 10: Collaboration diagram with other themes.

collaborative software, like LLM for the purpose of querying knowledge graphs. Moreover, we plan to co-author policy briefs, utilizing insights from multiple projects. We believe this powerful approach will help shape policy-making related to public health, justice, and environmental sectors among others. Finally, in alignment with Theme 3, we are committed to spearheading public engagement activities, such as community forums, public lectures, and educational events. We trust that these activities will amplify our project's visibility, stimulate public awareness and support, and provide channels for community participation and engagement.

4. Deliverables And Management Plan

Definitions of the alpha and beta level deliverables The alpha-level deliverables serve as an initial preview of the envisioned final product for our stakeholders. They encompass elements such as a detailed project roadmap, data models and schemas, an assessment of data quality, a plan for data integration, a prototype of a knowledge graph sample, a preliminary query interface, initial project documentation, mechanisms for user feedback, along with protocols for governance and security. Progressing from the alpha stage, the beta-level deliverables exhibit a more sophisticated and refined version of the project. They encompass the final versions of the data models and schemas, an established mapping of entities with other Knowledge Graph (KG) projects, an advanced query interface, the capability to integrate with external systems, feedback from user testing, finely-tuned governance and security protocols, and established performance metrics. The ultimate deployment of the proposed knowledge graphs is set to undergo iterative evaluation, as detailed in Section 5.

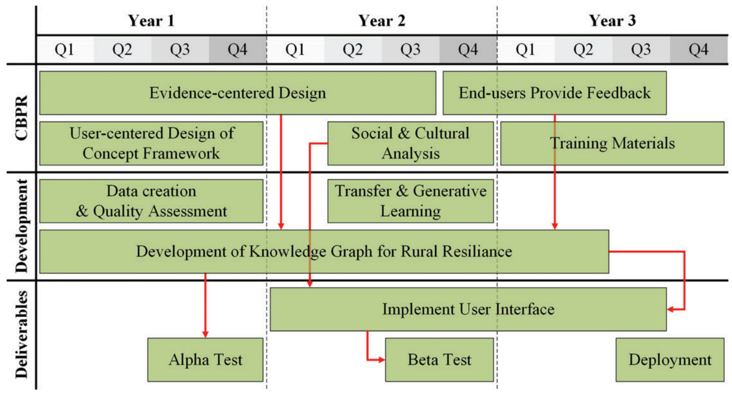

Figure 11: Project Timeline.

Project Timeline Figure 11 depicts the project's timeline, encompassing three key components: 1) Community-Based Participatory Research (CBPR) to facilitate engagement with end users, foster community partnerships, and incorporate social and cultural feedback. 2) The development of the Knowledge Graph (KG), which entails stages like data access, curation, quality assessment, transfer and generative learning, as well as the creation and refinement of user interfaces. 3) The creation and deployment of deliverables, which places emphasis on conducting standing meetings and workshops involving data providers, end users, and other relevant stakeholders.

Anticipating potential challenges within this timeline, we identify two primary areas of concern: 1) The establishment of protocols for Institutional Review Board (IRB), data access and governance, and security could exceed our estimated timeframe. Nevertheless, we plan to draw from our prior experiences, as demonstrated in our preliminary work [22], to navigate potential obstacles. 2) The recruitment of local community participants may require more time than anticipated. In response to this, we intend to collaborate closely with community health and social workers, as well as Alabama State Senator, to establish broader connections. By identifying these potential challenges in advance, we aim to develop effective strategies to ensure the project's timely progression.

Management Plan Our team is composed of specialists from various fields, providing us with a diverse range of skills and knowledge. The team includes computer scientists Gong and Hong, both proficient in data science and database software development. Social scientist Lee offers expertise in Community-Based Participatory Research (CBPR). Medical research scientists Geyer and Hudnall provide a deep understanding of health aspects within the context of rural resilience. Criminology scientist Dolliver lends his knowledge on justice aspects tied to rural resilience, and geoinformatics scientist Li contributes their skills in geoinformatics. Additionally, we have industrial partners, such as Cox, who bring a wealth of experience in health information technology. This diverse blend of backgrounds, experiences, and areas of expertise strengthens our project and enables a comprehensive, multi-dimensional approach.

