SWAPTIK CHOWDHURY

Incorporating Equity into Decision-Support Models

An Analytical Architecture and Application to Natural Gas Power Plant Retirement in California

This document was submitted as a dissertation in October 2026 in partial fulfillment of the requirements of the Frederick S. Pardee Ph.D. in Policy Analysis at the RAND School of Public Policy. The faculty committee that supervised and approved the dissertation consisted of Nidhi Kalra, Kelly Klima, and Piotr Zebrowski.

Dissertation

For more information on this publication, visit **www.rand.org/t/RGSDA5479-1**.

About the RAND School of Public Policy

The RAND School of Public Policy has specialized in graduate-level policy education since its founding in 1970. The RAND School is home to the Frederick S. Pardee Ph.D. in Policy Analysis, which is the original public policy Ph.D. program in the United States and the only Ph.D. program based at an independent public policy research organization. To learn more about the RAND School of Public Policy, visit www.rand.edu.

Published in 2026 by the RAND Corporation, Santa Monica, Calif.

RAND® is a registered trademark.

RGSD-A5479-1

Abstract

Environmental and climate-related decisions often have significant equity consequences. Those consequences are increasingly central to policy debate and can affect the legitimacy and durability of the decisions themselves. Yet the analytical models used to inform such decisions rarely incorporate equity in a transparent or systematic way. Even when equity is included, it is usually reduced to distributional equity, with much less attention given to contextual and procedural equity.

This dissertation addresses these limitations by developing and applying an analytical architecture for integrating equity into quantitative decision-support models. The architecture is applied to develop a phaseout plan for California's 96 natural gas powerplants to meet the state's environmental goals. The dissertation developed a method for using a retrieval-augmented large language model (LLM) pipeline to extract and synthesize literature-grounded judgments about indicator importance and equity type, retaining disagreement across models as a source of uncertainty. The resulting model evaluates phaseout priorities under utilitarian, sufficientarian, and egalitarian justice principles, which serve as proxies for stakeholder values and perspectives.

The analysis shows that incorporating equity more fully can produce meaningfully different phaseout rankings than an approach limited to distributional equity. It also shows that equity-informed phaseout rankings can distinguish phaseout decisions that are relatively robust to competing stakeholder values and perspectives from decisions that are much more sensitive to those values. Rankings converge most clearly at the top and bottom of phaseout ranking and diverge most in the middle. This pattern helps identify where stakeholder deliberation is most needed.

Table of Contents

Abstract iii
Table of Contents iv
Figures and Tables vi
Acknowledgements 1
Abbreviations 2
Executive Summary 4
Chapter 1. Introduction 6
Defining Equity and Justice 7
Why Energy Equity Matters 7
California's Commitment to Energy Equity 8
Existing Frameworks for Assessing Energy Equity 8
Limitations of Current Analytical Approaches 9
Case Study of Natural Gas Powerplant in California 11
Objective 12
Research Questions 12
Organization of this Thesis 12
Chapter 2. Methods 14
2.1 Shapes of Justice 17
2.2 Instantiating the Analytical Architecture for California Natural Gas Power Plant Phaseout 19
2.3 Equity Component Classification 45
2.4 Ranking Methodology and Sensitivity Analysis 47
2.5 Planning Exercise: CPUC Gas Capacity Not Retained Ranking 49
Chapter 3. Results 52
3.2 Plant Rankings Under Three Shapes of Justice 52
3.3 Referencing Against CPUC Gas Capacity Non-Retention Methodology 64
Chapter 4. Discussion and Implication 67
4.1 Can distributional, contextual, and procedural equity be operationalized in a quantitative decision-support model, and what design requirements does doing so impose? 67
4.2 Does incorporating those equity indicators into a decision-support model produce meaningfully different phaseout rankings than a model relying on distributional equity indicators alone? 69
4.3 Which plants are ranked similarly across different shapes of justice, making them stronger candidates for robust policy action, and which plants are ranked differently, suggesting a need for further stakeholder deliberation? 70
4.4 Commentary on the LLM-RAG-based Weight Extraction Method 73
4.5 Limitations and Future Work 75
4.6 Conclusion 77
Appendix A. Selected List of Indicators Not Included in the Model 79

Appendix B. Cross-Model Consistency of LLM-Based Weight Assigned to indicators and Metrics 83
Appendix C. Cross-Model Consistency of Equity Component Classification of Indicators and Metrics 87
Appendix D. Top 20 Powerplants by Median for Early Phaseout Under Each Shape of Justice. 90
Appendix E. Indicator Direction and Normalization Treatment 93
Appendix F. Prompts 95
References 101

Figures and Tables

Figures

Figure 1: A Framework for Equity 9
Figure 2: Analytical Architecture for incorporating equity into decision-making 16
Figure 3: Indicator Weights by Shape of Justice 44
Figure 4: Metric Weights by Shape of Justice 44
Figure 5: Top 20 Powerplants by Median Rank Under Each Shape of Justice 53
Figure 6: Powerplants by Median Rank between 25 to 75 Under Each Shape of Justice 56
Figure 7: Powerplants by Median Rank for Late Phaseout Under Each Shape of Justice 57

Tables

Table 1: Metrics and Associated Indicators used for NG Powerplant Phaseout Planning 23
Table 2: Summary of Indicators used for NG Powerplant Phaseout Planning 24
Table 3: Equity Component Classification of Metrics 27
Table 4: Global Sensitivity Analysis: Indicator Importance Scores Under Utilitarianism 59
Table 5: Global Sensitivity Analysis: Indicator Importance Scores Under Sufficientarianism 60
Table 6: Global Sensitivity Analysis: Indicator Importance Scores Under Egalitarianism 61
Table 7: Top Four Indicators by Importance Under Each Shape of Justice 61
Table 8: Potential Indicator and Their Exclusion Criteria 79
Table 9: Summary of Priority 1 vs Priority 2 patterns for indicators by model and justice type 83
Table 10: Summary of Priority 1 vs Priority 2 patterns for metrics by model and justice type 84
Table 11: Fleiss Kappa and ICC(2,k) by Shape of Justice for indicators and metrics 85
Table 12: Equity Component Classification of Indicator by Model 87
Table 13: Equity Component Classification of Metric by Model 88
Table 14: Summary of Priority 1 vs Priority 2 patterns for indicator equity classification by model 88
Table 15: Summary of Priority 1 vs Priority 2 patterns for metric equity classification by model 88
Table 16: Top 20 Powerplants by Median for Early Phaseout Under Utilitarianism 90
Table 17: Top 20 Powerplants by Median for Early Phaseout Under Sufficientarianism 90
Table 18: Top 20 Powerplants by Median for Early Phaseout Under Egalitarianism 91
Table 19: Indicator and their normalization direction 93

Acknowledgements

First, I would like to thank my maa, bapi, and my brother Swapneel. They are my rock, and they motivated me throughout this largest singular effort of my life, undertaken in an uncertain environment. This journey could not have been possible without them.

I am extremely thankful to my committee members. I am deeply grateful to my committee chair, Dr. Nidhi Kalra, for her constant guidance and mentorship. It was a project with her that first sparked my interest in studying equity in climate change. She has provided support and help in every part of my doctoral journey, and it has been a pleasure to learn research leadership from her. I am grateful to Dr. Kelly Klima for investing in both my intellectual and career development and for giving me the foundation to grow as a researcher. I learned critical skills about every part of the research pipeline from her, and I could not have asked for a better or kinder mentor. I am also thankful to Dr. Piotr Zebrowski for the many discussions about equity under a literal Habsburg castle. Our repeated discussions over the last three years helped me mature this study. I am also grateful to him for giving me the opportunity to participate in YSSP, for introducing me to IIASA, and for introducing me to one of my lasting obsessions, Vienna.

I want to thank Dr. Steven Popper, Dr. Rob Lempert, Dr. Tim Marler, Dr. Rhianna Rogers, Dr. Debra Knopman, Dr. Tepring Piquado, Dr. Todd Richmond, and Dr. Osonde Osoba for their continued mentorship and guidance. Research is an apprenticeship, and I could not have better teachers and mentors.

I am also grateful to David, Nihar, and James for good discussions on energy infrastructure and policy modeling. I am especially thankful to my friends Vikas, Aakash, Aastha, and Prishita, who supported me through the ups and downs of this journey. They let me vent, helped me work through the problems that came up along the way, and talked me through the decisions about what to do next. I could not have navigated this period as well without them. Finally, I want to thank Will, Ishita, Salil, Jim, Piyush, and Mugdha for the regular conversations and company that kept me grounded along the way.

I am also thankful to the Social and Economic Well-Being Center (erstwhile) and the Frederick S. Pardee Center for Longer Range Global Policy and the Future Human Condition at RAND Corporation. I am also grateful to the Roger Levien Fellowship and Friends of IIASA for providing the YSSP fellowship scholarship.

Abbreviations

Abbreviation	Definition
ACS	American Community Survey
ADI	Area Deprivation Index
BLS	U.S. Bureau of Labor Statistics
CAISO	California Independent System Operator
CEC	California Energy Commission
CES	CalEnviroScreen
CO_2	Carbon dioxide
CPUC	California Public Utilities Commission
DAC	Disadvantaged Community
FAISS	Facebook AI Similarity Search
GEM	Global Energy Monitor
GHG	Greenhouse gas
GOGPT	Global Oil and Gas Plant Tracker
GQA	Grouped-query attention
GSA	Global sensitivity analysis
ICC	Intraclass correlation coefficient
IQR	Interquartile range
JAEDI	Justice, Access, Equity, Diversity, and Inclusion
LCOE	Levelized Cost of Energy
LEAD	Low-Income Energy Affordability Data
LLM	Large language model
LQ	Location Quotient

MCDA	Multi-criteria decision analysis
MCDM	Multi-criteria decision-making
MDI	Mean Decrease in Impurity
MSA	Metropolitan Statistical Area
NAICS	North American Industry Classification System
NG	Natural gas
NO_x	Nitrogen oxide
NREL	National Renewable Energy Laboratory
OEWS	Occupational Employment and Wage Statistics
QFER	Quarterly Fuel and Energy Report
RAG	Retrieval-augmented generation
RF	Random Forest
RLHF	Reinforcement learning from human feedback
SLOPE	State and Local Planning for Energy
SMAA	Stochastic Multicriteria Acceptability Analysis
SO_2	Sulfur dioxide
SOC	Standard Occupational Classification
WSM	Weighted Sum Model

Executive Summary

Environmental and climate-related decisions often have significant equity consequences. Those consequences are increasingly central to policy debate and can affect the legitimacy and durability of the decisions themselves. Yet the analytical models used to inform such decisions rarely incorporate equity in a transparent or systematic way. Even when equity is included, it is usually reduced to distributional equity, with much less attention given to contextual and procedural equity. Models also tend to treat equity as a single, fixed concept, even though different justice principles define and prioritize it in different ways. As a result, the analytic basis for decisions often fails to capture the full range of equity considerations that policymakers claim to value.

This dissertation addresses these limitations by developing and applying an analytical architecture for integrating equity into quantitative decision-support models. The architecture is applied to a consequential policy decision facing California: the order in which its 96 natural gas power plants should be phased out to meet the state's environmental goals. This decision has significant equity implications, including environmental harm, reliability and affordability concerns, employment effects, and other impacts that fall unevenly across communities. It must also be addressed against the backdrop of longstanding energy inequities, including the disproportionate siting of energy infrastructure in historically marginalized communities. As the world's fifth-largest economy and a leader in climate policy, California's approach to advancing its climate and equity goals will also influence how other governments confront these issues.

The dissertation addresses three research questions. First, can distributional, contextual, and procedural equity be operationalized in a quantitative decision-support model, and what design requirements does that impose? Second, does incorporating those equity concepts produce meaningfully different phaseout rankings than a model relying on distributional equity indicators alone? Third, which plants are ranked similarly across different shapes of justice, making them stronger candidates for robust policy action, and which are ranked differently, suggesting a need for further stakeholder deliberation?

To answer these questions, this dissertation built an analytical architecture through several key steps: defining the shape of justice, identifying decision-relevant indicators and metrics, classifying the indicators and metrics by dimensions of equity, assigning weights to represent different shapes of justice, and producing rankings through a multi-criteria decision-making (MCDM) model. The scholarly literature does not provide a ready-made set of weights or equity classifications. The dissertation, therefore, developed a method for using a retrieval-augmented large language model (LLM) pipeline to extract and synthesize literature-grounded judgments about indicator importance and equity type. A protocol distinguishes between assignments directly supported by literature and those requiring more interpretive reasoning, and disagreement across

models is retained as a source of uncertainty in the analysis. The resulting model evaluates phaseout priorities under utilitarian, sufficientarian, and egalitarian justice principles and examines how rankings change across those principles under uncertainty and what drives that variation.

This dissertation makes three contributions that are consistent with the research questions. First, it provides novel analytical architecture for transparently and systematically integrating multiple equity dimensions into decision-support models. In terms of design requirements, it demonstrates how open-source LLMs can be used as a reproducible method for extracting literature-grounded weights while preserving uncertainty in the results. It also highlights a limitation of how decision support models can incorporate equity: 12 of the 16 indicators were classified as distributional equity, 3 as contextual equity, and none as procedural equity. Procedural equity remains difficult to represent as it is a measure of direct stakeholder participation in the decision process itself.

Second, it shows that incorporating equity more fully can produce meaningfully different phaseout rankings than an approach limited to distributional equity. A comparison with the CPUC's Gas Capacity Not Retained ranking shows substantial differences between the two approaches. These differences have practical implications. For example, this dissertation supports the early phaseout of Oxnard Power Station but assigns a lower phaseout priority to Alamitos Energy Center and Scattergood Power Station, two plants that have also received attention for early phaseout in California. Although the comparison has limitations, it suggests that a more systematic treatment of equity could materially change California's phaseout priorities.

Third, it shows that equity-informed rankings can distinguish between phaseout decisions that are relatively robust to model uncertainty and competing definitions of equity, and decisions that are much more sensitive to those assumptions. Rankings converge most clearly at the top and bottom, where some plants are consistently prioritized or deprioritized for phaseout across all three shapes of justice. They diverge most in the middle, where plant rankings are more sensitive to choices of weights and shapes of justice. This pattern helps identify where stakeholder deliberation is most needed.

