Trying to answer the questions:

1. What does it mean to have insights in these topics?
2. What methods are in these directions?

Research proposal should try to contribute one or more of the following:

Advancements in methodology, knowledge, or applications.

New knowledge on how to collect data (remote sensing)?

Is there a way to couple different platforms (AI methods, CV methods, LLM methods) to do something interesting?

Trends towards more students having publications. Not sure what to do couples research process. By the end of summer we want results to do reports.

**Current Idea:**

**Research Proposal:** Enhancing Wildfire Profiling and Spread Forecasting Using Multi-Modal Fusion of RGB, IR, and Descriptor Feedback. Framework for multi-modal models.

**Hypothesis:** Predicting wildfire descriptors (e.g., fire intensity, spread direction, fuel consumption) and feeding them back into the model enhances long-term fire spread forecasting accuracy. Descriptors act as an intermediate representation, reducing noise and improving the model’s ability to learn wildfire dynamics. Incorporating IR imagery alongside RGB images improves fire profiling and forecasting by differentiating between heat-emitting flames and cooler smoke clouds. Multi-modal fusion of RGB, IR, and descriptors leads to more robust fire detection and spread forecasting, particularly in smoke-obscured conditions.

\*Descriptors refer to characteristics of the fire (i.e. occupancy, color)

**Possible tests.**

1. **Baseline:** Pass consecutive RGB images into a CV model to get descriptors of the next scene. Compare with ground truth descriptors for accuracy.
2. **Intermediate Descriptors:** Pass consecutive RGB images into a model to get out generated descriptors, pass descriptors + images into a model to get descriptors of the next scene
3. **RGB + IR fusion:** Pass consecutive RGB w/ IR into a model to get more accurate descriptors of the next scene
4. **RGB + IR w/ intermediate descriptors:** Pass RGB w/ IR into a model to get more accurate descriptors, and pass both descriptors + images into a model to get descriptores of the next scene
5. **Intermediate IR and descriptors:** Pass RGB images into a model to get IR images, pass both in to get more accurate descriptors, pass both descriptors and images into a model to get descriptors of the next scene (less interesting, not clear why and if, correlation is fairly speculative)

Direction, fire characteristics, and color may be more useful. Foliage is more useful. Which descriptors make sense? Generating IR may not be useful for the field and no research to support why 5 is necessary.

**Questions:**

1. What is the timing between each image? How far out should we be predicting? shouldn't longer == more useful)
2. Does it make sense to pass in both images and descriptors or can descriptors alone (middle product) be used?
3. Methods for generating descriptors or IR images alone seems like a research proposal that could precede this work
4. Can we train a model to predict IR image to get back the

**Future Directions:**

1. Expanding to include additional sensor data to produce more detailed, accurate, or further out forecasting
2. Comparing different models to generate intermediate descriptions
3. Producing human interpretable warnings from the predicted behavior

**TOPICS:**

**Addressing Data Sparsity:** Use **GPT-4 Vision** or similar models to impute missing descriptors, evaluating accuracy and impact on downstream tasks. (Feasible)

Notes:

* **Ideas:**
  + Test on 100 images manually labeled by team. Then compare manual labels to generated one
    - Can consider and compare models outside of GPT-4 (i.e. Gemini)
  + Can look at both the relationship between RGB images and descriptors, IR images and descriptors, and Fire Pixels and descriptors to compare how descriptors predict features of fires and whether they get at some truth
* **Relevant Papers:**
  + [A Novel Evaluation Framework for Image2Text Generation](https://arxiv.org/abs/2408.01723)
    - Input image into LLM → Outputs descriptors → Input descriptors into LLM → Outputs image → compute similarity score between images
  + ​​[Assessing GPT-4 multimodal performance in radiological image analysis - PubMed](https://pubmed.ncbi.nlm.nih.gov/39214893/)
* **Applications:**
  + Wildfire data is noisy and limited. Can help produce more human interpretable wildfire models by supplying downstream features that are easily understood. Can also help humans characterize wildfires quickly and in detail
  + Can possibly be used to build and test wildfire models for multimodal data (providing key descriptors as opposed to an image)
  + Can build a tool to detect early signs of unintended wildfire using real-time wildfire surveillance imagery (later - *temporal analysis*)
    - Using predicted descriptors to form predictions on the stages of fire as well as recommended procedures for firefighters and other public service personnels
* **Strengths:**
  + Off-the-shelf Deep Learning LLMs are available online or [low cost](https://openai.com/api/pricing/)
  + Can start by working with existing data, so there are few barriers
* **Challenges:**
  + Depends on how well descriptors inform visual features → images of wildfires can be noisy and we have a limited batch posing a question about the prediction power of the descriptors and whether we have enough data to identify patterns
  + Limited image and descriptors may have limited explanatory power to describe the variance in wildfires, ideally want additional information such as environmental conditions and temperature data
    - Descriptors such as *Occupancy* might not offer critical information as they are dependent on global factors (how far is the camera from the scene, some photos are taken while zoomed in on the fire and others are not, etc.)
* **Research Gap:** Limited exploration exists in applying LLMs to domain-specific descriptor generation for disaster events like wildfires. While existing image-captioning benchmarks offer evaluation strategies, wildfire-specific semantics (e.g., "dense smoke overlay") require tailored approaches.
  + What do colors of fire, of smoke signify ([Fire Smoke Color Interpretation: What You Need to Know](https://smokedsystem.com/fire-smoke-color-meaning/))

**Expanding Dataset Variability:** Increase **low-light, distance, and terrain diversity** in RGB + IR images to enhance model generalization and reduce bias.