Gong, as the Collaboration Lead, will be overseeing the project. Working under Gong's supervision, postdoctoral researcher Dr. Chen Wang will assume the role of Integration Lead. Dr. Wang is expected to complete her graduation of Computer Science in August, aligning perfectly with the project's prospective start date in October, subject to the project's funding approval.

As depicted in Figure 11, Gong will spearhead tasks related to evidence-centered design, transfer and generative learning, and the developmental process of the Knowledge Graph (KG) for rural resilience. Hong will direct the creation of training materials, manage data access and curation, and oversee the design and refinement of user interfaces. Lee's focus will be on the user-centered design of the conceptual framework and social and cultural analysis within the CBPR process. On the other hand, Geyer will work to gather feedback from end-users and provide valuable input to the KG development and user interface enhancement. Hudnall will shoulder the tasks of data access, curation, and KG development in the health domain, while Dolliver will handle similar tasks but with a focus on the justice aspect. Li, specializing in geoinformatics, will manage data access, curation, and KG development for this specific aspect. Representing the industry, Cox will be in charge of the project's deliverables, which include the alpha, beta, and final deployment stages.

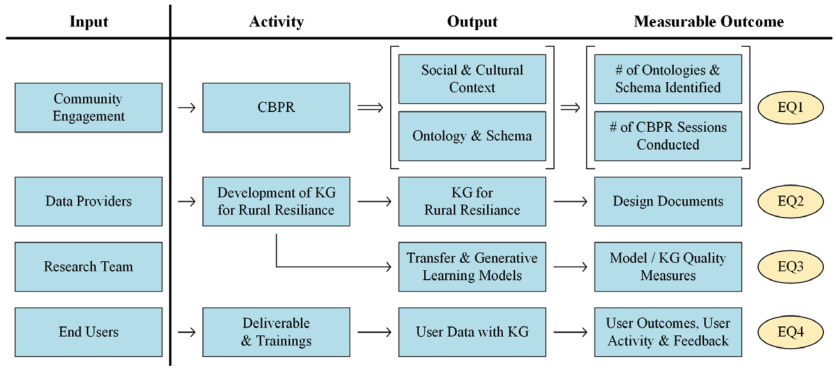

Figure 12: Logic model to support the Evaluation Plan.

5. Evaluation Plan

Our assessment approach consists of quarterly examinations conducted by our autonomous data governance board. This body will scrutinize the data, documentation, and project-generated artifacts. In addition, they will accumulate insights by conducting interviews with the PI and Co-PIs.

The assessment consultants will deliver feedback regarding how well the project adheres to its initial plan. Depending on the needs of the project, consultants will also lend their proficiency in data governance research, providing guidance and recommendations based on their prior experience in resolving issues in comparable research projects. As illustrated in Figure 12, the logic model outlines the activities, output, outcomes, along with corresponding evaluation questions (EQ).

- **EQ1**: (Year 1 and Year 2) Were the activities pertaining to Community-Based Participatory Research (CBPR) executed as per the plan? Did these activities yield adequate data or insights to inform the design of the knowledge graph, thus assisting the data-driven needs of end users' decision-making processes?
- **EQ2**: (Year 1) Were the planned design activities, such as workshops and meetings, carried out effectively? What was the participation level of community practitioners and end users at these workshops? Were these design activities successful in generating adequate data or insights to facilitate the development of the knowledge graph for rural resilience?

- **EQ3**: (Year 2) How closely did the data synthesis and quality assurance activities follow the planned schedule and were they executed as intended? How would one rate the quality of the knowledge graph in terms of its precision in assessing data quality, schema coherence, and mapping consistency? How effective was the knowledge graph in validating its reasoning against the KG entities as provided by experts, end users, and/or other KG projects?
- **EQ4**: (Year 2 and Year 3) Did the pilot implementation with beta-level deliverables proceed as anticipated? What was the count of trainings, end users, and practitioners involved in the pilot projects? How engaged were the users as measured by the duration spent interacting with the knowledge graphs and the distinct activities undertaken in relation to the beta-level deliverables? What feedback and recommendations for enhancement did users provide, as gathered from surveys and interviews?