This analysis has several limitations that point to future work. The findings are based on a single case study, so the framework should be applied and tested in other policy settings. The LLM pipeline also depends on retrieval quality, prompt compliance, and model instruction-following; although that variation is reflected as uncertainty, the outputs still depend on the reliability of the pipeline itself. Most importantly, the analysis does not yet incorporate direct stakeholder deliberation. In California, the rankings are best used to inform discussion about the role of equity in phaseout models and to focus deliberation on the plants shown most sensitive to competing assumptions. In general, future work should examine how this architecture can support deliberation by helping identify decisions that are stable and where disagreement over equity has the greatest effect. Even with these limitations, the dissertation provides a strong foundation for improving equity-informed decision-making.

Chapter 1. Introduction

Policies aimed at advancing equity in energy reflect a growing recognition of the systemic and historical injustices faced by marginalized communities in the United States (The White House, 2024; Chu and Cannon, 2021; Minow, 2021). California has demonstrated a strong commitment to equity across various sectors through initiatives such as the Equitable and Resilient Communities Initiative, Executive Order N-16-22, and Senate Bill 17 (Governor's Office of Planning and Research, 2017; Office of the Governor of California, 2022; California Legislature 2022). These efforts aim, respectively, to address social disparities through equity-centered planning, embed equity analysis in state policies and practices, and propose a racial equity commission to address systemic racial inequities.

Policies aimed at advancing equity in energy reflect a growing recognition of the systemic and historical injustices faced by marginalized communities in the United States (The White House, 2024; Chu and Cannon, 2021; Minow, 2021). However, most studies focus narrowly on *distributional* equity[1], often overlooking its *procedural*[2] and *contextual*[3] dimensions. Moreover, existing analytical tools used to inform policymaking rarely integrate equity systematically.[4]

This dissertation addresses these limitations by developing a novel analytical architecture for incorporating equity more completely and systematically into decision-support models. The architecture is applied to inform a natural-gas power-plant phaseout strategy for California. Key equity-relevant indicators are identified through a scoping literature review, and their relative importance (weights) under different resource allocation principles (or shapes of justice) are identified from literature using large language models (LLMs). Metric data and weight are used to create a multicriteria decision-making (MCDM) model. The model evaluates how different shapes of justice and equity metrics influence phaseout priorities. While LLM-based structured extraction from the literature guides the initial implementation (Lewis et al., 2020; Gao et al., 2023), the architecture is designed to accommodate broader stakeholder input and respond to common critiques of technocratic, top-down approaches. This work's main contribution is to advance the methodological foundations of equitable climate and energy policy by embedding a multi-

[1] Programs and policies that result in fair distributions of benefits and burdens across all segments of a community, prioritizing those with the highest need. (Park & Urban Sustainability Directors Network, 2014)

[2] Inclusive, accessible, authentic engagement and representation in processes to develop or implement sustainability programs and policies. (Park & Urban Sustainability Directors Network, 2014)

[3] Refers to the inclusion of initial social conditions in assessing the social impact or fairness of a project or policy (McDermott et al., 2013)

[4] See Limitations of Current Analytical Approaches in chapter 1 for a detailed review

dimensional conception of equity into decision-support models, thereby supporting more just and inclusive energy transitions.

The remainder of this chapter first defines the concept of equity and clarifies how it differs from justice. Second, it places equity within the broader context of energy transitions, illustrating how the shift to clean energy can exacerbate existing disparities. Third, it reviews key California policies that attempt to embed equity into energy planning. Fourth, it evaluates existing frameworks and analytical methods for assessing energy equity in policymaking. Finally, it identifies limitations in current modeling approaches that support decision-making but fail to meaningfully incorporate equity concerns, underscoring the need for new methods to include equity in analytical models

Defining Equity and Justice

Equity is a multifaceted and complex concept. It emphasizes fairness in process and outcomes across different groups, considering factors such as distribution, recipients, means, and historical context (Tarekegne et al. 2021; Crosman et al. 2022; Lewis, MacKenzie, and Kaminsky 2021). While often used interchangeably with justice, equity and justice are distinct concepts: justice focuses on people's rights, whereas equity is a comparative notion of fairness between groups (Grasso 2007; Minow 2021; Chu and Cannon 2021).Grasso (2007) distinguishes the three concepts hierarchically: Justice pertains to ethical principles that "exist independently before any process of judgment or interpersonal comparison has begun," equity provides the normative criteria for implementing principles or theories of justice in a given context, and fairness pertains to the "individual's perception arising from a judgmental process".

Why Energy Equity Matters

Energy transitions, particularly the shift to clean energy sources, are important for combating climate change (Maia and Garcia 2023; Yang et al. 2024). However, transitions also raise significant equity concerns. These actions can inadvertently create winners and losers, often disproportionately burdening already disadvantaged and vulnerable communities (Carley and Konisky 2020; Baker, Goldstein, and Azevedo 2021). For example, projects such as lithium mining and offshore wind development are frequently situated on Indigenous lands, exacerbating histories of exploitation and resource extraction, and fueling systemic inequities (Raimi, Ehrman, and Curley 2023). The inequitable impacts of the energy transition may also include the loss of jobs in the traditional energy sector and the worsening of energy poverty and energy insecurity for low-income communities (Chen et al. 2024).

Energy equity is about ensuring fair access for everyone to the benefits of electricity, natural gas, transportation fuels, and other energy-related uses, while fairly sharing the burdens of transitions (Kalra et al. 2022; Barlow, Tapio, and Tarekegne 2022; Crespo Montañés et al. 2023; Heffron, McCauley, and Sovacool 2015; Hall, Hards, and Bulkeley 2013). It calls for examining

the entire life cycle of energy resources to assess how energy systems impact different communities (Carley and Konisky 2020).

California's Commitment to Energy Equity

California provides an important case study of how equity considerations are being operationalized within energy policy. For almost a decade, the state has enacted various initiatives to ensure that the transition to clean energy does not perpetuate or intensify existing disparities. These include the California Energy Commission's (CEC) 2018 Energy Equity Indicators, which identify opportunities to improve access to clean energy technologies for low-income households and disadvantaged communities; the CEC's 2023 Justice, Access, Equity, Diversity, and Inclusion (JAEDI) Framework, which seeks to embed equity and environmental justice in energy planning; and Senate Bill 535 (SB 535), which directs resources toward communities disproportionately affected by climate change and environmental pollution (California Energy Commission n.d.; California Energy Commission 2023; Office of Environmental Health Hazard Assessment n.d.). However, a persistent challenge lies in translating broad aspirations of equity into equitable decisions.

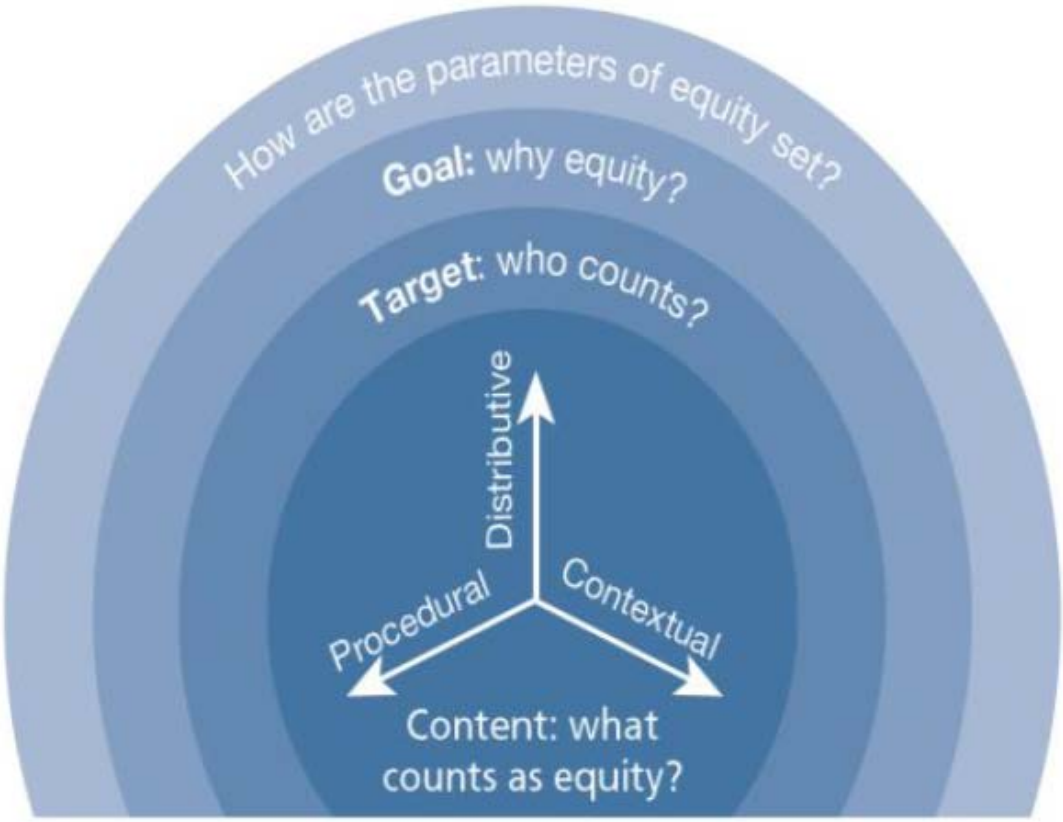
Existing Frameworks for Assessing Energy Equity

Efforts to define and assess energy equity frequently draw from energy justice principles (Sovacool and Dworkin 2015; Heffron and McCauley 2017; Heleno et al. 2022). The energy justice literature identifies three key justice principles: distributional justice, procedural justice, and recognition justice. Distributional justice examines the relative distribution of costs and benefits among different groups, emphasizing that no group should bear disproportionately high costs or accumulate disproportionately high benefits from the energy transition. Procedural justice ensures equal access and participation for all stakeholders in energy decision-making, particularly for communities that have historically faced systemic injustices, such as Indigenous communities, low-income individuals, and people of color. Recognition Justice requires understanding the role of historical and social inequalities and situating outcomes within their broader social context (McCauley et al. 2013; Sovacool et al. 2017; Sovacool and Dworkin 2015; Sovacool et al. 2016; Jenkins et al. 2021; Heffron and McCauley 2017). An additional principle, restorative justice, has also emerged, focusing on enabling disadvantaged communities to articulate and ameliorate historical harms (Heffron and McCauley 2017).

These principles have informed the development of approaches intended to embed equity into energy policy. For example, McDermott et al. (2013) developed a comprehensive and multidimensional equity framework comprising four interrelated parameters that define equity (Figure 1) (McDermott, Mahanty, and Schreckenberg 2013). The first parameter examines "what counts as equity" and consists of three dimensions: distributional equity, procedural equity, and contextual equity. The second parameter explores "who counts" and delves into the groups for

which equity is sought. The third parameter contextualizes the purpose of considering equity by asking "why equity," while the fourth parameter addresses "how the equity parameters are determined." Other energy equity approaches in the literature also align with the equity framework developed by McDermott et al (Energy Equity Project n.d.; American Council for an Energy-Efficient Economy n.d.; California Energy Commission 2023; Abdul-Khabir 2023; Forrester and Satchwell 2023).

Figure 1: A Framework for Equity



NOTE: McDermott et al. (2013)

Limitations of Current Analytical Approaches

The influence of energy equity principles on actual decision-making depends on whether and how they are represented in the analytical models used in decision-making. In the energy sector, policymakers may rely on results from quantitative models and computer simulations to inform

evidence-based decision-making (Susser et al. 2021; Mara et al. 2020; Prasinos et al. 2022; James et al. 2021). However, many analytical models used in climate and energy policy are not designed to prioritize equity (Giang et al. 2024; Weyant 2017). As a result, equity is often treated as a secondary concern, leading to a misalignment between the goal of equitable decision-making and the limitations of the analytical tools used for it. Even when equity is included, the approaches remain limited in scope.

For example, Jafino et al. (2021) sought to develop a practical approach for model-based policy analysis to support equitable delta adaptation planning. They identified the distributional impacts of climate change planning and used a watershed planning model to calculate equity outcomes. Yang et al. (2023), Jafino et al. (2022), and Ciullo et al. (2020) have adopted similar approaches. Some studies have also applied linear and multi-objective programming methods to consider equity in energy applications. For example, Heleno et al. (2022) have used linear models to optimize energy investment to minimize energy burden. Ma et al. (2020) employed a multi-objective programming model to incorporate equity principles for coal de-capacity allocation. In these efforts, analytical models focus primarily on distributional equity and often fail to integrate procedural and contextual equity. This limits the comprehensiveness of these assessments. The current modeling approach also masks disproportionate policy impacts on specific subgroups (Cardiff 2020; Schumacher 2018).

Additionally, stakeholders often hold diverse and conflicting perspectives on fairness, justice, and resource distribution, which can result in distinct "shapes of justice," meaning different ways in which justice principles are interpreted and applied (Hyatt and Gruenglas 2023; Williams et al. 2023; Müller and Renes 2021; Zimm et al. 2024). For example, some may favor utilitarianism, which focuses on maximizing total welfare, often by selecting economically optimal pathways based on assumptions about the relationship between consumption and well-being. Others may favor an egalitarian view, which focuses on reducing disparities among groups by ensuring equal distribution of resources. Still others may prefer sufficientarianism, which prioritizes meeting a minimum threshold of well-being for all individuals, ensuring that basic human needs are satisfied. Thus, differing perceptions of fairness and justice complicate the development of consistent metrics and methods for incorporating equity into analytical models. As Kalra et al. (2022) note, while literature acknowledges the need to harmonize different definitions of fairness and justice, it lacks a coherent process for doing so.

Further, analytical models often use a top-down approach, in which experts set the assumptions, variables, and priorities that shape decisions. As such, the models may not capture the needs and lived experiences of affected communities and may foster mistrust among them, reducing buy-in into the resulting policy (Porter and Dessai 2017; Mendoza and Prabhu 2005; Kiker et al. 2005).

