# Prompts for Discussion

- Discuss short-term and long-term AI research efforts for this application
  - What is the minimum viable research effort? What does it look like?
  - How do we build in R2O pathways?
  - What other considerations exist for applying AI techniques to this application?
- What strategies are common to multiple applications (e.g. a single effort can benefit multiple applications)?
- What should NOAA consider when prioritizing our AI research investments for this application?

* Underline the top 3 strategies in each column

# Verification / Testbeds

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Approach development from user perspective
  - Forecasters, NWS core partners
  - Involve social scientists, academia, developers in evaluations
  - Invest on case studies, realtime, visualization
  - Use Testbeds (+ other venues) for evaluation
- Review metrics used for NWP
  - Expand them for MLWP (time consistency, covariances, physical balance, etc.)
- Streamline coordination of AI-efforts
  - Within NOAA: Increase transparency, communication, collaboration
  - Externally, e.g.,
    - Partner with AI faculty.
    - Create an AI student cohort (students, postdocs)

## Long-Term Strategies

- Include user input in verification for continuous training of emulator
- Raise the bar on metrics to increase physicality of models

* Underline the top 3 strategies in each column

# Data Assimilation

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Plugging in AI-based models into JEDI
    - Forward operators. Exposing the states to observations.
    - Focus on some initial technical work that needs to be done to enable longer term strategy.
    - Make available to community. Provide benchmarks for evaluation.
- Use of emulators for generation of ensembles for background error. Particular focus on emulators that generate ensembles that handle representation of uncertainty
- Emulators for TL/AD to enable fast 4DVar
- Improved use of observations – fast forward operators, quality control, bias correction
- Building fast capabilities for assessment of the impacts from observing system (FSOI).
- Use of output from DA to drive ML-based model error estimation and correction.

## Long-Term Strategies

- Blending of ML with DA (e.g. "hybrid loss/cost functions"). Do we need to reimagine (or rewrite parts of) JEDI in ML framework or vice-versa?
- ML-generative models to represent high dimensional systems/uncertainty
- Toward efforts to remove layers of processing for things like AMVs to get toward using observations more directly in their native form
- Blue-sky attempts to go directly from observations to emulation
- Getting beyond weather to coupled earth system models
- Enabling continuous 4DDA

* Underline the top 3 strategies in each column

# Short Range CAMs

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Framework for testing AI methods
  - Sandbox
  - Could be housed at CIs
  - Focus on year-out
  - Indentfiy stakeholder groups
  - Increase accessibility of NOAA data to downstream users
  - Open access tools for playing with AI-driven models
- Need to improve signposting of R2O opportunities to community
  - NCO coding standards
  - Resourcing
  - O&M tail
  - Metrics awareness/use/update
- Need to leverage cloud resources more
  - Parallel works vs native environment (NSSL WoFS)
  - On-prem HPC 1:1 R:O
  - Lower barriers to data accessibility
- CAM training dataset procurement
  - HRRR analysis dataset (3-km)
  - Focus sub-CONUS
  - Hourly high-res data → only need a year?
  - Resource estimation for 1 km reanalysis dataset
- Post-processing is minimum viable R2O pathway
  - JTTI
  - Already transitioned ML from CSU (GEFS)
  - HAILCAST model
  - NWS use

## Long-Term Strategies

- Generative AI for data-driven NWP
  - CAM ensembles
  - Merge latest observations
  - Probe uncertainties for needed products
  - Generation of additional realizations
- Need quality (re)analysis dataset at 1-km
  - Severe storm attribute climatologies
  - Severe wind reports
- Impacts-space estimation
  - Number of tornado tracks, lengths, etc…
  - Floods, hail, lightning
- NWS use of TORP for improving warnings
  - Have saturated skill with current obs/models
- Probable Maximum Precipitation Estimation
  - Events that are not currently in record
  - Infrastructure (power stations, dams) impacts

* Underline the top 3 strategies in each column

# Global MRW / S2S

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Coordinate real-time parallels of MLWP models, and get the data in front of forecasters to evaluate performance.
- Evaluate Physical consistency MLWP models
- Part of the R2O process needs to be setting up training datasets ARCO formats for Reanalysis **and** re-forecasts.
- Coordinate Testing and Evaluation of MLWP models. Similar to UFS-R2O project which has a lot of coordination between NOAA's research labs, NCEP, and the academic community.
- Work on developing a hybrid ML, physical coupled model for S2S forecasts.
- Develop/train ML models to represent ensemble forecast.