Upon receiving the grant award notification, the data governance board will collaborate with the PI and CoPIs to fine-tune the evaluation queries, finalize the specific tasks required for evaluation, and review the schedule for the evaluation. The PI and CoPIs will assist in the evaluation process by periodically supplying the necessary research documentation or summaries, along with data derived from all research activities and pilot studies, including design documents, pilot study designs, student outcomes, and activity log data. In the project's concluding year, the evaluation consultants will compile a comprehensive summative evaluation. This document will record the degree to which the research project accomplished its objectives, address any unintended outcomes, and will include recommendations for enhancing future projects.

6. Scalability, Extensibility, And Sustainability Plan

AI for Rural Resilience This project represents one of the pioneering initiatives emerging from the currently under-construction AI research and education center at the University of Alabama. One of the primary goals of this center is to devise bespoke AI solutions, encompassing tools, systems, and packages that are specifically designed to cater to the needs of communities with low socio-economic status, a common occurrence in Alabama. The purpose is to acknowledge and address the unique challenges and necessities of these communities, equipping them with AI technology that has the potential to uplift their living conditions. The United Health Foundation (UHF) reports that numerous Southern States—including Louisiana, Mississippi, Alabama, Georgia, South Carolina, Tennessee, Arkansas, and Oklahoma—repeatedly score low in the U.S. for economic status, education, and health and wellness indicators. To address this, the project's scalability will augment the proposed knowledge graph system's ability to manage an escalating volume of data and user interactions across all these Southern U.S. states.

Sustainability Operational sustainability will be linked to the AI center that's currently under development, sharing staff resources and operational roles with this project. This collaboration ensures the long-term maintenance of the knowledge graph system, its cost-efficiency, and the provision of adequate staffing to keep the system running optimally. On the other hand, the task of keeping the knowledge graph data current and relevant will be bolstered by other projects, notably the use cases. Promoting rural resilience to public health and environmental crises—such as vaccine hesitancy, the opioid crisis, drought, and floods—is a vital research and community engagement focus supported by various funding agencies, including NOAA, USGS, and NIH. Significantly, the National Water Center resides on the University of Alabama campus, and the Cooperative Institute for Research to Operations in Hydrology (CIROH), a national consortium formed in partnership between NOAA and The University of Alabama, is dedicated to improving water prediction (including streamflow forecasts, extreme events like floods and droughts, and water quality) and enhancing community resilience to water-related challenges.

7. Data/AI Ethics Standards And Guidance

Our team, targeting government-related constituents like ADPH with the proposed OKN, will follow the principles from Executive Order 13960 for developing trustworthy AI. This commitment involves implementing CBPR, transfer learning, and fine-tuning LLMs. We will hold roundtable discussions under the guidance of the University administration and ethical workforce to uphold national values and legality. We will establish risk-benefit analysis methods for AI project phases and work with the University's Cybersecurity and Social work teams to address security concerns. Our development process will prioritize transparency, ensuring understanding among experts and users. Our commitment extends to ensuring the proposed OKN project aligns with these principles and is utilized in a way that suits the intended purposes of each Knowledge Graph (KG) application, especially in promoting rural resilience. We will ensure that KG design, development, acquisition, usage, and relevant inputs and outputs from specific KG applications are thoroughly documented and traceable to the best of our ability. These principles are consistent across various OKN projects and themes. Our team is actively interested in contributing to initiatives to alleviate ethical concerns and develop standards and guidelines. However, our project focusing on rural resilience confronts a unique and substantial issue - the Digital Divide [69–71]. Our team, through previous work, has explored several approaches to address this crucial challenge.

Digital Divide Our team brings to the table a wealth of experience in partnering with local communities, and we possess a profound understanding of the potential challenges that may arise alongside the advantages conferred by advanced digital technologies, especially in rural areas characterized by lower socioeconomic status. In these regions, a substantial digital divide exists, largely due to the unavailability and inaccessibility of affordable broadband internet [72, 73]. Today's digital health solutions typically presuppose user access to broadband internet, but the reality significantly diverges from these assumptions. In states like Alabama and Mississippi, nearly half of the rural inhabitants lack access to such services, underlining a stark disconnect between expectation and actuality [72, 73]. Given these circumstances, our project will be grounded in community-based participatory research activities. During our pilot study, we leveraged remote sensing data to identify all residents in Alabama State and collaborated with local community health and social workers to reach out to rural residents via postal mails. This strategy ensured the engagement of those affected by the digital divide, in line with our commitment to "Leave No One Behind". These postal engagement of the community members will not only solve the digital divide issue, but also reduce the over-reliance on technology, which might overlook the importance of human judgment and local knowledge in health and justice matters.