So, the models used for decision-making must go beyond assessing distributive outcomes. To support equitable energy policymaking, they must incorporate procedural and contextual dimensions of equity, including recognizing historical and structural disadvantages and integrating

community values into the modeling process. Therefore, methodologies that can systematically evaluate the equity implications of just transition policies are essential. This dissertation aims to achieve this.

Case Study of Natural Gas Powerplant in California

This thesis develops an analytical architecture that improves how equity is considered in policymaking. The architecture describes and demonstrates the data, metrics, models, and methodology to better incorporate equity into decisions. To demonstrate this analytical architecture, this thesis uses the decommissioning of natural gas (NG) power plants in California as its case study.

California has been the world's fourth-largest economy and the second-largest total energy consumer in the United States (U.S. Energy Information Administration, n.d.; State of California - Office of the Governor, 2024). Natural gas accounts for the largest share of installed in-state generation capacity, producing approximately 45 percent of the total capacity (California Energy Commission, 2023). The state has enacted several regulations aimed at reducing greenhouse gas emissions and achieving carbon neutrality. Examples include SB 32, which mandates a reduction in GHG emissions to 40% below 1990 levels by 2030; AB 32, which directs the California Energy Commission to assess the state's potential to meet that target; SB 100, which requires that retail electricity be sourced entirely from renewable energy and zero-carbon resources by 2045; and Executive Order B-55-18, which establishes a statewide goal of carbon neutrality by 2045 (California Global Warming Solutions Act of 2006: Emissions Limit., 2016; Executive Order B-55-18 to Achieve Carbon Neutrality, 2018; AB 32 Global Warming Solutions Act of 2006 | California Air Resources Board, 2018; California Energy Commission, n.d.). However, current trajectories suggest it may fall short of this target, making the transition from natural gas a significant and urgent policy priority (Lazo 2024; California Energy Commission n.d.).

California's continued reliance on natural gas also poses serious environmental and public health risks (Shetty et al., 2025). Natural gas power plants emit a range of pollutants, including carbon dioxide (CO_2), methane (CH_4), and nitrogen oxides (NO_x), which contribute to climate change, smog formation, and increased incidence of respiratory and cardiovascular diseases, as well as premature death. Furthermore, gas-fired power plants have been disproportionately sited in low-income and disadvantaged communities, exacerbating long-standing environmental injustices (Declet-Barreto and Rosenberg 2022).

As California pursues carbon neutrality by 2045, the state will need to determine how to sequence the retirement of its natural gas power plants. This requires deciding the order in which plants should be decommissioned, a decision with significant equity implications for the communities where these plants are located.

The increasing frequency and severity of climate-related hazards in California, including more severe wildfires and more frequent and intense droughts, heighten the urgency of reducing

greenhouse gas emissions for mitigating climate change (Crimmins et al. 2023; Petek 2022). The NG powerplant phaseout can support this goal by reducing emissions from fossil-fuel-based electricity generation. Also, given the shifting pace of federal commitment to climate policy, California's leadership in advancing equitable climate and energy planning is particularly significant (Cart et al. 2024). As a state that consistently prioritizes climate considerations in its policymaking, California is well-positioned to play an outsized role in both achieving national emissions goals and shaping more equitable approaches to energy transition.

Recent literature has also noted a partisan distortion in the public perception of extreme climate events, such as wildfires, and discrepancies in climate change coverage between urban centers and rural "heartland" areas (Hui et al. 2020; Bolstad and Victor 2024). These dynamics create a highly active and contested space for climate change analysis and policy deliberation in California. Also, data on power plants and emissions enable a robust analytical approach to examine equity in phasing out natural gas power plants during California's energy transition (Global Energy Monitor 2026; U.S. Energy Information Administration n.d.).

A key step in decommissioning is determining the order in which NG power plants can be phased out to meet the state's goals, including climate, energy security, and equity. Given the state's outsized economic and energy profile, this case study on California's natural gas power plant phaseout is particularly relevant for exploring equity in climate change mitigation.

Objective

The objective of this thesis is to develop an analytical architecture for integrating a more complete approach to equity into analytical models used for decision-making and to demonstrate its applicability and impacts in the context of California’s NG powerplant phaseout.

Research Questions

RQ1: Can distributional, contextual, and procedural equity be operationalized in a quantitative decision-support model, and what design requirements does doing so impose?

RQ2: Does incorporating those equity indicators into a decision-support model produce meaningfully different phaseout rankings than a model relying on distributional equity indicators alone?

RQ3: Which plants are ranked similarly across different shapes of justice, making them stronger candidates for robust policy action, and which plants are ranked differently, suggesting a need for further stakeholder deliberation?

Organization of this Thesis

These research questions are addressed across the remaining chapters. Chapter 2 presents the analytical architecture for integrating equity more broadly into a decision-support model. Chapter

3 presents the phaseout rankings of 96 California natural gas power plants under utilitarianism, sufficientarianism, and egalitarianism, examines ranking variability and its causes through global sensitivity analysis, and compares the dissertation's rankings with the CPUC reference. Chapter 4 interprets these findings in relation to all three research questions, evaluates the LLM-based weight extraction method, discusses limitations, and identifies directions for future work.

In addition, this thesis contains several appendices that provide detailed results. These are excluded from the main body to improve clarity and conciseness. Appendix A lists indicators considered during the scoping literature review but not included in the final model. Appendix B reports the cross-model consistency of LLM-based weight assignments to indicators and metrics. Appendix C reports the cross-model consistency of equity component classifications of indicators and metrics. Appendix D presents the top 20 powerplants by median rank for early phaseout under each shape of justice. Appendix E documents the indicator direction and normalization treatment used in the model.

Chapter 2. Methods

This chapter describes the analytical architecture developed in this dissertation to incorporate broader definitions and concepts of equity into decision-making models than is currently possible.

The analytical architecture (Figure 2) defines the relationships among equity-relevant constructs and organizes them into a sequence of modeling steps. The analytical architecture is not causal or deterministic (Jabareen 2009). It serves as a guide for organizing equity-relevant constructs within an analytical model. Once specific variables and data are defined for a given use case, the analytical architecture supports the development of an equity-integrated model.

As shown in Figure 2, the process begins by selecting a shape of justice, which defines how stakeholder preferences are represented in the decision process. Next, decision-relevant indicators and metrics are identified. Indicators are policy-specific factors important for making the decision under consideration and are also used to assess equity outcomes under the selected shape of justice. Metrics aggregate indicators that encode similar dimensions. The indicators and metrics are then combined with a decision-making model to evaluate equity outcomes under the selected shape of justice. The output is used to inform decision-making and prioritize actions. The entire process is repeated for each shape of justice, allowing comparison of outcomes under different resource allocation principles.

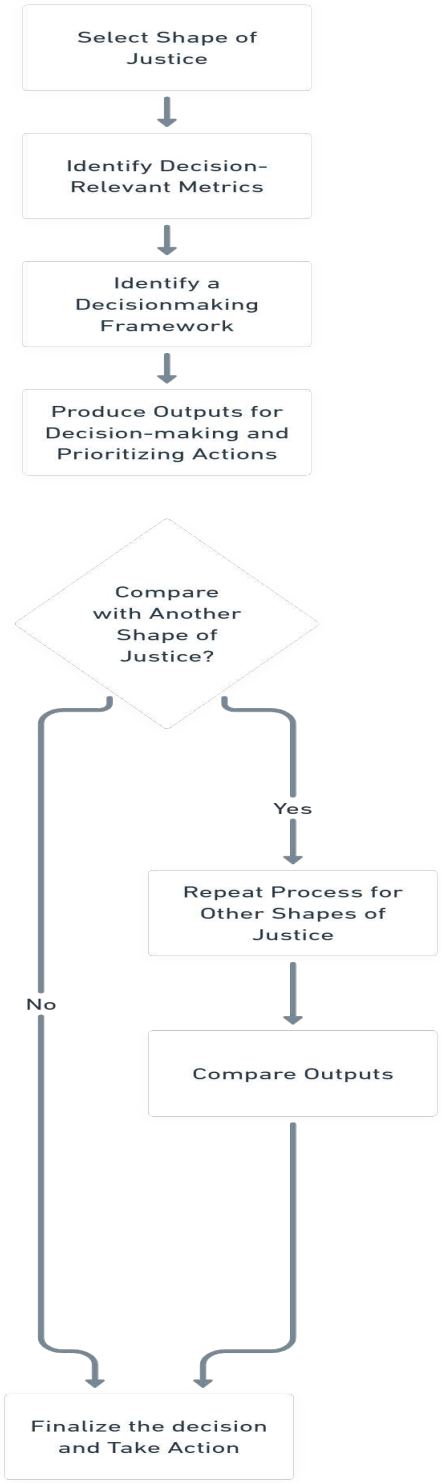
The shapes of justice affect the model through weights. Each shape of justice assigns different levels of importance to the same indicators and metrics, which changes the composite score and the resulting phaseout ranking. The three equity components enter the architecture as a classification layer: each indicator and metric is tagged as distributional, contextual, or procedural. This classification is used to assess whether the model covers the relevant dimensions of equity and to interpret which equity components drive ranking differences or rank variability. In this implementation, the classification does not directly change the composite score; the score changes through the weights assigned under each shape of justice.

To instantiate this architecture for the California case study, this chapter identifies 16 indicators and 6 metrics through a scoping literature review, describes the MCDM model including normalization and composite score construction, extracts indicator and metric weights under three shapes of justice using an LLM-based pipeline, classifies the indicators and metrics by equity component, and introduces the CPUC Gas Capacity Not Retained ranking as a planning-based reference point.

The remainder of this chapter elaborates on this architecture. Section 2.1 further describes the shapes of justice introduced in Chapter 1. Section 2.2 instantiates the analytical architecture for the California natural gas power plant phaseout case by specifying the MCDM model, identifying the indicators and metrics, describing the normalization and aggregation, and explaining the LLM-

based weight-assignment procedure. Section 2.3 describes how indicators and metrics are classified into distributional, contextual, and procedural equity components. Section 2.4 explains how composite scores are translated into phaseout rankings and how ranking robustness is assessed through global sensitivity analysis. Section 2.5 introduces the CPUC Gas Capacity Not Retained ranking as a planning-based reference point.

Figure 2: Analytical Architecture for incorporating equity into decision-making



REF: Developed by the Author

2.1 Shapes of Justice

As discussed in Chapter 1, stakeholders often hold different views on what constitutes an equitable or ethical outcome. These differences lead to different preferences for how indicators and metrics should be weighed in a decision-making model. The term "shape of justice" refers to a normative ethical principle, derived from moral premises about what ought to be (Zimm et al., 2024). This study uses it as a proxy for reflecting how stakeholders prioritize different indicators and metrics based on their interests and preferences. Adopting a specific shape of justice establishes a standardized basis for comparing outcomes without making purely subjective judgments about whose interests carry more weight. Three shapes of justice (utilitarianism, egalitarianism, and sufficientarianism) are used in this dissertation, each prioritizing different metrics and indicators and encoding distinct stakeholder values and perspectives in evaluating the NG powerplant phaseout.

Utilitarianism is defined as maximizing total welfare according to the policy goal. Sen (1979) defines utilitarianism as the principle that a state of affairs is better if and only if the sum total of individual utilities is at least as large as that in any alternative. Piacquadio (2017) extends this by noting that society uses utility functions to aggregate individuals' conflicting interests and to measure social welfare. Ferranna and Fleurbaey (2020) similarly describe the standard utilitarian social welfare function as one that adds up the expected well-being of all individuals to evaluate policies.

Sufficientarianism is defined as a pattern of justice in which priority is given to ensuring that basic or decent levels of needs are met by providing a threshold of goods and services. Zimm et al. (2024) describe this as a pattern in which a minimum threshold of goods and services is provided to ensure that everyone meets basic human needs. Once that threshold is met, sufficientarianism does not dictate how remaining resources should be distributed (Zimm et al., 2024). Frankfurt (1987) also argued that instead of focusing on equality, society's goal should be to "distribute the available resources in such a way that as many people as possible have enough or, in other words, to maximize the incidence of sufficiency". Roger Crisp (2003) builds on this by defining sufficientarianism through his "Sufficiency Principle" and "Compassion Principle," which prioritize those below a certain threshold of well-being. It should be noted that direct empirical literature applying sufficientarianism specifically to decarbonization-related planning remains limited, and the weighting scheme derived from this shape of justice in this dissertation reflects a reasoned interpretation of the principle rather than an established precedent.

Egalitarianism minimizes differences in how resources are obtained and ensures that everyone receives the same quantity. In this context, each community bears an even and proportionate share of the costs and benefits of the intervention. In its narrow form, this means ensuring that everyone receives the same quantity of a given metric, as Zimm et al. (2024) state, where "egalitarian patterns strive to minimize differences among people by making sure that everyone receives the same quantity" of a given metric, whether that is caloric intake, income, or overall utility. This

principle is frequently used to critique utilitarianism, which relies on sum-ranking and fails to account for cases in which one community bears all the costs while another receives all the benefits (Sen, 1979).

In the context of natural gas decommissioning, utilitarianism would seek to achieve plant closures at minimum overall direct cost, prioritizing phaseout sequences that maximize total welfare across all affected communities. Sufficientarianism would place the greatest weight on the concerns of the worst-off communities, ensuring that basic needs and minimum thresholds of well-being are met before considering broader efficiency gains. Egalitarianism would seek to ensure that each community bears an even and proportionate share of the costs and benefits of the phaseout, minimizing disparities in how the transition affects different populations.

Please note that, as mentioned earlier, the three shapes of justice are operationalized in this dissertation as distinct weight profiles applied to a common set of indicators and metrics, combined through a weighted sum (discussed in greater detail in Section 2.2.1). An unweighted sum is the functional form associated with utilitarian aggregation, in which a social objective is represented by adding up individual values (Adler 2024). Prioritarian principles correspond to a different functional form, in which values are combined through a strictly increasing and concave transformation before being summed, so that lower values carry greater weight (Adler 2024). A strict egalitarian rule (such as the maximin or leximin function) departs further from summation by ranking outcomes according to the worst-off position rather than the total (Sen 1979; Adler 2024).[5] Applying a weighted sum to all three shapes of justice therefore represents them through differences in weights rather than differences in aggregation form. This can be argued as applying a utilitarian aggregation logic to all three.

