|  |  |
| --- | --- |
| **Principal Investigator** | |
| PI’s/PIs’ Name: | Phone: |
| E-mail: | Department:  College: |
| Primary Campus: ☐ Main Campus ☐ Other: | |
| ☐ Faculty ☐ Doctoral ☐ Specialist ☐ Masters ☐ Undergraduate ☐ Other: | |
| **Co-Investigator(s)** | |
| Co-I’s Name(s):  *(By each name indicate: F(Faculty), D(Doctoral), S(Specialist), M(Masters), U(Undergraduate), O(Other))* | E-mail: |

|  |  |
| --- | --- |
| **PROJECT INFORMATION** | |
| Title: Generating Textual Descriptions for Grid Heatmap using Encoder Decoder LSTM with Attention-based Mechanism | |
| Number of Subjects (Maximum): 1 | |
| **COMPLIANCE INFORMATION** | |
| Do you or any investigator on this project have a financial interest in the subjects, study outcome, or project sponsor? (A disclosed conflict of interest will not preclude approval. An undisclosed conflict of interest will result in disciplinary action.). ☐ Yes ☐ No | |
| ☐ Self-funded/non-funded  ☐ Internally funded | ☐ External Funding (*You are responsible for duplicate or additional approval submissions required by funders.)*  Funding Source: ☐ Government ☐ Scholarship ☐ Institution ☐ Contract ☐ Sponsorship  Funding Agency:  Status: ☐ Pending Submission ☐ Submitted ☐ Funded    Grant Title: ☐ Same as above **OR** Enter here:  ☐ Funding application scope of work attached |

|  |
| --- |
| **CERTIFICATIONS** |
| I hereby certify that all information above is true and correct to the best of my knowledge and belief. As I understand the necessity of Ethics Clerance Approval, I will submit a UREC application through the research ethics process. I further certify that all personnel listed on this application have reviewed and approved of their described engagement in this research.  I understand that I am responsible for the conduct of all researchers engaged in this project.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Signature of Primary Investigator Date  If the PI is a student:  By signing this cover page, I acknowledge that I have reviewed and approved this narrative for accuracy. I further acknowledge that I approve of the ethical basis for the study. I understand that my student may need to apply to the UREC for approval.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Typed/Printed Name  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | Signature  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | | Date  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | | |  |  |  |  | |  |  | |

|  |
| --- |
| 1. **COMMON RULE DETERMINATION** |
| 1. Does this research involve any of the following?   ☐ Food ☐ Animals ☐ Literary works  ☐ New drug or drug use ☐ Bodies of Water ☐ Media-related  ☐ Investigational medical devices ☐ Micro-organisms ☐ If not in the following, state: \_\_\_\_\_\_\_\_\_\_  ☐ Plants ☐ Public documents |

|  |
| --- |
| 1. **RESEARCH DETERMINATION** |
| 1. Is the data being studied in this project obtained in a systematic manner?   Yes  No |
| 1. Is the intent of this data collection to contribute to generalizable knowledge? (E.g., applicable to situations beyond the study or designed to draw general conclusions, inform policy, or generalize findings beyond a single individual or an internal program)   ☐ Yes  ☐ No. Explain: |