8. Broader Impacts

A comprehensive and accessible knowledge graph will act as a navigational tool, enabling effective resource allocation, policy development, and partnership establishment within rural communities. It will notably empower underrepresented groups, providing them a platform to assert their perspectives and influence decisions of substantial significance. To ensure the widespread dissemination of the knowledge products derived from this project, our outreach initiatives will target a diverse group of stakeholders, including interdisciplinary researchers, educators, students, school administrators, and industry partners across multiple practice domains. This project will generate a robust collection of learning science methodologies, computational models, design frameworks, software artifacts, and empirical data, thus providing a rich resource for the interdisciplinary educational research community to further advance teaching and learning in knowledge graphs.

References Cited

[1] L. H. Downey, "Rural populations and health: determinants, disparities, and solutions," *Preventing Chronic Disease*, vol. 10, 2013.

[2] R. S. Cox and M. Hamlen, "Community disaster resilience and the rural resilience index," *American Behavioral Scientist*, vol. 59, no. 2, pp. 220–237, 2015.

[3] G. Kim, W. Kang, and J. Lee, "Knowledge structures and components of rural resilience in the 2010s: Conceptual development and implications," *Sustainability*, vol. 12, no. 22, p. 9769, 2020.

[4] A. R. Fleming, N. A. Ysasi, D. A. Harley, and M. L. Bishop, "Resilience and strengths of rural communities," *Disability and vocational rehabilitation in rural settings: Challenges to service delivery*, pp. 117–136, 2018.

[5] B. O'sullivan, J. Leader, D. Couch, and J. Purnell, "Rural pandemic preparedness: the risk, resilience and response required of primary healthcare," *Risk Management and Healthcare Policy*, pp. 1187–1194, 2020.

[6] S. Freeman, R. Weaver, and S. Freeman, "Rural resilience through covid-19," *Innovation in Aging*, vol. 5, no. Supplement_1, pp. 417–417, 2021.

[7] H. Love and M. Powe, "The necessary foundations for rural resilience: A flexible, accessible, and healthy built environment," 2020.

[8] D. Hartley, "Rural health disparities, population health, and rural culture," *American journal of public health*, vol. 94, no. 10, pp. 1675–1678, 2004.

[9] W. Caldwell, E. Ferguson, E. Lapierre-Fortin, J. Ball, S. Reid, P. Kraehling, E. Marr, J. Devlin, C. White, T. McQuail *et al.*, *Planning for rural resilience: Coping with climate change and energy futures.* Univ. of Manitoba Press, 2015.

[10] J. D. Freeman, S. Kadiyala, J. F. Bell, and D. P. Martin, "The causal effect of health insurance on utilization and outcomes in adults: a systematic review of us studies," *Medical care*, pp. 1023–1032, 2008.

[11] D. Anderson, "Social justice in health: the patient-centered medical home and health disparities," Ph.D. dissertation, Rutgers University-Graduate School-Newark, 2021.

[12] H. Ibrahim, X. Liu, N. Zariffa, A. D. Morris, and A. K. Denniston, "Health data poverty: an assailable barrier to equitable digital health care," *The Lancet Digital Health*, vol. 3, no. 4, pp. e260–e265, 2021.

[13] R. Marchesi, N. Micheletti, G. Jurman, and V. Osmani, "Mitigating health data poverty: Generative approaches versus resampling for time-series clinical data," *arXiv preprint arXiv:2210.13958*, 2022.

[14] M. S. Al-Zaman, "A review of literature on information poverty," *GeoJournal*, pp. 1–14, 2023.

[15] A. H. Iskandar, A. F. Suja'ie, and I. Agusta, "Redesigning consolidated data for handling extreme poverty in rural areas based on sdgs desa," *Journal of Community Positive Practices*, no. 1, pp. 96–115, 2023.

[16] M. Cafarella, M. Anderson, I. Beltagy, A. Cattan, S. Chasins, I. Dagan, D. Downey, O. Etzioni, S. Feldman, T. Gao *et al.*, "Infrastructure for rapid open knowledge network development," *AI Magazine*, vol. 43, no. 1, pp. 59–68, 2022.