This is a deliberate modeling choice. The weighted additive value function is a standard aggregation method in multi-criteria decision analysis and is valid under the assumption of mutual preferential independence among criteria, meaning that the contribution of one criterion does not depend on the level of another (Dyer and Sarin 1979). This dissertation adopts that additive form and uses the shapes of justice to vary the relative importance assigned to indicators and metrics, consistent with the treatment of each shape of justice as a proxy for how stakeholders prioritize different criteria, as described earlier in this section.[6] As such, this dissertation adopts a narrow interpretation of the shapes of justice, treating them as proxies for different stakeholder values rather than as full social welfare functions. Representing the shapes of justice through aggregation

---

[5] Maximin: A rule that ranks outcomes according to the position of the worst-off, so the preferred outcome is the one that maximizes the minimum value.

Leximin: An extension of maximin that, when the worst-off positions are equal, compares the next-worst-off, and continues down the order to break ties

[6] A more direct representation of egalitarianism and sufficientarianism could instead embed each shape of justice in the indicators themselves, for example by including measures of inequality or dispersion across groups for egalitarianism, or indicators defined relative to a threshold for sufficientarianism. This dissertation does not take that approach. Instead, it assigns stronger weights to indicators that are close in spirit to each shape of justice.

forms that match their formal definitions would require specifying a utility function for each affected individual in the population, which is not feasible at the scale of this analysis.

2.2 Instantiating the Analytical Architecture for California Natural Gas Power Plant Phaseout

This section applies the analytical architecture to the phaseout of natural gas (NG) power plants in California. It explains how the requirements for that architecture are operationalized for the NG powerplant phaseout planning. Section 2.2.1 describes the decision-making framework selected for this dissertation: a multi-criteria decision-making (MCDM) model. Section 2.2.2 describes the scoping literature review used to identify candidate indicators and the exclusion criteria applied during indicator selection. Section 2.2.3 defines the six metrics used to group the 16 indicators. Sections 2.2.4 through 2.2.9 describe each metric and its associated indicators. Section 2.2.10 explains the normalization and aggregation process used to produce the final phaseout ranking.

2.2.1 MCDM Framing and Model Specification

MCDM is used for ranking alternatives when multiple, often conflicting criteria are involved (Rojas-Zerpa and Yusta 2015; Taherdoost and Madanchian 2023).[7] MCDM converts multiple indicators into a ranking by weighting and summing normalized indicator values into a composite score for each alternative.

MCDM is an appropriate method for NG decommissioning, which involves deciding how to order decommissioning 96 natural gas power plants in California based on multiple technical, economic, and equity indicators relevant to the decision (as described in Section 2.2.2). An indicator is a single measurable variable that captures a specific attribute relevant to the phaseout decision, such as SO2 emissions or energy burden. One method within MCDM that can be used is the Weighted Sum Model.[8] Here, after indicators are identified, the weights assigned to each indicator reflect their relative importance under a given shape of justice. As each shape of justice

[7] Multi-criteria decision-making (MCDM) is a branch of operational research concerned with finding optimal results in complex scenarios involving multiple conflicting objectives and criteria (Pohekar and Ramachandran 2004). Multi-criteria decision analysis (MCDA), while often used interchangeably, refers more specifically to the subset of MCDM methods that evaluate, compare, and rank discrete alternatives (Taherdoost and Madanchian 2023).

[8] The WSM was selected from among several available MCDA aggregation methods, such as the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon 1981), VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) (Opricovic and Tzeng 2002), Elimination Et Choix Traduisant la REalité (ELECTRE) (Roy 1968), and the Analytic Hierarchy Process (AHP) (Saaty 1980). The WSM is a popular MCDA aggregation method across multiple fields, including sustainable energy and sustainable development (Kandakoglu et al. 2019). Its transparency, wherein weights and criterion scores are explicitly defined and directly traceable to the final ranking, promotes transparency (Esangbedo and Che 2016). Moreover, as this dissertation also analyse how ranking outcomes change across multiple weight configurations, the WSM's direct linear relationship between weights and composite scores ensures that observed ranking variation is attributable to the weights themselves rather than to nonlinear properties of the aggregation method.

assigns importance differently, the WSM is applied separately for each shape of justice, producing three distinct sets of phaseout rankings from the same underlying data.

MCDM does not require organizing indicators into a structure suitable for weighted aggregation, but WSM does. Freudenberg (2003) describes a two-step weighting approach in which indicators are first aggregated within thematic sub-groups, and the resulting sub-indices are then weighed to produce the overall composite. This structure avoids overweighting components with more indicators. Freudenberg further notes that thematic grouping based on domain knowledge is a standard approach in composite indicator construction, particularly when indicators are drawn from heterogeneous data sources and are not expected to be statistically correlated. Pakkar (2016) demonstrates a similar hierarchical approach in which indicators with similar characteristics are grouped into subcategories, with weights assigned at each level. Cui et al. (2021) adopt the same approach for coal phaseout ranking, grouping plant-level variables into technical, economic, and environmental categories based on the domain each indicator measures. This dissertation follows the same two-tier structure, in which 16 indicators are organized into six corresponding metrics. A metric is defined as a thematic grouping of indicators that address similar constructs.[9]

As shown in Figure 3, the composite score for each plant is computed in two steps. In the first step, indicator values within each metric are combined using their corresponding indicator weights to produce a metric score. In the second step, the metric scores are combined using metric weights to produce a composite score for each plant.

The two-tier structure of MCDM ensures that each metric contributes in proportion to its assessed importance rather than the number of indicators it contains. A similar approach is used by Cui et al. (2021) in their plant-by-plant decision-making for coal phaseout.

For example, suppose one plant is evaluated using two metrics: Economic Impact and Environmental Impact. Economic Impact comprises two indicators: Total Employment (normalized value: 0.60) and Relative Compensation (normalized value: 0.80). If their weights are 2 and 1, the Economic Impact score is:

$$\frac{(2 \times 0.60) + (1 \times 0.80)}{2 + 1} = \frac{2.00}{3} = 0.67$$

Environmental Impact comprises two indicators: NOx Emissions (normalized value: 0.30) and CO_2 Emissions (normalized value: 0.50). If both weights are 3, the Environmental Impact score is:

[9] Freudenberg (2003) discusses factor analysis and cluster analysis as robustness checks in composite indicator construction. This dissertation groups indicators thematically based on domain knowledge, consistent with the approach used by Cui et al. (2021). The absence of a post-hoc statistical validation of these groupings is a limitation and a useful direction for future work

$$\frac{(3 \times 0.30) + (3 \times 0.50)}{3 + 3} = \frac{2.40}{6} = 0.40$$

If the metric weights are 2 for Economic Impact and 3 for Environmental Impact, the composite score is:

$$\frac{(2 \times 0.67) + (3 \times 0.40)}{2 + 3} = \frac{2.54}{5} = 0.51$$

The composite score of 0.51 represents the plant's weighted performance under this weight configuration. A higher composite score indicates a lower phaseout priority as higher normalized values indicate more desirable conditions. Plants are then ranked in ascending order of composite score, so the plant with the lowest composite score receives Rank 1.

However, note that this weighted linear aggregation approach relies on a standard simplifying assumption in MCDM. In MCDM, indicators are treated as mutually independent. In simple terms, this means that the contribution of one indicator to the composite score does not depend on the level of another indicator. This assumption allows the model to aggregate normalized indicator values transparently and consistently, but it may not capture interactions among indicators. For example, the effect of Energy Burden may differ depending on the Area Deprivation Index or the Disadvantaged Community status. Modeling such interactions is outside the scope of this dissertation, but remains an important direction for future methodological refinement

The equations make explicit what the model requires: (a) a set of indicators, (b) metrics that group related indicators, (c) indicator weights and metric weights for each shape of justice, and (d) normalized indicator values. Each requirement is addressed in the remaining subsections. The literature review and indicator selection procedure is described in Section 2.2.2. The six metrics are defined in Section 2.2.3, and the indicators grouped under each metric are described in Sections 2.2.4 through 2.2.9. Normalization and aggregation are described in Section 2.2.11. Indicator and metric weights are extracted through the LLM-based procedure described in Section 2.3.

2.2.2 Literature Review

The selection of phaseout-relevant indicators was informed by the scoping literature review (Tricco et al. 2018). The review covered environmental justice, energy equity, energy systems modeling, climate adaptation planning, fossil fuel phaseout, community resilience, and decarbonization. Literature was identified using Google Scholar and Web of Science, with search terms combining keywords such as energy justice, energy equity, just transition, fossil fuel phaseout, power plant retirement, environmental justice, community resilience, composite indicators, multi-criteria decision analysis, distributive justice, social welfare function, utilitarianism, egalitarianism, sufficientarianism, energy burden, and socioeconomic deprivation. Included sources were peer-reviewed journal articles and white literature from the U.S.

Department of Energy (DOE), the National Institute of Standards and Technology (NIST), the Organization for Economic Co-operation and Development (OECD), the International Energy Agency (IEA), the American Council for an Energy-Efficient Economy (ACEEE), and the National Bureau of Economic Research (NBER). General gray literature, such as news coverage or opinion pieces by recognized experts, was used only as complementary support for claims already grounded in academic or institutional sources.

The review prioritized literature published from 2020 onward, reflecting the recent growth of scholarship on energy equity, clean energy transitions, power plant retirement, and equity-centered modeling. Older foundational works on topics such as normative justice and welfare theory, composite indicator construction, and multi-criteria decision analysis were included where they provided core conceptual or methodological grounding. No formal review protocol was registered. The initial literature search for developing the analytical architecture yielded 58 sources, of which 46 were retained after title and abstract screening. The review was then conducted iteratively throughout the course of the dissertation rather than through a single structured search protocol and continued iteratively until additional sources no longer yielded distinct information. As such, for later phases of review, the exact number of sources screened at each stage was not recorded.

The specific literature supporting the inclusion of each final indicator is discussed in the corresponding indicator sections below. Two levels of exclusion criteria were applied during indicator selection. At the first level, where two candidate indicators reflected closely overlapping constructions, one was removed to avoid redundancy. At the second level, indicators such as minority employment in power generation, loss of retail and commercial activity, and job retraining capacity were excluded when suitable data were unavailable for California natural gas power plants and a suitable proxy could not be identified. The final set of 16 indicators was organized under six thematic metrics, which are defined in the subsequent section. The full candidate indicator list and the rationale for inclusion or exclusion are provided in Appendix A.

Indicator selection was applied at the indicator level before indicators were grouped into the six metrics described in Section 2.2.1. Selecting indicators under a data-availability constraint is precedented in literature. For example, The OECD/JRC Handbook states that a lack of relevant data limits which indicators can be included and that indicator selection requires compromises between what is desirable to measure and what the available data support (OECD and JRC 2008). Similarly, Apergi et al. (2024) also report limited availability of secondary data for measuring some elements of energy justice as a stated limitation of their index.

The two exclusion criteria used here remove indicators for different reasons and have different consequences for the model. Indicators removed for closely overlapping construction are represented by a retained indicator. As such their exclusion does not impact the results of the model . Indicators removed for lack of data are not represented by any retained indicator, so their exclusion removes a dimension that would have provided decision relevant signal. However, as the data for these indicators are absent, the resulting loss cannot be quantified, and the following discussion identifies which dimensions are affected.

As shown in Appendix A, some of the excluded indicators relate to fiscal effects on local governments, such as Loss of Tax Revenue and City Services. In the coal-closure context, Jolley et al. (2019) found that decommissioned plants stopped paying tangible personal property taxes, producing $8.5 million in lost tax revenue for local governments in Adams County, Ohio. Others reflect indirect economic impacts, such as Loss of Small Business and Loss of Retail or Commercial Employment, which correspond to the indirect and induced employment effects that Jolley et al. (2019) measured. Some, such as Loss of Culture, relate to community identity; Carley, Evans, and Konisky (2018) found that coal communities undergoing closures experience a social transition that involves reshaping their culture and sense of identity. The retained indicator set reflects the dimensions for which data are available at the plant or county level, and the excluded indicators pertaining to fiscal, indirect economic, and community-level impacts are areas the current model is less able to capture directly.

2.2.3 Indicators and Metrics

As shown in Table 1, the 16 indicators were organized into six metrics through a qualitative coding process in which indicators addressing similar themes were grouped together. The six metrics are defined as follows.

Table 1: Metrics and Associated Indicators used for NG Powerplant Phaseout Planning

Metric	Description	Associated Indicators
Economic Impact	Effects on jobs, income, local business activity, and fiscal/tax base linked to powerplant phaseout.	Total Employment by Powerplant, Location Quotient, Relative Compensation
Environmental Impact	Effects of powerplant phaseout on air pollutants, greenhouse gases, and related local environmental quality.	SO2 Emissions, NOx Emissions, CO2 Emissions
Community Impact	Availability and strength of local social, civic, and institutional support surrounding the powerplant.	Healthcare and Social Assistance; Religious, Grant-making, Civic, Professional, and Similar Organizations
Affordability	Costs of energy borne by households and the broader system in the vicinity of each plant.	Energy Burden, Average Annual Energy Cost, Levelized Cost of Energy (LCOE)
Plant Characteristics	Technical and operational attributes of the powerplant.	Capacity, Age of Plant, Heat Rate
Living Standards	Socioeconomic disadvantages and well-being status of communities surrounding the powerplant.	Disadvantaged Community, Area Deprivation Index.

Note: Metrics and indicators were identified through the scoping literature review described in Section 2.2.2. Each metric aggregates indicators that capture a related dimension of equity-relevant impact. The descriptions reflect how each metric is operationalized in the context of California's natural gas powerplant phaseout.

