Very good outline.

**A review of technical and scientific issues related to contemporary "Big Data" observational astronomy and surveys, specifically focused on the Rubin LSST survey and the Square Kilometer Array (SKA).**

- Abstract:
  - Summarize the importance of large scale astrophysical surveys
  - Talk about how the LSST and SKA are in the frontier for Big data in Astronomy
    - Mention specifics on how they are, the scope of the data they plan on collecting
  - Issues
    - Issues with real time processing,
    - How data is stored efficiently both in size and cost,
    - With Big Data structure and techniques becoming more advanced, people have to learn more
    - Transparency in the data pipeline,
    - lastly, talk about how important open source data is for not monopolizing data
  - Conclusion
    - Conclude with discoveries in Big data, algorithms for compressing data even more, machine learning, the future of Big Data in Astronomy
- Intro:
  - Historical Context:
    - Talk about how in the past, Big data was mostly not a big part of the budget. There were some problems with older telescopes such as the Slow and Digital Sky telescope but it wasn't as big as LSST and SKA.
    - Moore's Law
      - Talk about Moore's Law, how the data acquisition/transformation part is only going to develop faster and faster and it will be more expensive and it's not gonna be sustainable
      - So they have to come up with a solution like nuclear physicists have, use them as an example
  - Goal of Paper:
    - Talk about the Goal of the paper explicitly: To review the technical and scientific issues of big data in astronomy using LSST and SKA, summarize solutions that are being used right now, and challenges right now, and identify future trends in trying to solve the challenges or ways I might think they will.
  - Breaking Point in Data acquisition
    - LSST and SKA signify the breaking point in the importance of Big Data both in terms of the budget for the project and time so I can compare the 3 to show how much their goal and scales changed
    - Maybe split this into two parts, one for LSST, and one for SKA if its too long
- Methods:

Discuss the technical and scientific drivers of high/er data rates. eg. sensitivity, wide-field imaging, surveys, and technical advances. Can also mention some precursors (SDSS for LSST; MeerKAT/ASKAP for SKA)

LSST and SKA are canonical Big Data telescopes. They will transform obs. astronomy.

- Approach to collecting Data
    - Talk about how the LSST and SKA are collecting data
        - The nature of their data, ie the scope, type, etc
            - LSST is optical imaging
            - SKA is radio interferometry
        - How they're storing/archiving the data.
        - How they're generally processing their data

- Results:
    - Open Source Policies → move to Discussion
        - List what the LSST has done/plans to do concerning open source data
        - List what the SKA plans to do concerning open source data
    - Transparency of their respective Data pipeline
    - Real time processing techniques **(NOTE: NOT THE SAME AS THE METHODS DATA PROCESSING, THIS ONE IS FOR REAL TIME PROCESSING)**
        - Show new techniques being developed to combat issues such as optimization and time important issues → move to Methods.

Results.
- LSST commissioning / early science data achievements.
- SKA data challenges / tests.
- Brief precursor data achievements (SDSS, MeerKAT, ASKAP)

- Discussion:
    - Open source Data and Monopolization of data
        - Talk about how important it is that the data is open source so the data isn't monopolized
        - This is important for independent scientists, students, other people with limited funding/resources
    - The Challenge of the increase in skill needed
        - Talk about how as people join, one issue is that they have to learn more and more as the data pipeline and data in general becomes more complex
    - The Future of Big Data:
        - Talk about recent developments and the way AI and ML are going with the data rn
        - How LSST and SKA may be fighting against challenges such as open source data
        - Other potential future technologies, better algorithms, etc

- Conclusion:
    - Talk about how Big data is transforming in astronomy, with LSST and SKA pioneering it
    - Talk about how Big data is causing a lot of problems
        - Not sustainable
        - Open source data
        - Increasing need on both knowledge and budget
    - What's new
        - New techniques in data acquisition and AI/ML
    - End off on a good note
        - These problems

Move to Discussion. Conclusion rehashes key points of Discussion, but does not introduce new material.

- The fact that we are facing these problems is a good thing. It means we are advancing into a new age of data

**Key References and the Relevance of Them**

Big telescope, big data: towards exascale with the Square Kilometre Array - Astrophysics Data System

Relevance: Describes the SKA Data flow both in general and great detail, Gives the data rate. Explains the physics behind the Interferometric image formation SKA uses. Gives challenges that the SKA is facing. Important for the Methods section for SKA data processing and data storage.

The LSST Data Management System - Astrophysics Data System
Relevance: As the name suggests, it describes the LSST Data management System. The only thing is that it isn't as quantitative as I want it to be. I can just use it as a base to use its references. Important for Methods section for LSST how they're storing data and the scope of it

SKA Science Data Challenge 2: analysis and results - Astrophysics Data System
Relevance: It was an experiment run to test datasets with similar size of what they'd expect. They test processing algorithms to see how successful it would be. Very quantitative. Important for both methods and results because it shows how their algorithms can handle pressure

The LSST DESC DC2 Simulated Sky Survey - Astrophysics Data System
Relevance: Same thing as the above paper but for LSST concerning the data pipeline, seeing how good it can handle realistic sizes of data. Also talks about how being transparent in the data pipeline is good. Good for the methods and results section, also for discussion concerning open source data

Square Kilometre Array Science Data Challenge 1: analysis and results - Astrophysics Data System
Relevance: This scholarly paper describes the mission of SKA, and the issues in the scale of SKA's data acquisition in astronomy. Talks about the issue of Astronomy software being dated. Also talks about making the data accessible on an international level. Good for my introduction concerning SKA

AAS2RTO: Automated Alert Streams to Real-Time Observations: Preparing for rapid follow-up of transient objects in the era of LSST - Astrophysics Data System
Relevance: Presents a new python tool called AAS2RTO. It's an alert system for LSST for events where time is crucial using machine learning techniques. This is honestly so important for my real time processing section for LSST.

Machine-learning-based Brokers for Real-time Classification of the LSST Alert Stream - Astrophysics Data System

Relevance: Explains why LSST needs ML to automate such a big scope of data. Talks about a new brokering system called ANTARES. This is very important for my Discussion section for the future of ai/ml section, specifically LSST, as it is still not done yet.

Cataloguing the radio-sky with unsupervised machine learning: a new approach for the SKA era - Astrophysics Data System
Relevance: Explains why we need ML when it comes to LSST. Talks about using ML such as neural networks to optimize the data processing aspect. They trained a model by the three steps they talk about in section 2.1. Describes the challenges of their approach. Overall a good paper for my discussion on AI/ML section concerning SKA.

LSST: From Science Drivers to Reference Design and Anticipated Data Products - Astrophysics Data System
Relevance: Explains the goals for LSST, explains how its being built and why, describes anticipated data volumes, catalogs that will be available publicly. Describes the impact LSST will have on society and education. A great paper for both my intro to explain LSST, and discussion for societal impact of LSST