[17] X. Ma, "Knowledge graph construction and application in geosciences: A review," *Computers & Geosciences*, p. 105082, 2022.

[18] C. Galeone and R. Bonzi, "Open data: applications in social care and health," *Statistica Applicata-Italian Journal of Applied Statistics*, no. 2, pp. 125–139, 2020.

[19] K. Janowicz, P. Hitzler, W. Li, D. Rehberger, M. Schildhauer, R. Zhu, C. Shimizu, C. Fisher, L. Cai, G. Mai *et al.*, "Know, know where, knowwheregraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence," *AI Magazine*, vol. 43, no. 1, pp. 30–39, 2022.

[20] W. Maurer, "Building resilient communities: Lessons learned from long term recovery," *MUP Capstone*, 2012.

[21] E. Chen, C. Leos, S. D. Kowitt, and K. E. Moracco, "Enhancing community-based participatory research through human-centered design strategies," *Health promotion practice*, vol. 21, no. 1, pp. 37–48, 2020.

[22] X. Wu, S. Cao, H. Y. Lee, and J. Gong, "Let every voice be heard: Developing a cost-effective community sampling frame in rural alabama to combat covid-19," in *2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2022, pp. 174–175.

[23] H. Estiri, S. Vasey, and S. N. Murphy, "Generative transfer learning for measuring plausibility of ehr diagnosis records," *Journal of the American Medical Informatics Association*, vol. 28, no. 3, pp. 559–568, 2021.

[24] S. Yao, Q. Kang, M. Zhou, M. J. Rawa, and A. Abusorrah, "A survey of transfer learning for machinery diagnostics and prognostics," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 2871–2922, 2023.

[25] P. Mujuru, B. Jean-Francois, and E. J. Pérez-Stable, "National institute on minority health and health disparities specialized centers of excellence on minority health and health disparities," *American Journal of Preventive Medicine*, vol. 63, no. 1, p. S6, 2022.

[26] N. Pourat, X. Chen, C. Lee, W. Zhou, M. Daniel, H. Hoang, and A. Sripipatana, "Expanding capacity in hrsa-funded health centers to address the national opioid and substance abuse crisis," in *APHA's 2018 Annual Meeting & Expo (Nov. 10-Nov. 14)*. APHA, 2018.

[27] B. Bondaruk, S. A. Roberts, and C. Robertson, "Assessing the state of the art in discrete global grid systems: Ogc criteria and present functionality," *Geomatica*, vol. 74, no. 1, pp. 9–30, 2020.

[28] L. Mikesell, E. Bromley, and D. Khodyakov, "Ethical community-engaged research: A literature review," *American journal of public health*, vol. 103, no. 12, pp. e7–e14, 2013.

[29] R. W. Cooksey, *Illustrating statistical procedures: For business, behavioural & social science research*. Tilde University Press, 2007.

[30] R. T. Warne, *Statistics for the social sciences: a general linear model approach*. Cambridge University Press, 2020.

[31] R. Cattell, *The scientific use of factor analysis in behavioral and life sciences*. Springer Science & Business Media, 2012.

[32] R. Angles, H. Thakkar, and D. Tomaszuk, "Rdf and property graphs interoperability: Status and issues." *AMW*, vol. 2369, 2019.

[33] C. Hondros, "Standardizing legal content with owl and rdf," in *Linking Enterprise Data*. Springer, 2010, pp. 221–240.

[34] R. Abraham, J. Schneider, and J. Vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *International journal of information management*, vol. 49, pp. 424–438, 2019.

[35] A. Reisser and T. Priebe, "Utilizing semantic web technologies for efficient data lineage and impact analyses in data warehouse environments," in *2009 20th International Workshop on Database and Expert Systems Application*. IEEE, 2009, pp. 59–63.

[36] E. D. Foster and A. Deardorff, "Open science framework (osf)," *Journal of the Medical Library Association: JMLA*, vol. 105, no. 2, p. 203, 2017.

[37] L. Gamm, S. Stone, and S. Pittman, "Mental health and mental disorders—a rural challenge: A literature review," *Rural healthy people*, vol. 2, no. 1, pp. 97–114, 2010.

[38] B. Showers, D. Dicken, J. S. Smith, and A. Hemlepp, "Medication for opioid use disorder in rural america: A review of the literature." *Journal of Rural Mental Health*, vol. 45, no. 3, p. 184, 2021.