Table 2: Summary of Indicators used for NG Powerplant Phaseout Planning

Indicator	Definition	Data Source	Units (before normalization)	Equity Component
Total Employment	The approximate number of employees directly engaged in the operation, control, and maintenance of machinery at each natural gas power plant.	BLS OEWS (SOC 51-8013); capacity-to-employment ratio (Wagman)	Number of employees	Distributional
Job Share by Powerplant (Location Quotient)	A measure of the local concentration of power plant operator employment relative to the national average.	BLS OEWS (SOC 51-8013)	Ratio	Distributional|Contextual
Relative Compensation	The ratio of the annual mean wage of power plant operators in a county to the per capita income of California.	BLS OEWS; ACS 5-Year Estimates, Table B19301	Ratio	Distributional
SO_2 Emissions	The plant-level sulfur dioxide emissions associated with the air pollution burden of the plant.	EPA Clean Air Markets Division; Decarb My State	Tons	Distributional
NO_x Emissions	The plant-level nitrogen oxide emissions associated with	EPA Clean Air Markets Division; Decarb My State	Tons	Distributional

	the air pollution burden of the plant.			
CO_2 Emissions	The plant-level carbon dioxide emissions associated with the greenhouse gas contribution of the plant.	EPA Clean Air Markets Division; Decarb My State	Tons	Distributional
Health Care and Social Assistance	The number of health care and social assistance establishments in the county where the plant is located, used as a proxy for local support infrastructure.	Economic Census, NAICS Sector 62	Number of establishments	Contextual
Religious and Civic Organizations	The number of religious and civic organization establishments in the county where the plant is located, used as a proxy for local social capital.	U.S. Census Bureau, NAICS 81 (Other Services, except Public Administration)	Number of establishments	Contexual
Energy Burden	The percentage of gross household income spent on energy costs in the area surrounding the plant.	DOE LEAD Tool	Percentage of household income	Distributional
Average Annual Energy Cost	The total household expenditure on energy, including electricity, utility gas,	DOE LEAD Tool	Dollars	Distributional

	bottled gas, and other fuels, in the area surrounding the plant.			
Levelized Cost of Energy (LCOE)	The lifetime cost of generating electricity on a per-megawatt-hour basis, used to assess the cost of replacing the plant's generation.	NREL SLOPE dataset	Dollars per MWh	Distributional
Capacity	The generating capacity of each natural gas power plant.	Global Energy Monitor, Global Oil and Gas Plant Tracker	Megawatts (MW)	Distributional
Age of Plant	The age of the plant, calculated from its start year.	Global Energy Monitor, Global Oil and Gas Plant Tracker	Years	Distributional
Heat Rate	The thermal efficiency of the plant, expressed as the fuel energy required to generate one kilowatt-hour of electricity.	California Energy Commission, Quarterly Fuel and Energy Report (QFER)	Btu/kWh	Distributional
Disadvantaged Community	A binary locational indicator of whether the plant's census tract qualifies as a disadvantaged community under California's SB 535 framework.	CalEPA / OEHHA CalEnviroScreen 4.0	Score	Contextual
Area Deprivation Index	A composite measure of neighborhood-level socioeconomic	University of Wisconsin Neighborhood Atlas (ADI)	State decile (1–10)	Contextual

	disadvantage for the area surrounding the plant.			

Note: All indicators are normalized to a [0,1] scale, as described in 2.2.10. Note that the equity classifications are independent of the shape of justice. None of the models classified any indicator or metric as procedural equity

Table 3: Equity Component Classification of Metrics

Metric	Definition	Equity Component
Economic Impact	Effects on jobs, income, local business activity, and the fiscal and tax base linked to powerplant phaseout.	Distributional
Environmental Impacts	Effects of powerplant phaseout on air pollutants, greenhouse gases, and related local environmental quality.	Distributional
Community Impact	Availability and strength of local social, civic, and institutional support surrounding the powerplant.	Contextual
Affordability	Costs of energy borne by households and the broader system in the vicinity of each plant.	Distributional
Powerplant Characteristics	Technical and operational attributes of the powerplant.	Distributional
Living Standard	Socioeconomic disadvantages and well-being status of communities surrounding the powerplant.	Distributional

Note: Equity classifications are independent of the shape of justice. None of the models classified any indicator or metric as procedural equity.

The following subsections are organized around these six metrics to clarify how each indicator contributes to the metric it informs and, in turn, to the composite phaseout score. Each subsection introduces one metric and describes the indicators grouped under it, including the rationale for inclusion, the definition, and the data source for each indicator. Table 2 provides a summary of all 16 indicators, including their definitions, data sources, raw units before normalization, and equity component classification. Table 3 summarizes the six metrics, including their definition and equity component classification. The equity component classification described in detail in Section 2.3.

2.2.4 Metric: Economic Impact

Economic Impact captures the local economic consequences of powerplant phaseout for the surrounding community, specifically the effects on employment, income, and the local fiscal and tax base. The indicators grouped under this metric are Total Employment, Job Share by Powerplant (Location Quotient), and Relative Compensation.

Total Employment

Total Employment approximates the number of employees directly engaged in the operation, control, and maintenance of machinery at individual natural gas power plants in California. Its inclusion is supported by the literature, which highlights total employment as a primary consideration in phaseout strategies. Hanson (2022) identifies worker displacement from natural gas-related activities as among "the most immediate impacts of the energy transition," establishing workforce scale as a direct and measurable consequence of retirement decisions. Similarly, at the local level, Funderburg (2024) quantifies plant-level job losses as a primary economic impact of fossil-fuel plant closures, and Jolley et al. (2019) document the disproportionate regional burden associated with fossil-fuel employment.

Given that plant-level employment figures are not directly reported at the facility level, two complementary estimation rubrics were employed, and their outputs were cross-validated.

The primary estimation method draws on the U.S. Bureau of Labor Statistics (BLS) Occupational Employment and Wage Statistics (OEWS) data for May 2023, specifically the occupational category *Power Plant Operators* (SOC 51-8013) (U.S. Bureau of Labor Statistics 2024). It is defined by the BLS as workers who "control, operate, or maintain machinery to generate electric power," including auxiliary equipment operators, but explicitly excluding Nuclear Power Reactor Operators (SOC 51-8011). Since this data is reported at the Metropolitan Statistical Area (MSA) level rather than the facility level, MSA-level employment figures were first disaggregated to the county level. The total installed generating capacity across all natural gas plants in each county was then calculated, and total county-level employment was allocated to individual plants proportionally, based on each plant's share of the countywide generating capacity.

Where MSA-level data were unavailable, a secondary rubric was applied: a facility requiring approximately 565 MW of generating capacity employs around 27 full-time personnel (*Wagman 2017).* This capacity-to-employment ratio was used to estimate plant-level staffing from each facility's reported nameplate capacity. Both methods were tested against the same facilities to assess consistency. For the Carson CoGen Plant (56 MW), the BLS-based method yielded an estimated 2.60 employees, while the rubric-based method produced 2.67 ($56/565 \times 27$). For the CalPeak Power Enterprise Plant (59 MW), the respective estimates were 3.98 and 2.80. While minor discrepancies exist at some facilities, the Marsh Generating Station (207 MW) had estimates of 20 and 10 employees, respectively, and the Crockett Cogen Project (247 MW) had 23 and 12.

Job Share by Powerplant (Location Quotient)

The Location Quotient (LQ) measures a region's industrial specialization relative to a larger geographic unit, typically the nation (U.S. Bureau of Economic Analysis n.d.). It is computed as an industry's share of regional employment divided by that industry's share of national employment. An LQ greater than 1.0 indicates that the occupation is more concentrated in the local area than the national average, while an LQ less than 1.0 indicates below-average local concentration. The LQ is also an important indicator in phase-out decision-making, as it serves as

a proxy for assessing the localized significance of NG power plant employment. Gottlieb and Goetz (2020) identify the LQ as "a recognized measure of regional industry concentration," noting its advantages of straightforward calculation and interpretability for capturing spatial economic specialization. Jolley et al. (2019) operationalize the location quotient (LQ) to assess the impacts of fossil fuel plant closures in Appalachian Ohio.

In this study, the LQ for the occupational category *Power Plant Operators* (SOC 51-8013) is drawn from the BLS Occupational Employment and Wage Statistics for May 2023, reported at the MSA level (U.S. Bureau of Labor Statistics 2024). Since facility-level LQ data are not available, MSA-level values were disaggregated to the county level, and natural gas power plants located within the same county were assigned the same LQ value.

Relative Compensation

Relative Compensation is defined as the ratio of the annual mean wage of power plant operators in a given county to the per capita income of California, expressed as follows:

$$Relative\ Compensation = \frac{Annual\ Mean\ Wage\ of\ Powerplant\ Operators\ (County)}{Per\ Capita\ Income\ of\ California}$$

Relative Compensation is important for phaseout decision-making as it captures the wage premium that power-plant employment may provide in the local labor market. A plant that employs relatively few workers but pays wages substantially above the local per capita income represents a qualitatively different transition burden than one whose wages are closer to the regional norm. Jolley et al. (2019) found that displaced fossil fuel plant workers transitioning to occupations with comparable skill requirements nonetheless experienced wage decreases.

Per capita income for California, measured in inflation-adjusted dollars for the past 12 months (2022), was obtained from the American Community Survey (ACS) 5-Year Estimates (2023 release), Table *B19301 (U.S. Census Bureau 2022).*A ratio greater than 1.0 indicates that power-plant operator wages in that county exceed the statewide per capita income benchmark, suggesting that plant employment offers above-average compensation within the local economic context. A ratio below 1.0 indicates the reverse.

2.2.5 Metric: Environmental Impact

Environmental Impact captures the effects of powerplant phaseout on air pollutants, greenhouse gases, and related local environmental quality. The indicators grouped under this metric are SO2 Emissions, NOx Emissions, and CO2 Emissions.

SO2 Emissions

SO_2 emission is a well-established indicator for prioritizing fossil fuel plant retirement due to its direct contribution to fine particulate matter ($PM_{2.5}$) formation and the associated public health burden. Sergi et al. (2020) found that long-term exposure to $PM_{2.5}$ produced by SO_2 and NO_x

emissions is strongly linked to premature death and other adverse health outcomes, with the social costs of health effects attributable to U.S. power-sector emissions estimated at $60–130 billion annually. Campos Morales et al. (2024) show that the sequencing power plant retirements produces measurably different air-quality and health outcomes, depending on which emissions criteria are used to rank facilities. They further find that strategies prioritizing plants with higher SO_2 emissions and co-pollutants deliver larger reductions in pollution exposure and associated mortality. Burney (2020) similarly provides empirical evidence that decommissioning fossil-fuel plants can yield substantial air-quality improvements, supporting SO_2 emissions as a decision-relevant indicator for phaseout planning.

Plant-level SO_2 emissions were obtained from the U.S. Environmental Protection Agency (EPA) Clean Air Markets Division Annual Power Plant Emissions data for 2022, reported in tons (U.S. Environmental Protection Agency 2023). The EPA dataset is organized at the plant level, with underlying records collected across individual turbine units. When multiple turbines operate within a single plant, total plant-level emissions are distributed among turbines in proportion to their generating capacity.

For a small number of facilities for which EPA data were unavailable, such as the AltaGas Ripon Energy power station, SO_2 estimates were sourced from *Decarb My State n.d.* The same capacity-proportional distribution method was applied to this supplementary data.

NOx Emissions

Nitrogen oxide (NO_x) emissions are also a well-established indicator for prioritizing fossil fuel plant retirements. For example, Sergi et al. (2020) establish that $PM_{2.5}$ generated from NO_x and SO_2 emissions is strongly associated with premature mortality and adverse health outcomes, with power-sector-attributable health costs estimated at $60–130 billion annually. Declet-Barreto et al. (2022) also highlight that the burden of mortality and morbidity from pollutants such as NOx is well established and that equitable decarbonization policy must account for the distributional health benefits of NO_x reductions.

Plant-level NO_x emissions were obtained from the EPA Clean Air Markets Division Annual Power Plant Emissions data for 2022, reported in tons (U.S. Environmental Protection Agency 2023). Similar to SO_2 emissions, for power plants where EPA data were unavailable, data were sourced from Decarb My State (n.d.).

CO2 Emissions

CO2 emissions are an important decision-making indicator for the phaseout of NG power plants. For example, Pollin and Callaci (2019) identified that the CO_2 from fossil fuel combustion accounts for approximately 75 percent of all greenhouse gas emissions, driving climate change. Sergi et al. (2020) framed power plant replacement as an optimization problem structured around meeting a CO_2 reduction target, and Fofrich et al. (2020) extend this logic to retirement scheduling, demonstrating that achieving rapid emissions reductions requires the early retirement of natural

gas-fired power plants. Cui et al. (2021) also incorporated CO_2 emission rate as one of three core criteria in their multi-criteria retirement ranking algorithm, alongside technical attributes and profitability. Plant-level CO_2 emissions were obtained from the EPA Clean Air Markets Division Annual Power Plant Emissions data for 2022, reported in tons (U.S. Environmental Protection Agency 2023). The dataset is structured at the plant level. For power plants where EPA data were unavailable, data was sourced from Decarb My State (n.d.).

2.2.6 Metric: Community Impact

Community Impact captures the availability and strength of local social, civic, and institutional support surrounding the powerplant. The indicators grouped under this metric are Healthcare and Social Assistance, and Religious and Civic Organizations.

Healthcare and Social Assistance

This indicator captures the number of health care and social assistance establishments in the county where a natural gas power plant is located. It is used as a proxy for the local support infrastructure available to workers and residents during plant retirement. Community resilience research supports its inclusion by showing that health care and social assistance capacity is an important factor in how communities absorb and recover from economic shocks. For example, Burton (2015), drawing on Hurricane Katrina recovery data, identifies health services and social assistance programs per capita as core components of social resilience, arguing that communities providing their residents with access to health and well-being support exhibit greater capacity to withstand pre-disaster impacts and sustain post-disaster recovery. This observation is directly applicable to the economic disruption associated with the NG power plant phaseout. Gu et al. (2023) validate this within community resilience measurement frameworks, identifying employment in the health care and social assistance sector as a commonly used and empirically validated proxy for community resilience capacity across multiple established frameworks. This validation supports the use of establishment counts in this sector as a reliable indicator of local support capacity. Carley, Evans, and Konisky (2018), examining energy transition in American coal communities, argue that government intervention must extend beyond environmental regulation to address health, education, and public services in fossil fuel-dependent regions.

The data are drawn from the 2022 Economic Census, NAICS Sector 62 (Health Care and Social Assistance), reported at the county level (U.S. Census Bureau, 2022; U.S. Census Bureau 2026). The unit of measurement is the number of establishments. Because the data are reported at the county level, all power plants located within the same county are assigned the same value.