|  |
| --- |
| 1. **HUMAN SUBJECTS DETERMINATION** |
| 1. Does the project/activity involve obtaining information about human participant through intervention *(e.g., physical procedures, manipulations of the subject or subject’s environment)* or interaction *(e.g., communication with or data collected from individuals)*?   ☐ Yes  ☐ No. Does the project involve obtaining protected health information including (PHI) about deceased individuals or biospecimen?  ☐ Yes ☐ No |
| 1. Is the data collected by interaction with human subjects **only** seeking data that is not about the human subjects?   *(E.g., survey about business policy, practice, or characteristics without obtaining data about the characteristic or opinions of the individuals providing the data or other individuals, E.g., survey of human resource offices asking about the types of benefits offered but not the desirability or effectiveness of the benefit program.)*  ☐ Yes \*If yes, attach the research instrument (survey questionnaire, interview, data fields, etc.)  ☐ No  ☐ N/A: Not collecting data by interaction with human subjects. |
| 1. Will all the data collected be gathered from published sources? (*e.g. library books, journal articles, open websites*)   ☐ Yes  ☐ No  **Provide website link here if applicable: https://www.gapminder.org/data/** |
| 1. Will all the data collected be oral history that is designed solely to create a record of specific historical events where the interviewees are aware of the method of publication intended for their stories? “*The interview must allow the subject (narrator) to tell their story without analysis, manipulation or content editing and be collected with fully informed consent to include consent for archiving, presentation or publication, and subsequent sharing of the collected story. Oral history projects must adhere to the principles and best practices established by the Oral History Association.”*   ☐ Yes  ☐ No |
| 1. Is the sole purpose for the interview production of a news article for publication in a paper or online newspaper?   ☐ Yes  ☐ No  ☐ N/A: Not conducting interviews. |
| 1. **PROJECT DESCRIPTION** |
| 1. Explain the project below in enough detail to allow someone not familiar with your project to understand the interactions and intent. (*Briefly discuss the Background, Objectives, and Significance of the Study.)* |
| Data visualizations are commonly utilized to effectively communicate large amounts of information (Holland, 2017). Recent developments in politics, health, finance, and business analytics underscore the significance and widespread use of online data visualizations (Sharif, Zhang, Shih, Wobbrock, & Reinecke, 2022). Screen readers face significant challenges in accessing and interpreting graphical data due to limited or totally inaccessible current chart visualization. With the rapid shift towards digital literacy, especially during/post-COVID-19 pandemic, where most critical information such as trends, guidelines, and policies, is digitally served, presses the need for improvement in the BVIP accessibility scenario more than ever. Addition of textual caption for graphical content is a valuable step for the improvement of Web Content Accessibility (WCA). To achieve this, manual annotations in form of tag, hint, and alternative text are imperative in textual components. Although proven to be effective, this manual method is impractical to apply on all documents; hence making these annotations commonly lacking in the document (Farahani, Adibi, Ehsani, Hutter & Darvishy, 2023).  Some have utilized traditional CNNs in conjunction with other methods to extract relevant features for use in their template-based descriptions (Mishra et al., 2022; Belle et al., 2022). Another approach involves using Encoder-Decoder LSTM with an attention mechanism, where only pre-trained CNNs or combined with other feature extraction methods are used as the encoder, and LSTMs with varying attentions serve as the decoder for extracting and generating textual descriptions, as demonstrated by works such as those by Xu, Ba, Kiros, Cho, Courville, Salakhutdinov, Zemel & Bengio (2015), Chen, Zhang, Ko, Kim, Cohen & Rossi (2020), and Qian et al. (2021). Some studies produce textual descriptions in templated formats, limiting overall interpretation (Mishra et al., 2022; Belle et al., 2022), others employ Encoder-Decoder LSTMs to extract relevant features from general images like photographs (Xu et al., 2015), or separately from textual and graphical data in chart images such as bar charts (Chen et al., 2020; Qian et al., 2021).  This variability underscores the diverse nature of image features, yet there remains a gap in effectively extracting essential features from images, particularly in the case of grid heatmaps. Moreover, findings from Chen et al.'s (2020) study suggest that distinct features in each chart type necessitate different attention mechanisms, potentially paving the way for developing specific attention mechanisms for other chart types, such as grid heatmaps. Despite the widespread use of heatmaps in statistical visualization (Gu, 2022), there is currently a dearth of studies focusing on or recommending textual data generation for heatmaps, leaving this area largely unexplored.  The objective of this study is to develop a model that generates descriptive text for grid heatmaps using an encoder-decoder architecture with an attention-based mechanism. This model is specifically designed to improve accessibility for Blind and Visually Impaired People (BVIP), as well as users who rely on screen readers. The significance of the study lies in its potential to greatly benefit these key stakeholders by making complex data visualizations more accessible. This enhancement in accessibility will improve their interaction with information technology, empowering them to engage more effectively with data-driven insights. |