Notes:

* Ideas:
  + Squishy fire images?
  + Data augmentation
  + Fire image sourcing online
  + Can we artificially generate RGB + IR data?
  + Can we use RGB + IR images from other situations to train a model to generate IR images to create more data
  + Use another model to take in RGB + IR images and generate accurate descriptors
  + Could even use one model to train on RBG + IR images (not wildfire) and then use for the RGB + IR images pipeline (transfer learning)
* Relevant Papers:
  + [WildfireSpreadTS: A dataset of multi-modal time series for wildfire spread prediction](https://proceedings.neurips.cc/paper_files/paper/2023/hash/ebd545176bdaa9cd5d45954947bd74b7-Abstract-Datasets_and_Benchmarks.html)
    - Purpose is for next day prediction which may be too coarse

**Improving Smoke and Fire Differentiation:** Enhance **superposition and occupancy calculations** with **multi-modal segmentation** (RGB + IR fusion) for better fire detection.

Notes:

* **Relevant Papers:**
  + [Enhancing Fire and Smoke Detection Using Deep Learning Techniques †](https://www.mdpi.com/2673-4591/62/1/7)
    - "The in-depth examination will highlight why image-based detection can be more efficient than currently employed sensors."
    - Proposed system achieves "an impressive 99% accuracy by harnessing the power of EfficientNet and YOLOv5 deep learning models".
  + [Improving Fire Detection Accuracy through Enhanced Convolutional Neural Networks and Contour Techniques](https://www.mdpi.com/1424-8220/24/16/5184)
  + [Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset](https://ieeexplore.ieee.org/abstract/document/9953997)
    - Early vs. late fusion of IR. Late fusion appears more robust because it can compensate when one channel is weak.
  + [Fire and Smoke Detection in Complex Environments](https://www.mdpi.com/2571-6255/7/11/389)
    - "Our modified model leverages the attention-based feature extraction capabilities of ViTs to improve detection accuracy, particularly in complex environments where fires may be occluded or distributed across large regions."
* **Ideas:**
  + Compare different fusion methods (early and late) and test the results. See if late fusion can improve prediction of RGB only models
* **Applications:** 
  + Smoke adds noise to the process of identifying fire pixels. This effort can improve the accuracy of fire identification and reduce noise in fire characterization.
  + This can be applied to prediction models to aid in the identification of fires in real time.
* **Challenges:**
  + Smoke may be useful for alerting to the presence or onset of fires, so on top of differentiating, identifying fire and smoke individually in the presence of each other and outside of the presence of the other is important
  + [Studies](https://www.mdpi.com/2673-4591/62/1/7) suggest fog, mist, and other similar phenomena are difficult to differentiate from smoke and our photos lack these. A whole other differentiation task is needed to improve the accuracy and utility of this type of model for real world conditions.
* **Research Gap:** Limited availability of annotated smoke and fire masks, especially under fog or mist conditions, hampers generalizability. Developing robust labeling tools or semi-supervised learning approaches could address this.

**Temporal Fire Analysis:** Assess how **sequential RGB + IR images** generalize across fire conditions using **LSTMs or Transformers** for wildfire progression modeling.

Notes:

* **Ideas:**
* **Relevant Papers:**
  + [Wildfire Danger Prediction and Understanding with Deep Learning](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022GL099368)
    - Research using deep learning for wildfire forecasting.
    - Points to relevant features
    - "We find that ML models focus more on soil moisture and relative humidity for their predictions. They also take into account the cumulative temperature and precipitation, as well as the last day's wind speed and the relative humidity in the 2 days preceding fire ignition. These associations lead to new, data-driven ways to anticipate wildfires."
  + [A combined real-time intelligent fire detection and forecasting approach through cameras based on computer vision method](https://www.sciencedirect.com/science/article/pii/S0957582022005675)
    - "Through the ratio conversion between the image and the actual size, the key fire developing characteristics can be obtained (fire spread position, fire front spread speed and flame width). Finally, ResNet takes these extractions as inputs and predicts those data after 30 s"
  + [Leveraging Large Language Models for Enhanced Classification and Analysis: Fire Incidents Case Study](https://www.mdpi.com/2571-6255/8/1/7)
  + [A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework](https://www.sciencedirect.com/science/article/pii/S0301479723006965?casa_token=rsh61h8vxzAAAAAA:V1hpl1AhejhyxPQGzXb7V_Ciag1hHGDsDIKSCRUdCqwrzZ2jxzavHgXGvCD9yA6DZzwvxw64t3Y)
* **Research Gap:** While temporal models exist for weather-informed fire prediction, vision-based temporal models remain scarce, particularly those combining IR and RGB modalities.
* **Applications**:
  + Using temporal analysis to determine how fast the captured fire is spreading
    - Smoke volume and velocity ([Reading Smoke Signals: How Firefighters Use Smoke Signals - BME Fire Trucks](https://www.bmefire.com/reading-smoke-signals/))
* **Challenges**:
  + Sequence photos in the Corsican dataset are likely contained fire, which can be useful in analyzing properties of fire in general but not an accurate representation of how wildfires progress

**Testing Model Generalization:** Benchmark **pretrained models** (Vision Transformers, YOLO, EfficientNet) to evaluate robustness across diverse fire descriptors and conditions.