Religious and Civic Organizations

This indicator measures the number of religious and civic organization establishments in the county where a natural gas power plant is located. It serves as a proxy for local social capital and

community organizational capacity during an energy transition. Prior research supports using such organizations as indicators of a community's ability to manage and recover from the economic disruption associated with power plant retirement. Edgemon et al. (2020) identify affiliation with religious and civic organizations as a proxy measure for social connectedness and the degree to which a community can rely on reciprocity and cooperative action among residents. Burton (2015) validates religious and civic organizations as indicators of community capital and social resilience, identifying religious, social advocacy, and civic organizations per capita as distinct components of a community's capacity to sustain resilience before and after disruptive events. Carley, Evans, and Konisky (2018) further provide support from coal community transitions, highlighting that successful adaptation in energy-impacted communities involved coordinated action across community groups, nonprofits, universities, and local *businesses.*

Ideally, the data for this indicator would have matched NAICS 813 (Religious, Grantmaking, Civic, Professional, and Similar Organizations). Because county-level NAICS 813 data are frequently suppressed for confidentiality reasons (Isserman and Westervelt, 2006), this study uses NAICS 81 (Other Services, except Public Administration) as the closest consistently available proxy at the county level (U.S. Census Bureau 2022).

2.2.7 Metric: Affordability

Affordability captures the costs of energy borne by households and the broader system in the vicinity of each plant. The indicators grouped under this metric are Energy Burden, Average Annual Energy Cost, and Levelized Cost of Energy (LCOE).

Energy Burden

Energy burden is defined as the percentage of gross household income spent on energy costs, calculated as the ratio of average household energy expenditure to average annual household income (Heleno et al., 2022; Parker et al., 2023). A household energy burden at or above 6% is classified as high energy burden, a threshold operationalized by the U.S. Department of Energy's Low-Income Energy Affordability Data (LEAD) Tool, which reports a national average energy burden of 3% (Heleno et al., 2022).

Energy burden is relevant to phaseout planning because retiring a natural gas power plant can reduce local generation capacity, which may increase electricity prices for households in the surrounding area. Households that already spend a high share of their income on energy are more vulnerable to these price increases. Bohr and McCreery (2020) establish that energy burden is an important dimension of household economic vulnerability, noting that policies internalizing the environmental costs of energy production may increase prices and leave some households vulnerable. Drehobl, Ross, and Ayala (2020) document that this burden is not uniformly distributed, with low-income, Black, Hispanic, Native American, renter, and older adult households carrying disproportionately higher energy burdens. Heleno et al. (2022) operationalize

energy burden within a quantitative decision-support model for energy justice, using the 6% threshold as a primary eligibility criterion for targeted energy policy interventions.

Energy burden data were obtained from the DOE LEAD Tool, accessed in October 2024, at the address level, with the base layer data structured at the census tract level (U.S. Department of Energy n.d.). For power plants where the LEAD Tool returned no data, likely due to an absence of residential households in the immediate vicinity, a value of zero was assigned. This assignment reflects the absence of a residential energy burden exposure rather than a data gap, and is consistent with the LEAD Tool methodology, which suppresses estimates where the underlying household sample is insufficient (U.S. Department of Energy 2024).

Average Annual Energy Cost

Average annual energy cost is defined as the total household expenditure on energy, encompassing electricity, utility gas, bottled gas, and other fuels, including fuel oil and wood. It is a direct measure of the absolute dollar cost of energy borne by households, distinct from energy burden, which expresses that cost as a proportion of income.

The average annual energy cost is relevant to phase-out planning because clean energy transition policies have direct household cost implications, particularly for lower-income households that cannot readily switch away from fossil fuels (International Energy Agency 2024). Areas where households already face high absolute energy costs are more exposed to these cost changes. Parker et al. (2023) identify cost per kilowatt-hour and related expenditure metrics as primary affordability indicators for assessing whether electric services are provided at a cost that promotes universal service. Average annual energy cost provides a direct measure of where that exposure is concentrated.

Energy burden and average annual energy cost are retained as separate indicators because they capture different dimensions of affordability. Energy burden measures relative affordability as the share of income spent on energy costs and is sensitive to income levels. Average annual energy cost measures absolute expenditure and is sensitive to consumption levels, housing characteristics, and local energy prices. A community with a moderate energy burden but high absolute costs face a different risk than one with a high burden and low absolute costs.

Data were obtained from the DOE LEAD Tool at the address level, with the base layer structured at the census tract level, accessed in October 2024 (U.S. Department of Energy, 2024). For power plants for which the LEAD Tool returned no data, a value of zero was assigned, indicating the absence of a residential population in the immediate vicinity.

Levelized Cost of Energy (LCOE)

Levelized Cost of Energy (LCOE) is a standard metric that expresses the lifetime cost of generating electricity from a given source on a per-megawatt-hour basis, enabling cost comparisons across technologies and locations on a consistent footing.

LCOE is important for phaseout planning because it determines how economical it would be to replace the retiring plant's generation with new capacity in the same location. Parker et al. (2023) classify LCOE as a necessary system-level economic criterion in any planning approach that includes affordability as a core objective, noting that it should be complemented by household-level metrics to capture distributional impacts on specific communities. Carley, Evans, and Konisky (2018) demonstrate this by mapping minimum LCOE by technology across U.S. counties, noting that because power plants are built to last 30 to 40 years, utilities rely on regional LCOE projections and price differentials when prioritizing investments in new facilities. Fofrich et al. (2020) reinforce this, identifying the levelized cost of new generating infrastructure as a primary economic determinant for deciding when to retire existing fossil fuel generators.

County-level LCOE values were derived from the National Renewable Energy Laboratory (NREL) State and Local Planning for Energy (SLOPE) dataset, which provides Median Total LCOE projections in dollars per MWh at the county level. NREL projections for 2020 and 2030 were used to derive a linear interpolation equation in Excel, which was then applied to estimate county-level LCOE values for 2024. All power plants within a given county were assigned the same projected LCOE value. A later version of the NREL SLOPE dataset no longer includes the Total Median LCOE projection series or provides the data in a comparable format. Accordingly, this analysis relies on the earlier version of the dataset, which supported this projection approach.

2.2.8 Metric: Plant Characteristics

Plant Characteristics captures the technical and operational attributes of the powerplant. The indicators grouped under this metric are Capacity, Age of Plant, and Heat Rate. The list of active natural gas power plants in California was obtained from the Global Energy Monitor (GEM) Global Oil and Gas Plant Tracker (GOGPT) in the February 2024 release (Global Energy Monitor n.d.). The dataset was filtered to natural gas units in California listed as operating, recording attributes such as capacity, location, status, owner, and start-up year, yielding a list of 96 NG power plants. GOGPT provides both plant-level and unit-level records; this dissertation used plant-level data.

Capacity

Plant capacity, measured in megawatts (MW), represents the generating capacity of each natural gas power plant. Values were obtained directly from the Global Energy Monitor dataset and required no additional derivation.

Plant capacity directly reflects both the contribution a plant makes to the regional grid and the scale of the economic and reliability consequences associated with its retirement. From a reliability perspective, the U.S. Department of Energy (2024) identifies natural gas generators as a historically important source of firm capacity in resource adequacy planning. Similarly, Schleifer et al. (2023) argue that public utility commissions evaluating fossil plant retirements must assess

whether resource adequacy can be maintained after closure. From a community and labor market perspective, Burke, Best, and Jotzo (2019) find that closures involving cumulative capacity above 180 MW are associated with statistically significant local unemployment effects, increasing the local unemployment rate by about 0.9 percentage points. Cui et al. (2021) also treats plant size as a relevant decision factor in plant-level retirement analysis, including it as one of eight criteria in their ranking algorithm and grouping it with technical attributes such as age and combustion technology.

Age of Plant

Plant age was calculated by subtracting the plant's start year, as recorded in the Global Energy Monitor dataset, from 2024. Age of Plant is a recognized criterion in phaseout planning. Cui et al. (2021) include it as one of eight criteria in their plant-by-plant retirement ranking algorithm for Chinese coal capacity, classifying it alongside plant size and combustion technology under the technical attributes dimension. Burke, Best, and Jotzo (2019) provide additional support, reporting that the median age at closure for retired coal-fired power stations was 43 years, with many remaining plants nearing the end of their operational lifetimes. Nyberg (2020) provides California-specific grounding, documenting that aging natural gas plants in the state averaged a heat rate of 13,212 Btu/kWh while holding 16 percent of gas-fired capacity but producing only 3 percent of generation, identifying older plants as the least thermally efficient units in the fleet.

Heat Rate

Heat rate (Btu/kWh) expresses the thermal efficiency of a power plant as the quantity of fuel energy, in British thermal units (Btu), required to generate one kilowatt-hour (kWh) of electricity. Lower heat rates indicate greater fuel efficiency, while higher heat rates indicate greater fuel consumption per unit of output (Energy Information Administration 2024). Literature has provided support for including this indicator. Cui et al. (2021) include heat rate as one of the unit-level variables in their plant-by-plant retirement ranking algorithm for Chinese coal capacity, alongside capacity, vintage year, combustion technology, and location. Nyberg (2020) further documents that aging natural gas plants in California carry the highest heat rates in the state fleet, averaging 13,212 Btu/kWh, and that cycling and ramping demands from increasing renewable penetration further degrade thermal efficiency over time. Plants with persistently high heat rates consume more fuel per unit of output, resulting in higher operating costs and greater emissions intensity. Heat rate data for 2022 were obtained from the California Energy Commission Quarterly Fuel and Energy Report (QFER 2) (California Energy Commission n.d.).

2.2.9 Metric: Living Standards

Living Standards captures the socioeconomic disadvantage and well-being status of communities surrounding the powerplant. The indicators grouped under this metric are the Disadvantaged Community and Area Deprivation Index.

Disadvantaged Community

Disadvantaged Community (DAC) status is a binary locational indicator derived from California's SB 535 designation framework, administered by the California Environmental Protection Agency (CalEPA). As of May 2022, a census tract qualifies as a DAC if it meets any one of four criteria: it falls within the top 25 percent of CalEnviroScreen 4.0 (CES 4.0) scores; it ranks among the worst 5 percent for pollution burden but lacks a complete CES score; it was previously designated as a DAC; or it constitutes tribally controlled land (California Office of Environmental Health Hazard Assessment n.d.).

Literature provides strong support for the indicator. Tarekegne et al. (2022) documented that natural gas Peaker plants are disproportionately located near disadvantaged communities, and that retirement and replacement decisions therefore carry equity implications that must be accounted for in prioritization. Specht and Wisland 2018 further found that as many as 23 natural gas plants located in disadvantaged communities can be retired by 2030, with a positive effect on air quality in communities with persistent pollution burdens. Goforth et al. (2023) situate this within the broader energy justice literature and argue that awareness of which communities carry cumulative air pollution burdens should inform which communities have their environmental burden reduced first. This directly supports the use of CES-based DAC status as an important indicator for phaseout decision-making.

For this study, CES 4.0 scores were collected from the official ArcGIS dashboard by entering each power plant's address and recording the corresponding tract-level result. Decile ranges were converted to single integers by dividing the upper bound of each range by 10 (e.g., 90 to 100 becomes 10, 80 to 90 becomes 9). Where the map suppressed a tract due to sparse data, the highest decile value was assigned to avoid understating potential disadvantage in data-limited areas.

Area Deprivation Index

The Area Deprivation Index (ADI) is a composite measure of neighborhood-level socioeconomic disadvantage developed by the University of Wisconsin Neighborhood Atlas (Kind and Buckingham, 2018). It combines 17 census variables spanning income, education, employment, and housing quality into a single score for every U.S. census block group. ADI values are published as state-level deciles from 1 (least disadvantaged) to 10 (most disadvantaged). It combines 17 census variables spanning income, education, employment, and housing quality into a single score for every U.S. census block group. ADI values are published as state-level deciles from 1 (least disadvantaged) to 10 (most disadvantaged).

ADI is relevant to phaseout decision-making because it identifies communities with structural socioeconomic disadvantage that may be less equipped to absorb the economic disruption of plant closure. Trinidad et al. (2022) show that indices such as ADI are actively used to direct public resource allocation toward high-need communities, citing the Centers for Medicare & Medicaid Services' adoption of these indices to adjust reimbursement rates under equity-focused care models. Martenies et al. (2019) support this in the fossil fuel retirement context, using

socioeconomic disadvantage variables that constitute the core components of the ADI, including median income, educational attainment, and employment rate, in their environmental justice analysis of coal plant retirements in Colorado. DAC and ADI are retained as separate indicators because they operationalize distinct constructs: DAC status reflects cumulative pollution and environmental burden, while ADI captures structural socioeconomic disadvantage independent of environmental exposure.

This dissertation uses the 2022 ADI release (v4.0.1), derived from the 2018 to 2022 American Community Survey five-year estimates mapped to 2022 census block group boundaries. Each plant was assigned to its county's average ADI score, so all plants within the same county share one value. Block groups suppressed by the Atlas due to fewer than 100 residents or 30 housing units were assigned the maximum disadvantage score of 10 to avoid understating deprivation in data-limited areas

2.2.10 Normalization and Aggregation

Once the 16 indicators were defined and the data collected, the indicator values were normalized and aggregated to a common plant-level representation before entering the MCDM model. This subsection describes both procedures.

Normalization. Before normalization, each indicator was assigned a direction based on whether a higher raw value represented a more desirable or less desirable condition for phaseout planning.[10] All indicators were then transformed to a common [0,1] scale using one rule: higher normalized values represent more desirable plant-level conditions. Indicators where higher raw values were desirable, such as Health Care and Social Assistance, were normalized directly. Indicators where higher raw values were undesirable, such as SO_2 Emissions, NO_x Emissions, CO_2 Emissions, Energy Burden, Heat Rate, and Age of Plant, were reverse coded before normalization. This direction-setting step is necessary because the composite score combines indicators with different substantive meanings. After this transformation, a higher composite score indicates a more desirable plant-level profile and, therefore, a lower phaseout prior. Further details are presented in Appendix E

Aggregation. Indicator data were then aggregated to produce a single plant-level representation for each facility. Capacity and total employment were aggregated as sums. Binary or categorical indicators, such as disadvantaged community status, were aggregated using the maximum value across units. All remaining indicators were aggregated using capacity-weighted means. After aggregation, plant-level values were renormalized to a [0,1] scale. This dataset of indicator data for 96 power plants in California served as input to the MCDM model.