| 1. Summarize the planned activity for which you are seeking a UREC determination. (*Discuss your Methodology*) |
| The research methodology of the study is structured as true experimental design aimed at evaluating the effectiveness of the proposed model for generating textual descriptions of grid heatmaps. The researchers have chosen a posttest-only group design, which means that the performance of the proposed Encoder-Decoder Long Short-Term Memory (LSTM) model with an Attention-based Mechanism will be assessed after the experiments are conducted, allowing for a clear comparison with a baseline model— the CNN-LSTM-Attn.  To test the hypothesis if there is no significant difference in performance between the two models, the researchers will conduct statistical analyses using a McNemar’s Test on the results obtained from the experiments. This approach is crucial for determining the validity of the proposed model in generating accurate textual descriptions.  The study is designed to carefully assess the proposed model's capabilities. By systematically evaluating the model's performance, the researchers aim to contribute valuable insights into the field of data visualization accessibility. |
| B.1 Additional Literature/References |
| Obtaining approval from the University Research Ethics Center is essential as it affirms that the research study has gone through ethical research practices and is aligned with guidelines established by the institution. As discussed by Singh et al. in 2023, assistive technologies play a vital role in supporting BVIPs, but they also raise questions regarding their effectiveness. By prioritizing these ethical considerations, the research study can be of great contribution to the advancement of knowledge and improvement of the user experience, particularly the BVIPs and SRUs, by enhancing their access and engagement with data visualizations. |
| 1. How will the project be conducted? (*Discuss your Study Procedure)* |
| The research will be conducted in a several key phases, each designed to ensure the successful development and evaluation of a model. The first phase involves collection of data. The researchers will utilize the extensive resources available from Gapminder, which provides a wealth of statistical information across various domains. The data collected will be converted into 1,710 grid heatmap figures.  Once the data is gathered, the next phase focuses on feature extraction. This step is essential for preparing the heatmaps for analysis. The researchers will employ a systematic approach to identify and classify color legends, extract relevant colors, and recover color mappings from the heatmaps. This process will ensure that the model has access to the critical visual features needed to generate accurate and meaningful textual descriptions.  With the data and features in place, the researchers will move on to model development through using Encoder-Decoder LSTM model with enhanced Attention-based Mechanism. This architecture will allow the model to process grid heatmap images and produce corresponding textual descriptions. The attention mechanism will be particularly important, as it will enable the model to focus on specific features of the heatmap during the generation of each word, thereby improving the relevance and quality of the output.  Following the development of the model, the researchers will enter the training phase. During this stage, the model will learn to identify features extracted from the heatmaps with their corresponding textual descriptions. This will then be validated to assess its performance and make necessary adjustments to enhance accuracy.  After the completion of the training is the evaluation phase. The researchers will compare the performance of their proposed model against a baseline model, the CNN-LSTM-Attn, through an evaluator that will be undergoing a Turing test. Measuring the accuracy of the textual descriptions generated by both models and usage of McNemar’s test, to determine if there is a significant difference in performance, will be conducted afterwards.  Finally, the researchers will analyze the results and draw conclusions about the model's performance and its implications for improving accessibility in data visualizations. The findings will be documented in a comprehensive report, detailing the results of the model. |
| 1. Types of data to be studied: |
| ☐ Quantitative  ☐ Qualitative |
| 1. What data will be accessed? |
| ☐ Observation ☐ Simulation ☐ Other: \_\_\_\_\_\_\_\_\_\_\_\_\_  ☐ Experiment ☐ Secondary Data  ☐ Case study ☐ Samples |
| 1. How will your study team obtain the rights to access this data? |
| The researchers will use data source that is openly accessible which is the Gapminder. This platform provides statistical data that is freely available for public use. The researchers have reviewed the terms of use on the Gapminder website confirming that the data can be utilized for sharing and adaptation given that appropriate credit is given to the original source. |
| 1. How and where will the data be collected originally? (*Discuss your Research Locale & Data Collection)* |
| Data collection will come from credible sources, such as Gapminder, which contains a repository of global indicators. These data in CSV format will then be converted into a grid heatmap by utilizing an R package designed for creating complex heatmaps. This tool will allow the researchers effectively generate grid heatmaps. After generating the heatmap, experts in the field of statistics will provide the necessary ground truth by annotating textual descriptions for all the grid heatmaps created. This ensures quality interpretation and textual descriptions with the heatmaps. |