[39] Y. Sano and S. Mammen, "Mitigating the impact of the coronavirus pandemic on rural low-income families," *Journal of Family and Economic Issues*, vol. 43, no. 2, pp. 227–238, 2022.

[40] C. E. Peterson, A. Silva, H. K. Holt, A. Balanean, A. H. Goben, and J. A. Dykens, "Barriers and facilitators to hpv vaccine uptake among us rural populations: a scoping review," *Cancer Causes & Control*, vol. 31, pp. 801–814, 2020.

[41] P. Lal, J. R. Alavalapati, and E. D. Mercer, "Socio-economic impacts of climate change on rural united states," *Mitigation and Adaptation Strategies for Global Change*, vol. 16, pp. 819–844, 2011.

[42] K. S. Gutierrez and C. E. LePrevost, "Climate justice in rural southeastern united states: a review of climate change impacts and effects on human health," *International journal of environmental research and public health*, vol. 13, no. 2, p. 189, 2016.

[43] T. M. Sullivan, R. J. Limaye, V. Mitchell, M. D'Adamo, and Z. Baquet, "Leveraging the power of knowledge management to transform global health and development," *Global Health: Science and Practice*, vol. 3, no. 2, pp. 150–162, 2015.

[44] S. Dolley, "Big data's role in precision public health," *Frontiers in public health*, vol. 6, p. 68, 2018.

[45] R. Bhatia and A. Wernham, "Integrating human health into environmental impact assessment: an unrealized opportunity for environmental health and justice," *Environmental Health Perspectives*, vol. 116, no. 8, pp. 991–1000, 2008.

[46] S. McCormick, M. Peterson-Badali, and T. A. Skilling, "Mental health and justice system involvement: A conceptual analysis of the literature." *Psychology, Public Policy, and Law*, vol. 21, no. 2, p. 213, 2015.

[47] H. Y. Lee, J. Eyer, D. Lee, C. R. Won, M. Hudnall, and D. S. Cain, "The opioid crisis in alabama: Actionable recommendations for prevention, treatment, and recovery in rural communities," *Best Practices in Mental Health*, vol. 16, no. 1, pp. 1–11, 2020.

[48] H. Budhwani, A. C. Maragh-Bass, E. E. Tolley, M. L. G. Comello, M. C. Stoner, M. A. Larsen, D. Brambilla, K. E. Muessig, A. Pettifor, C. L. Bond *et al.*, "Tough talks covid-19 digital health intervention for vaccine hesitancy among black young adults: Protocol for a hybrid type 1 effectiveness implementation randomized controlled trial," *JMIR Research Protocols*, vol. 12, no. 1, p. e41240, 2023.

[49] H. Budhwani, T. Maycock, W. Murrell, and T. Simpson, "Covid-19 vaccine sentiments among african american or black adolescents in rural alabama," *Journal of Adolescent Health*, vol. 69, no. 6, pp. 1041–1043, 2021.

[50] M. Nabavi, E. Kidd, C. Simma, A. Sullivan, D. Coombs, and J. Waterbor, "You are not alone: A depression and suicide prevention public service announcement for alabama," in *APHA's 2020 VIRTUAL Annual Meeting and Expo (Oct. 24-28)*. APHA, 2020.

[51] K. M. Nowotny and A. Kuptsevych-Timmer, "Health and justice: framing incarceration as a social determinant of health for black men in the united states," *Sociology Compass*, vol. 12, no. 3, p. e12566, 2018.

[52] M. L. Rosino and M. W. Hughey, "The war on drugs, racial meanings, and structural racism: A holistic and reproductive approach," *American Journal of Economics and Sociology*, vol. 77, no. 3-4, pp. 849–892, 2018.

[53] J. H. Boman and O. Gallupe, "Has covid-19 changed crime? crime rates in the united states during the pandemic," *American journal of criminal justice*, vol. 45, pp. 537–545, 2020.

[54] H. J. Pandit, D. O'Sullivan, and D. Lewis, "Using ontology design patterns to define shacl shapes." in *WOP@ ISWC*, 2018, pp. 67–71.

[55] C. Debruyne and K. McGlinn, "Reusable shacl constraint components for validating geospatial linked data," in *4th International Workshop on Geospatial Linked Data (GeoLD) Co-located with the 18th Extended Semantic Web Conference (ESWC 2021)*. CEUR, 2021.