---

[10] Please note that the indicator directions where higher value is better is an analytical assumption. For example, in this dissertation a higher Total Employment value is assigned to be better because it reflects the potential loss of local jobs after phaseout. This may not hold uniformly across communities, as a higher powerplant employment in a wealthier community with different alternative job options may matter less than in a community where the powerplant is one of the few employers.

2.2.11 Weight Assignment to Metrics and Indicators

The remaining inputs required by the MCDM model are the weights for indicator and metric under each shape of justice. The weights for the metrics and indicators were identified using a large language model (LLM) pipeline. The objective was to identify weights that are either explicitly supported by existing scholarship or, where such support is unavailable, derived through structured theoretical reasoning consistent with the shape of justice under consideration.

Note that another approach to weight assignment would be expert or stakeholder elicitation. However, eliciting weights for 16 indicators and six metrics across three shapes of justice creates a substantial scaling challenge, particularly when the goal is to preserve uncertainty rather than force consensus. The value of preserving rather than collapsing that uncertainty is well established across the decision-analysis literature.

In Multi-Criteria Decision Analysis (MCDA), uncertainty is unavoidable because the criteria weights and the performance scores of alternatives are frequently elicited from stakeholders rather than objectively measured (Broekhuizen et al. 2015). Briggs et al. (2012) identified four types of uncertainty which were then adapted by Broekhuizen et al. (2015) for multi criteria decision making. These are:

a) Stochastic uncertainty: Refers to the random variability in criteria weights when they are assigned by otherwise identical individuals.
b) Parameter uncertainty: Represents the estimation error of a specific quantity, such as the mean weight assigned to a criterion by a group of decision-makers.
c) Heterogeneity: Explainable variation in weights that can be attributed to the different background characteristics of the people providing the weights.
d) Structural uncertainty: Arises when decision-makers are unsure whether all the relevant decision criteria have been included in the model, or if the criteria are structured and transformed properly.

Broekhuizen et al. (2015), also, identify five primary approaches used in the MCDA literature to handle this uncertainty. This includes fuzzy set theory, deterministic and probabilistic sensitivity analysis, Bayesian frameworks, and grey theory.

A common principle across adjacent fields is to represent the full spread of plausible inputs instead of reducing them to a single estimate. In climate projection, multi-model ensembles are utilized to investigate structural uncertainty by aggregating projections from multiple models rather than relying on a single definitive outcome. However, rather than being strictly independent, these collections function as 'ensembles of opportunity' that frequently share underlying code and correlated biases, which complicates how the resulting spread of uncertainty should be interpreted (Tebaldi and Knutti 2007; Knutti et al. 2009; Parker 2013). This also applies to structured expert elicitation as well. If consensus is not the objective, the spread of individual expert estimates is retained and reported to convey inter-expert uncertainty (Morgan 2014; Hemming et al. 2018; O'Hagan 2019). Within Multi-Criteria Decision Making (MCDM) specifically, Stochastic

Multicriteria Acceptability Analysis (SMAA) operationalizes this by exploring the full space of feasible weights and reporting how alternatives rank across it, rather than requiring decision-makers to agree on a single weight vector (Lahdelma and Salminen 2010).

This principle is also reflected in the broader literature on decision making under deep uncertainty (DMDU). Lempert, Popper, and Bankes (2002) demonstrate that an ensemble of plausible scenarios provides a more representative characterization of available information than any single best estimate or probabilistic forecast. Lempert (2025) also argues that robust decision-making should consider a wide range of plausible futures and seek policies that perform well across them. Extending it to MCDA applications, Beutler et al. (2024) find that collapsing diverse preferences into equal weights obscured meaningful differences between decision options, whereas retaining the full range of stakeholder preferences allowed clearer discrimination among alternatives.

Therefore, this dissertation retains uncertainty in the weight assignment. Direct elicitation was attempted but proved impractical at this scale. The LLM-based pipeline is therefore used as a proof-of-concept extraction method, not as a substitute for stakeholder judgment. Section 4.4.1 returns to this issue in greater detail.

Background on LLM-RAG. A large language model (LLM) is a neural network built on the transformer architecture (Vaswani et al. 2017) that uses a mechanism called self-attention to weigh the relevance of each word in a passage against every other word when constructing a representation of meaning. Unlike earlier architectures that processed text sequentially, transformers process all positions in a sequence in parallel, allowing them to capture long-range dependencies within a text. By scaling up model capacity and training on massive text corpora, these models exhibit emergent abilities such as in-context learning. Their specific ability to follow instructions and interact safely with users is developed through subsequent alignment stages, such as supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF) (Huang et al., 2024). However, because LLMs generate output based on statistical patterns learned during training, they can produce content that is fluent and syntactically correct but factually inaccurate or unsupported by evidence, a phenomenon known as hallucination (Huang et al., 2024).

Hallucination is a particular concern in applications that require factual grounding, such as the extraction of normative weights from academic literature in this dissertation. Retrieval-augmented generation (RAG) addresses this limitation by augmenting the model's input with relevant text passages retrieved from an external document corpus at query time (Huang et al., 2024). This conditions the model's output on specific source documents rather than on its parametric knowledge alone (Lewis et al., 2020). Lewis et al. (2020) introduced the RAG framework and demonstrated that models augmented with retrieved passages produce more specific, diverse, and factual outputs than models relying solely on their parameters.

Gao et al. (2024) provides a comprehensive survey of RAG architecture and confirms that retrieval-augmented generation enhances the accuracy and credibility of generated outputs, particularly for knowledge-intensive tasks. The retrieval component of a RAG pipeline relies on

converting text into numerical vectors, called embeddings, that represent the semantic content of a passage in a continuous vector space. Embedding assumes that semantic relationships between linguistic units can be captured as geometric relationships between vectors, where words or passages with similar meanings will be mapped to nearby points in the vector space (Mikolov et al. 2013). Reimers and Gurevych (2019) extended this principle from individual words to sentences and paragraphs by developing Sentence-BERT, a transformer-based model that produces fixed-length sentence embeddings optimized so that semantically similar texts have high cosine similarity. Cosine similarity measures the cosine of the angle between two vectors, producing a value between 0 and 1, and is the standard metric for comparing document representations in information retrieval.

In a RAG pipeline, this means that when a query is submitted, the system identifies the most semantically relevant passages from the corpus by comparing the query's embedding against all stored document embeddings and returning those with the highest cosine similarity scores. The retrieved passages are then provided to the LLM as context for the generation task. The use of LLMs for structured information extraction from scientific text is supported by a growing body of methodological literature (Dagdelen et al. 2024; Shamsabadi, D'Souza, and Auer 2024; Polak et al. 2024). This study leverages LLMs to extract normative weights from the academic literature identified in Section 2.2.2.

Note that for indicators and metrics where the literature supporting the RAG pipeline is sparse or unavailable, the models may default to their base training. This may produce plausible weight assignments that are not verifiable against the corpus for hallucinations. The priority-based assignment protocol described in Section 2.2.11 requires each model to tag whether a given weight assignment is supported by the retrieved literature (Priority 1) or derived from the model's own reasoning (Priority 2), providing an auditable record of evidentiary grounding for each output.

Literature retrieval. The academic literature identified in Section 2.2.2 was processed with PyMuPDF (Welcome to PyMuPDF, 2026) to extract text from PDF documents. The extracted text was segmented into semantically coherent chunks at the paragraph and section level to preserve the argumentative structure of each source. Each chunk was then embedded as a 384-dimensional vector using the sentence-transformers/all-MiniLM-L6-v2 model (Sentence Transformers 2026). This model is a variant of the Sentence-BERT architecture **(**Reimers and Gurevych, 2019**)**, which uses a Siamese network to produce embeddings optimized for semantic similarity tasks. The resulting vectors were indexed using FAISS (Facebook AI Similarity Search), an open-source library for efficient similarity search over dense vector collections (Douze et al., 2024). For each weighting task, a retrieval query was constructed from the indicator or metric definition and embedded using the same model. The query embedding was compared against all indexed literature chunks using cosine similarity (Salton and McGill 1983)[11], and the top 10 most similar

[11] Several families of similarity metrics exist for comparing LLM outputs, each suited to different output types. Cosine similarity, used in this dissertation's retrieval step, measures the geometric proximity of two vector representations in

chunks were retrieved. It is important to note that cosine similarity is used here solely as the retrieval mechanism; it determines which passages from the literature are provided to the model as context. It is not used to evaluate the model's outputs, which are categorical weight assignments (low, medium, high) assessed through inter-rater reliability statistics (see Appendix B). Only chunks retrieved through this pipeline were made available to the models; no additional context or external information was provided. The same retrieval query was used across all three shapes of justice to ensure that differences in extracted weights are attributable to the justice principle rather than to variation in the retrieved literature.

Note that the retrieval corpus is the same set of academic literature identified through the scoping review in Section 2.2.2. The same corpus was used because the scoping review was designed to capture the literature relevant to the indicators and metrics being weighted, and the weight-extraction task requires the models to draw on that same body of evidence. No documents outside this set were added to the retrieval corpus, so the weights and classifications are grounded only in the literature identified in Section 2.2.2.

Model selection. Weights were extracted using three open-source LLMs accessed via the Hugging Face model repository: Mistral-7B-Instruct-v0.3 (Jiang et al. 2023), Gemma-2-9b-it (Gemma Team et al. 2024), and Qwen2.5-7B-Instruct (Yang et al. 2024). All models were loaded using the Hugging Face Transformers library (Wolf et al. 2019) and run on Google Colab with 4-bit quantization to fit within the available GPU memory. All three are dense transformer models in the 7–9 billion parameter range, which ensures that differences in outputs are not attributable to large disparities in model capacity. At the same time, they originate from three independent development organizations (Mistral-7B-Instruct-v0.3 is from Mistral AI, Gemma-2-9b-it is from

---

embedding space and is suited for comparing dense semantic representations of text passages (Salton and McGill, 1983). BLEU (Bilingual Evaluation Understudy) measures surface-level lexical overlap between a generated text and a reference text by computing n-gram precision, the proportion of word sequences in the generated output that also appear in the reference(Mahmud et al. 2025). However, to address the shortcomings of mere token matching in programming languages, variants like CodeBLEU extend this metric by integrating syntactic structure and semantic data-flow matching (Mahmud et al. 2025). ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is another standard metric that measures the overlap of n-grams between generated texts and reference texts. Both BLEU and ROUGE are traditional lexical comparison metrics for evaluating free-text outputs, with BLEU primarily used for machine translation and ROUGE widely used for automatic summarization (Xu et al., 2024). Xu et al. (2024) demonstrate that these traditional unsupervised NLP metrics have significant limitations in specialized domains because they focus on lexical overlap and fail to capture deeper semantic meaning; they propose using LLM-generated categorical labels as an alternative basis for similarity assessment, an approach conceptually parallel to the categorical weight assignments used in this dissertation. Smith et al. (2025) provide a comprehensive analysis of LLM output similarity across 12 models and approximately 3 million generated texts, finding that outputs from the same LLM are more similar to each other than to human-written texts, and that inter-model similarity varies substantially across model families.

None of these text-based similarity metrics is applicable to the outputs of this dissertation's pipeline. Similarity metrics such as cosine similarity, BLEU, or ROUGE for LLM outputs are designed to compare free text (for example, to evaluate how closely a machine-generated translation matches a reference translation). The outputs of this dissertation's pipeline are categorical labels and, as such, can't be compared for linguistic similarity. They can be thought of as independent raters performing the same structured classification task and thus can be compared using inter-rater reliability metrics.

Google DeepMind, and Qwen2.5-7B-Instruct is from Alibaba Qwen Team) and differ in their architectural details, training data, and fine-tuning procedures, thereby introducing meaningful diversity in inductive bias.

Mistral-7B is a 7-billion-parameter model that uses grouped-query attention (GQA) for efficient inference and sliding window attention (SWA) to handle long sequences at reduced computational cost (Jiang et al. 2023). Gemma-2-9b is a 9-billion-parameter model developed by Google DeepMind that employs GQA and additionally interleaves global and local sliding-window attention layers; notably, the 9B variant was trained via knowledge distillation from a larger 27B model rather than standard next-token prediction (Gemma Team 2024). Qwen2.5-7B is a 7-billion-parameter model from the Qwen series that uses GQA, the SwiGLU activation function, rotary positional embeddings (RoPE), and QKV bias in its attention mechanism; it was pre-trained on 18 trillion tokens spanning web text, mathematical reasoning, and programming code (Yang et al. 2024).

All three models share the same transformer architecture with grouped-query attention, which supports broadly consistent prompt processing and interpretation of literature. They differ in training corpora, fine-tuning strategies (standard instruction tuning vs. knowledge distillation), and attention mechanisms (sliding window vs. interleaved local-global), which introduces variation in how each model handles ambiguous or underspecified evidence. This architectural diversity reduces the likelihood that extracted weights reflect the systematic behavior of any single model rather than the content of the retrieved literature and preserves inter-model disagreements as a signal of genuine uncertainty.

Prompting strategy. Each model received the same prompt for comparability. For each run, the model was given the definition of the relevant shape of justice, definitions of the indicators or metrics to be weighted, and the retrieved literature chunks.[12] The prompt instructed the model to treat the supplied definitions as authoritative. All the prompts are presented in Appendix F.

Priority-based weight assignment. Each model assigned weights under two mutually exclusive conditions. Under Priority 1 (Literature-based), the model assigned a weight only when explicit textual support was identified in the retrieved literature. Under Priority 2 (LLM-Reasoning-based), if no explicit support was identified, the model was permitted to assign a weight using its base-trained reasoning. This protocol enforces transparency regarding the evidentiary basis for each weight and identifies areas where the existing literature does not provide direct guidance. It also distinguishes between weights grounded in the domain literature and those that rely on the model's parametric knowledge.