|  |  |
| --- | --- |
| 1. **DATA ACCESS & OTHER ATTACHMENTS** | |
| *Attach other research ethics documentary requirements (Specific format is not required.)* | |
| ☐ | Letter of Intent attached. (If educational data is requested, the permission must include a statement indicating if the data can be accessed without parental permission). |
| ☐  ☐  ☐  ☐  ☐ | Curriculum Vitae attached.  Certificate of Validity attached. (If the research questionnaire is self-administered/researcher-made/modified).  Informed Consent Form to include consent for archiving, presentation or publication, and subsequent sharing of the data.  ICF For Publishing. (documents user access and ability to publish/if the researchers plan to publish the paper)  N/A - Because only published material; open websites (no pass required) & no data agreement or application is requested by the owner. This still applies if the only pass required is payment of reasonable subscription price. (Note – if this is a data set, Exemption may be more appropriate.) |
| Was any member of the research team associated with the original research from which the data is being gathered or the individuals whose information will be studied?  ☐ Yes  ☐ No  ☐ N/A – Original data collection | |



**Reminder:** No research can be undertaken until your proposal has been approved by UREC.

****

|  |
| --- |
| **GUIDED QUESTIONS** |
| **Does the study involve living material (such as micro-organisms, plants, and/or animals)?**  ☐ Yes  ☐ No |
| **Will the participants give their explicit consent – on a voluntary basis – either digitally or on paper?**  ☐ Yes  ☐ No |
| **Does the study involve direct participants? *(those who are vulnerable or unable to give informed consent)***  ☐ Yes  ☐ No |
| **May the research or design procedure cause harm or discomfort to the participant in any way?**  ☐ Yes  ☐ No |
| **Is the study invasive? (i.e. affects the body such as taking blood or other body material from the participants)?**  ☐ Yes  ☐ No |

|  |
| --- |
| **TO BE FILLED OUT BY THE EVALUATOR** |

**Recommendation:** ☐ Approved

☐ Major Revisions Required

☐ Minor Revisions Required

☐ Disapproved

Remarks/ *Reasons for disapproval:*

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |  |
| --- | --- |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  *Signature over Printed Name of Reviewer* | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  *Review Date* |

|  |  |
| --- | --- |
| **Exempted Review** | Projects which involve the collection data from publicly available databases or public documents are exempted from review |
| **Expedited Review** | Projects posing minimal risk to research subjects go through expedited reviews. Projects qualifying for expedited review are those that involve:   * Research involving minor changes in previously approved research projects; * Research involving analysis of information without interaction with subjects; * Research, where informed consent is needed from the subjects and the informed consent process, will be correctly and appropriately applied, and that the researchers will be taken appropriate measures to protect the privacy of the subjects; * Research which is a local portion of a multi-center or multi-national research project has already received a full review from another research ethics committee or institutional review board. |
| **Full Review** | Research projects which pose a more than “minimal risk” to research participants or subjects are subjected to a full review. Risk is minimal when “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical examinations or tests” (U.S. Department of Health and Human Services, 1994, p.6, as cited by Hadjistavropoulos, & Smythe, 2001).   * Research involving vulnerable groups, such as the elderly, youth-at-risk, special children, or individuals who are in inequitable relationships; * Research involving sensitive topics, such as substance use, sexual behaviors, or criminal or politically sensitive behaviors; * Research with groups which necessitate permission to acquire access to them, such as research with indigenous communities; * Research which will require deception or which will be conducted without the participants’ full and informed consent at the time data is to be collected; * Research that will require access to personal and confidential information of identifiable individuals, such as genetic or biological information, medical records, or psychological assessment records; * Research that will cause physical and/or psychological harm or pain, or will cause humiliation, stress or anxiety; * Research that will involve intrusive interventions, such as hypnotherapy, drug administration, or vigorous exercise, which may cause participants to reveal information about themselves they otherwise would not normally want revealed in their everyday lives. * Research involving respondents through the internet * Research involving deceased persons, body parts or other human elements |