## Long-Term Strategies

- Physical model are still foundational to MRW/S2S forecasts and still need appropriate investments to be continued to develop.
- NOAA cannot to hire and retain AI experts, need to partner with the outside.
- Physical constraints reduce the amount of training data.
- Bring in NCO early to build infrastructure.
- R2O will have to be sped up (and can).

* Underline the top 3 strategies in each column

# Hurricane / Tropical

* Mark in red text strategies common to other applications

## Short-Term Strategies

- "Quick" wins
  - Unified physics emulation applied across regional system
  - Automated TCvitals
  - ML-based vortex initialization
  - ML-based RI, track, intensity models (advance current statistical models)
  - DA
  - Optimize consensus TC forecast (understand uncertainty of inputs)
- AI-enabling. Provide datasets, tools, and communicate needs to the broader community
  - TC-Primed
  - Hurricane reanalysis dataset
  - Global high-res datasets

## Long-Term Strategies

- Data-driven ensemble generation
- Transformative/High risk
  - High resolution global model emulation
  - High resolution hurricane model emulation (or regional model). Domain-wide or inner nest.
  - Train to predict threats/products based on initial state

* Underline the top 3 strategies in each column

# Reanalysis

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Understand requirements for training data
    - Do we need different reanalyses to train different emulators for different applications?
- Need for high-quality ensembles/TL/AD
    - Need testbed to determine confidence, maybe using hierarchy of simplified models/ob networks up to full complexity/density
- Publicize observational datasets for experiments to use AI to replace/augment DA
- Determine set of confidence-building steps to evaluate different ways of producing reanalysis (from obs -> reanalysis, ie no DA?)
- But need to determine what that reanalysis would be used for as well; can we make one that's all of climate record, initialization for reforecasts, etc?
- Toy benchmarking setup for AI research to explore idea of using obs data as input (e.g. replace DA); maybe a set of synthetic observations from ERA5, or surface pressure from 20CR.

## Long-Term Strategies

- Test a hybrid AI/physics based emulator to produce a reanalysis
- What does a reanalysis that uses an AI-only emulator look like vs a reanalysis that is used for AI training?
    - Determine experiments, concept of operations for how to bring AI in to reanalysis production
- Huge amount of person-years spent on QC and ob processing could be replaced with AI?
- Reconsider assumptions behind traditional DA
- Remain cognizant of vast use cases of reanalysis
- What are different possibilities of combining of AI/physics-based models/traditional DA in training AI?

* Underline the top 3 strategies in each column

# Group 1 Remote - All

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Provisioning / collecting satellite data across met agencies for training
- Defining the tech stack. e.g. GPU, Pytorch, etc.?
- There will be a proliferation of experiments. Who in NOAA is coordinating approaches/results? NCAI?
- Post processing is *still* easiest win
- Revisit and validate existing verification metrics for each application. Ensure we are optimizing against the correct metrics (with forecaster input)
    - Forecasters care about extreme events, not RMSE. We need to identify the correct metrics based on future role of the forecaster

## Long-Term Strategies

- Building reanalysis datasets of different resolutions, frequencies, esp at high resolution
    - 3 km reanalysis
- Try to replace model physics in current NWP with NN?
    - Hybrid systems will take time since many advancements are currently in pipeline (JEDI)
- Need to take bigger risks over the long term (NWS is moving in this direction)
- Process improvement in NOAA: It is no longer R2O, it is I2O (innovation to operations). Gradual RL advancement doesn't make sense in this space
- Identify ways to quantify the value of these new forecast systems

* Underline the top 3 strategies in each column

# Group 2 Remote - All

* Mark in red text strategies common to other applications

## Short-Term Strategies

- Get data in order first (AI-ready data). Nothing is successful without training data.
- Identify bite sized projects, and ensure room to fail. Ensure we learn a lot from our first projects
- Conduct evaluations of existing AI models including in testbeds.
- Immediately establish a team of 3-5 from across the agency and set some clear short term objectives

## Long-Term Strategies

- Government partnerships with academia and industry.
    - Government has benefit of path to operations
    - Explore mechanisms like NSF industry-university cooperative research centers
- Continue to develop relevant metrics, and optimize for these metrics (e.g. high impact events)
- Rapid reforecasts
- Ability to generate more rapid ensembles