[56] P. S. Gupta, A. M. Mohareb, C. Valdes, C. Price, M. Jollife, C. Regis, N. Munshi, E. Taborda, M. Lautenschlager, A. Fox *et al.*, "Expanding covid-19 vaccine access to underserved populations through implementation of mobile vaccination units," *Preventive Medicine*, vol. 163, p. 107226, 2022.

[57] R. Agarwal, M. Bjarnadottir, J. Clark, L. Rhue, and G. G. Gao, "Data science for health equity: Perils and promise of ai/ml in healthcare," *ML in Healthcare (April 24, 2023)*, 2023.

[58] J. Engelmann and S. Lessmann, "Conditional wasserstein gan-based oversampling of tabular data for imbalanced learning," *Expert Systems with Applications*, vol. 174, p. 114582, 2021.

[59] G. Ghosheh, J. Li, and T. Zhu, "A review of generative adversarial networks for electronic health records: applications, evaluation measures and data sources," *arXiv preprint arXiv:2203.07018*, 2022.

[60] H. Wang, G. Tao, J. Ma, S. Jia, L. Chi, H. Yang, Z. Zhao, and J. Tao, "Predicting the epidemics trend of covid-19 using epidemiological-based generative adversarial networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 2, pp. 276–288, 2022.

[61] H. Bao, X. Zhou, Y. Zhang, Y. Li, and Y. Xie, "Covid-gan: Estimating human mobility responses to covid-19 pandemic through spatio-temporal conditional generative adversarial networks," in *Proceedings of the 28th international conference on advances in geographic information systems*, 2020, pp. 273–282.

[62] A. Bej, U. Maulik, and A. Sarkar, "Time-series prediction for the epidemic trends of covid-19 using conditional generative adversarial networks regression on country-wise case studies," *SN Computer Science*, vol. 3, no. 5, p. 352, 2022.

[63] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, and N. Zhang, "Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities," *arXiv preprint arXiv:2305.13168*, 2023.

[64] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying large language models and knowledge graphs: A roadmap," *arXiv preprint arXiv:2306.08302*, 2023.

[65] P. Christ, "Generation of meaningful sql-query exercises using large language models and knowledge graphs."

[66] H. Tran, L. Phan, J. Anibal, B. T. Nguyen, and T.-S. Nguyen, "Spbert: an efficient pre-training bert on sparql queries for question answering over knowledge graphs," in *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part I 28*. Springer, 2021, pp. 512–523.

[67] S. A. Khowaja, P. Khuwaja, and K. Dev, "Chatgpt needs spade (sustainability, privacy, digital divide, and ethics) evaluation: A review," *arXiv preprint arXiv:2305.03123*, 2023.

[68] I. Solaiman and C. Dennison, "Process for adapting language models to society (palms) with values-targeted datasets," *Advances in Neural Information Processing Systems*, vol. 34, pp. 5861–5873, 2021.

[69] B. Kelley and L. Sisneros, "Broadband access and the digital divides. policy brief." *Education Commission of the States*, 2020.

[70] H. Whiteside and B. Beavers, "Social justice and the digital divide: A narrative inquiry of how digital inequities contribute to the underrepresentation of low-income african americans in stem careers," in *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 2022, pp. 294–298.

[71] C. N. Eruchalu, M. S. Pichardo, M. Bharadwaj, C. B. Rodriguez, J. A. Rodriguez, R. W. Bergmark, D. W. Bates, and G. Ortega, "The expanding digital divide: digital health access inequities during the covid-19 pandemic in new york city," *Journal of Urban Health*, vol. 98, pp. 183–186, 2021.

[72] E. Y. Choi, S. Kanthawala, Y. S. Kim, and H. Y. Lee, "Urban/rural digital divide exists in older adults: Does it vary by racial/ethnic groups?" *Journal of Applied Gerontology*, vol. 41, no. 5, pp. 1348–1356, 2022.

[73] H. Y. Lee, S. Kanthawala, E. Y. Choi, and Y. S. Kim, "Rural and non-rural digital divide persists in older adults: Internet access, usage, and attitudes toward technology." *Gerontechnology*, vol. 20, no. 2, 2021.