Notes:

* Shape what contribution the work can provide
* Relevant prediction papers on the topic can be outside of wildfire paper
  + Image quantification data
* **Relevant Papers:**
  + [YOLO-Based Models for Smoke and Wildfire Detection in Ground and Aerial Images](https://www.mdpi.com/2571-6255/7/4/140)
    - "These models demonstrated their significant potential in handling challenging scenarios, including detecting small fire and smoke areas; varying fire and smoke features such as shape, size, and colors; the complexity of background, which can include diverse terrain, weather conditions, and vegetation; and addressing visual similarities among smoke, fog, and clouds and the the visual resemblances among fire, lighting, and sun glare."
* **Research Gap:** Existing wildfire detection benchmarks rarely consider semantic fire descriptors beyond basic classification (fire/no fire). A descriptor-level benchmark could shift research towards more interpretable and task-specific modeling.

**Hypothesis Formation:**

1. **Enhancing Fire Detection Accuracy with RGB + IR Image Fusion and Descriptor-Based Augmentation**
   1. **Can LLMs accurately provide fire scriptors? Would adding IR improve predictions?**
2. **Leveraging Large Language Models to Generate Fire Descriptors for Enhanced Wildfire Prediction**
   1. **Need to create ground truth first. Prompt engineering. Talk about the ways to improve prompt engineering.**
   2. **How do we determine what is useful? Compare different models. What do we want people to learn that will still be useful later on? Are some ways of prompting better than others? Are prioritizing different descriptors?**
3. **Generating Infrared (IR) Fire Images from RGB Data to Improve Fire Identification Models**

**Chain of reasoning or thought? Step by step compare to**

**1 and 3 more general. 2 most tied to things that will be outdated.**

### **Multimodal Descriptor Learning as a Foundation for Wildfire Prediction and Understanding (Combine Direction 1,4,5)**

#### **Step 1: Data Preparation**

* **Dataset: Use an existing wildfire dataset with aligned RGB + IR imagery (e.g., satellite or drone-based).**
* **Annotations:**
  + **Manually label a subset of ~100–200 images with fire-related descriptors (e.g., “dense smoke,” “flame edge,” “low thermal intensity”).**
  + **These serve as the ground truth for evaluating descriptor generation.**

#### **Step 2: Descriptor Generation**

* **Use vision-language models (e.g., GPT-4 Vision, BLIP-2, Gemini) to generate textual descriptors from the same images.**
* **Output: A set of auto-generated descriptors per image.**

#### **Step 3: Descriptor Quality Evaluation**

* **Compare auto-generated descriptors to human annotations using:**
  + **Textual similarity metrics (BLEU, BERTScore, CLIPScore)**
  + **Human expert review: Assess correctness, fluency, and fire-domain relevance.**
* **Purpose: Assess which model produces the most accurate and domain-useful descriptors.**

#### **Step 4: Static Fire Classification Modeling**

**Train and compare two distinct models:**

1. **Image-Based Model (Baseline)**
   * **Input: Raw RGB + IR images**
   * **Architecture: ViT or EfficientNet + classifier head**
   * **Output: Binary or multi-class fire classification**
2. **Descriptor-Based Model**
   * **Input: Tokenized descriptors (either human or LLM-generated)**
   * **Architecture: BERT or LSTM-based text classifier**
   * **Output: Same classification task as above**

**Goal: Compare whether visual or semantic (descriptor) features are more predictive.**

#### **Step 5: Temporal Fire Spread Prediction**

**Use sequential image data (e.g., 5–10 frame clips) to train two sequence models:**

1. **Image Sequence Model (Baseline)**
   * **Input: Sequence of RGB + IR frames**
   * **Architecture: CNN-LSTM or ConvLSTM**
   * **Output: Predict next-frame fire mask or spread region**
2. **Descriptor Sequence Model**
   * **Input: Sequence of descriptors generated per frame**
   * **Architecture: Transformer or LSTM**
   * **Output: Same fire spread prediction task**

**Goal: Assess whether temporal modeling using descriptors can capture progression as effectively as raw image sequences.**

#### **Step 6: Evaluation**

**Static Models (Classification):**

* **Accuracy, Precision, Recall, F1 Score**

**Temporal Models (Spread Prediction):**

* **Intersection over Union (IoU)**
* **Hausdorff Distance (boundary error)**
* **Qualitative assessment of prediction stability and interpretability**