AI for Operations

Not limited to agents

Logistics

- Notes (and your suggestions) can be found here: https://docs.google.com/document/d/1-SsV3pCR2bsPEL9FGeG__Zw2aGcSZ1fxlmtD1gJZZxQ/edit?usp=sharing

AI ready infrastructure

- Documentation readiness (cleaned to avoid old data)
  - Common format across experiments?
- Lower-level data readiness (e.g., detector control system data)
  - Quite experiment specific but cross experimental discussions might help specific developments
- What levels of resources are needed for the types of workflows envisioned?
  - What types of resources, if any, are not readily available now?
  - (crossover with infrastructure session) How much commonality between platforms can we envision across experiments?

Parallel or complementary thrusts/themes

- Detector and facility operations
  - Accelerators, ATLAS, CMS, DUNE, Mu2e, Cosmic frontier experiments
  - How can smaller experiments effectively interface with the larger dedicated operations programs?
- Software and computing operations
  - Data management, job submissions, debugging, porting of software, user support
- Common approaches where complementary needs to be defined
  - Agentic orchestrators
  - Human log ingestion
  - Software and application log ingestion
  - Anomaly detection
  - Interpretability

Foundations of collaboration : continuously collecting information across experiments

- “Repository” of needs
  - Help identify common needs
- Repository of models used
  - Including data structure ingested by models
  - Performance metrics…
- Benchmarks
  - Help avoid redoing previous work

Other discussion points

1) Humans in the loop and truly autonomous agents– when do we decide we trust an agent as much or more as a human? What validation steps and levels of rigor are required? How do we ensure appropriate safeguards until we get to that point so that we don't lose data? (this last one also has connections to training).
2) Knowledge preservation: feedback for later fine tuning or additional context.
3) Cost optimization (for lack of a better phrase): how do we decide when we need frontier models vs. others? What levels of reliability can we accept to make things more cost-effective?