**CRITERIA FOR NON-HUMAN DETERMINATIONS**

|  |  |
| --- | --- |
| **STANDARD DETERMINATION (NEW)** | **SCENARIO** |
| Non-human subjects research  Self Determination HS1001 – Literature Review  – does not constitute human subjects research or require UREC review | Research that only collects data from published books, journals, or public facing websites that do not require a password or a data use agreement or other permission to access and where the posters have the authority to legally post the information. |
| Non-human subjects research –  Self Determination HS1002 – Oral History  – does not constitute human subjects research or require UREC review | Oral history interviews seek an in-depth account of personal experience and reflection, with sufficient time allowed for the narrators to give their story the fullness they desire. The content of oral history interviews is grounded in reflections on the past as opposed to commentary on purely contemporary events.” (Oral History Association – 2018)  **To claim the oral history self-determination**, the interview must allow the subject (narrator) to tell their story without analysis, manipulation or content editing and be collected with fully informed consent to include consent for archiving, presentation or publication, and subsequent sharing of the collected story.  Oral history projects must adhere to the principles and best practices established by the Oral History Association.  Reference: [Oral History Association Principles and Best Practice](file:///C:\Users\Acer%20User\Downloads\Oral%20History%20Association%20Principles%20and%20Best%20Practice),  [https://www.oralhistory.org/about/principles-and-practices/](https://www.oralhistory.org/principles-and-best-practices-revised-2018/) |
| Non-human subjects research –  Self Determination HS1003 – **Publicly Available** Data Sets Cleared as Non-human Subjects Research  – does not constitute human subjects research or require UREC review | Research that only collects data that is readily available to the public domain, such as websites that do not require a password or a data use agreement or other permission to access and where the posters have the authority to legally post the information. |
| Non-human subjects research –  Self Determination HS1004 – Market Research  – does not constitute human subjects research or require UREC review | Gathering information about customer or client needs and preferences for the purpose of improving the service provided.  This information is not generalizable beyond the market.  This self-determination also applies to data gathered for the purposes of accreditation documentation. |
| Non-human subjects research – Self Determination HS1005 – De-identified pre-existing data (Does not apply to clinical data.)  – does not constitute human subjects research or require UREC review **(Please read full description)**  (Does not apply to clinical data; Waiver of Authorization may apply.) | De-identified data: If the dataset has been stripped of all identifying information and there is no way that it could be linked back to the subjects from whom it was originally collected (through a key to a coding system or by any other means), its subsequent use by the lead researcher or another investigator would not constitute human subjects research, since it is no longer identifiable.  Identifiable means the identity of the subject is known or may be readily ascertained by the investigator or associated with the information. In general, information is identifiable when it can be linked to specific individuals by the investigator(s) either directly or indirectly through coding systems, or when characteristics of the information obtained are such that by their nature a reasonably knowledgeable person could ascertain the identities of individuals. Therefore, even though a dataset may have been stripped of direct identifiers (names, addresses, student ID numbers, etc.), it may still be possible to identify an individual through a combination of other characteristics (e.g., age, gender, ethnicity, and place of employment) and may require UREC review. |

***\*This form is adapted from the Non-human Subjects Determination IRB Form of Georgia Southern University. \****