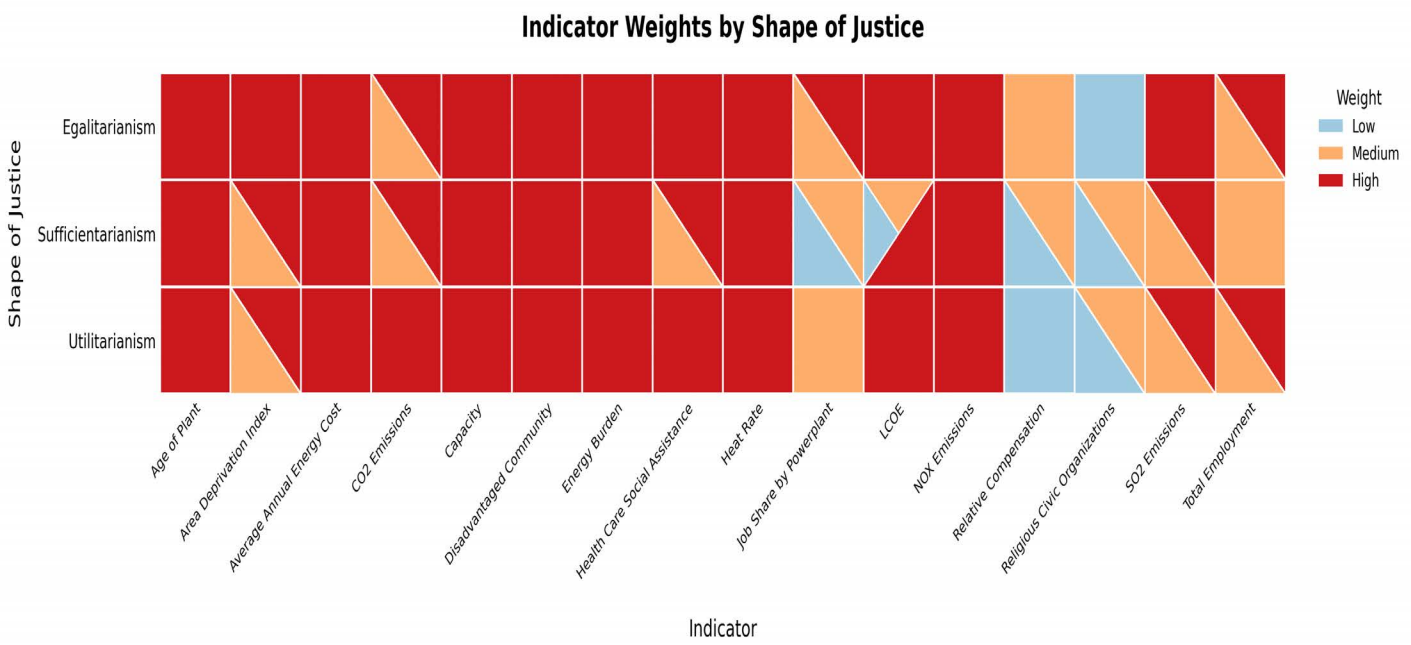
[12] All three shapes of justice were evaluated using the same indicators and metrics, with only the weights varied across them. This was done to test how the same decision-relevant signal receives different importance under each shape of justice. Adding or removing indicators across shapes of justice was avoided because an indicator that is unimportant under a given shape of justice can be assigned a weight approaching zero, so varying the weights rather than the indicator set achieves the same effect while keeping the comparison consistent.

Aggregation across models. Weights produced by the three models were aggregated using a rule that prioritized evidence-supported assignments. Priority 1 weights were given precedence over Priority 2 weights. If at least one model assigned a Priority 1 weight to an item, all Priority 2 weights for that item were disregarded, and the Priority 1 value was taken as the final weight. If two or more models assigned different Priority 1 weights to the same item, the item was classified as having an uncertain weight assignment. If all three models assigned Priority 2 weights and those weights were consistent, the shared value was selected as the final weight. If all three models assigned Priority 2 weights but the values differed, the item was classified as uncertain. No averaging or interpolation was performed. Uncertainty was retained as an analytic feature and propagated into the weight configurations described below.

Weight configurations. The procedure produces qualitative weight levels, low, medium, and high, for each indicator and metric under each shape of justice. If the procedure produced a consistent evidence-supported value, a single weight was assigned and if it produced uncertain assignments, multiple admissible values were retained. Qualitative levels were then mapped to ordinal numerical values, with low = 1, medium = 2, and high = 3. A weight configuration is defined as one complete assignment of numerical weights to all indicators and metrics under a given shape of justice. Weight configurations were generated by expanding uncertain qualitative weights into all admissible values. For example, a weight listed as *medium*|*high* was treated as two possible values, while a weight listed as *low*|*medium*|*high* was treated as three possible values. Under utilitarianism, four indicator weights have two admissible values each, producing $2^4 = 16$ indicator-weight combinations, and two metric weights have two admissible values each, producing $2^2 = 4$ metric-weight combinations. This gives $16 \times 4 = 64$ total weight configurations. Under sufficientarianism, the uncertain indicator weights produce $2^7 \times 3 = 384$ combinations, while metric weights are fixed, producing 384 total configurations. Under egalitarianism, three indicator weights produce $2^3 = 8$ combinations, and two metric weights produce $2^2 = 4$ combinations, giving $8 \times 4 = 32$ total configurations. The weights are used in the ranking and sensitivity analysis described in Section 2.3.

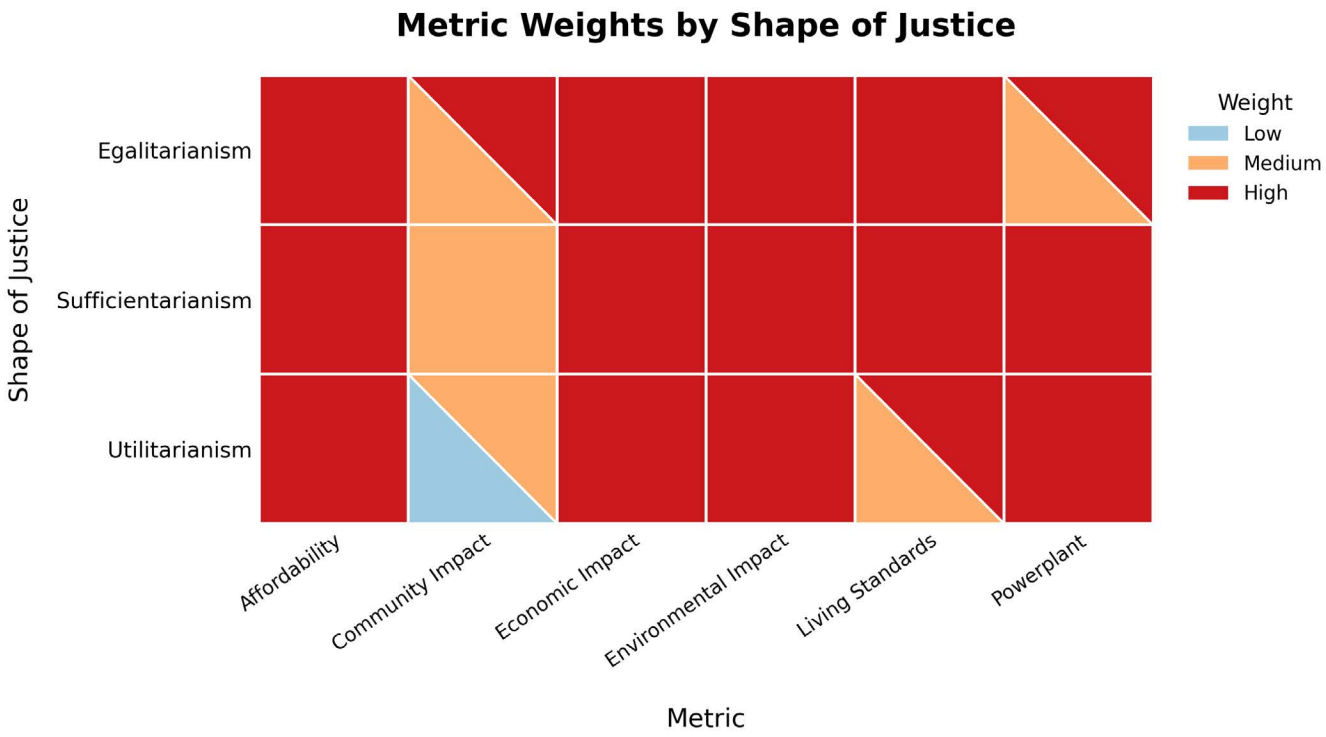
Weight Extraction Results. Figure 3 presents the indicator weights extracted under each shape of justice, and Figure 4 presents the metric weights. In both figures, the y-axis lists the three shapes of justice (utilitarianism, sufficientarianism, egalitarianism) and the x-axis lists the indicators (Figure 3) or metrics (Figure 4). Low weights are shown in blue, medium in light orange, and high in red. Cells with uncertain weight assignments are split to show the two or three plausible weight values produced by the LLM extraction procedure.

Figure 3: Indicator Weights by Shape of Justice



REF: Developed by the Author

Figure 4: Metric Weights by Shape of Justice



REF: Developed by the Author

Two patterns are visible across both figures. First, the same indicator or metric can receive substantially different weights under different shapes of justice, which is the core mechanism by which shapes of justice enter the MCDM model. Second, weight uncertainty is not evenly distributed: it concentrates on a subset of indicators and metrics under each shape of justice, while the remainder receive single, evidence-supported values.

As shown in Figures 3 and 4, Energy Burden, Average Annual Energy Cost, NOx Emissions, Disadvantaged Community, Age of Plant, Capacity, and Heat Rate receive a "high" weight under all three shapes of justice. These indicators form a stable set that the literature treats as decision-relevant for the NG powerplant phaseout regardless of shape of justice. Note that the inclusion of these stable metrics and indicators is deliberate. Even when all three shapes of justice agree on an indicator's importance, that agreement helps identify plants whose phaseout priority is robust across normative framings. Excluding these indicators would remove the analytical basis for distinguishing robust from contested rankings, a key objective of this dissertation.

Eight of 16 indicators have uncertain weights under sufficientarianism, compared with only four under utilitarianism. The lower indicator-level uncertainty under utilitarianism likely reflects a more mature body of literature that provides clear guidance on how to weight individual indicators.

The Community Impact metric weight shifts in a single direction across the shapes of justice, moving from “medium|low” under utilitarianism to “medium” under sufficientarianism to “medium|high” under egalitarianism. A similar directional pattern is visible at the indicator level for Relative Compensation, which moves from “low” under utilitarianism, to “medium|low” under sufficientarianism, to “medium” under egalitarianism. Similarly, Job Share by Powerplant moves from “medium” under utilitarianism to “medium|low” under sufficientarianism, and to “medium|high” under egalitarianism. Relative Compensation and Job Share by Powerplant represent how concentrated and well-compensated power-plant employment is in the local economy, which determines how heavily a community would feel the loss of these jobs in a phaseout. This suggests that the literature consistently associates community-level support and labor-market vulnerability with egalitarianism

All three Economic Impact indicators carry uncertainty under at least one shape of justice, even though the metric weight itself is fixed at "high" across all three. The literature consistently treats Economic Impact as a high-priority consideration in NG phaseout decisions but does not converge on how its underlying indicators should be weighed across shapes of justice.

The agreement among the three models on these weight assignments is reported in Appendix B.

2.3 Equity Component Classification

The same LLM pipeline used for weight extraction in Section 2.2.11 was also used to classify each of the 16 indicators and 6 metrics into one of three equity components: distributional,

contextual, or procedural. The computational setup, including the three open-source LLMs, the retrieval-augmented generation pipeline, and the cross-model aggregation rule, is identical to that described in Section 2.2.11, except for changes to the prompt and the output category. The remainder of this section describes those differences, presents the resulting classifications, and observes the patterns that emerge.

As the next sections describe, eleven of the 16 indicators and five of the six metrics were classified as distributional; four indicators and one metric were classified as contextual; the Location Quotient was retained as Distributional|Contextual; and no indicator or metric was classified as procedural. The implications of the procedural equity gap and the role that multiple shapes of justice play in partially addressing it are discussed in Section 2.3.3.

2.3.1 Method differences from weight extraction

The classification task uses the same retrieval pipeline, the same three LLMs (Mistral-7B-Instruct-v0.3, Gemma-2-9b-it, Qwen2.5-7B-Instruct), the same Priority 1 (Literature-based) and Priority 2 (LLM-Reasoning-based) tagging protocol, and the same cross-model aggregation rule described in Section 2.2.11. The only difference is the prompt that instructs each model to classify the indicator or metric into one of three equity categories, distributional, contextual, or procedural, rather than assign weight.

2.3.2 Equity component classifications

Tables 2 and 3 present the equity component classifications for the 16 indicators and 6 metrics, respectively. As evident from the tables, most indicators (12 of 16) and metrics (5 of 6) are classified as distributional. The indicators classified as contextual share a common characteristic: they measure place-based conditions or community-level attributes rather than population-level outcomes. None of the models classified any indicator or metric as procedural, indicating that the indicator set used in this case study captures distributional and contextual equity but does not capture procedural equity through any plant-level variable. The implications of this absence are discussed in Section 2.3.3.

Please note that the Living Standards metric was classified as distributional, but its two constituent indicators, Disadvantaged Community and Area Deprivation Index, were classified as contextual. This is because the metric was interpreted as describing outcomes distributed across populations, whereas the two indicators were interpreted as describing place-specific conditions and neighborhood-level disadvantages. This ambiguity is a normative judgment that the model cannot resolve on its own and is an area where stakeholder engagement could directly inform the classification.

Similar to weight assignment, model-level classifications and the Priority 1 / Priority 2 distribution are reported in Appendix C.

2.3.3 Absence of Procedural Equity

As shown in Tables 1 and 2, none of the 16 indicators or 6 metrics was classified as procedural by any of the three models. This dissertation, therefore, does not directly operationalize procedural equity through its indicators. Procedural equity is fundamentally about process: who has access to and agency in the decisions that affect them. This makes sense: no indicator measuring plant characteristics or community conditions can capture whether affected communities had a genuine seat at the table in phaseout decisions.

The use of multiple shapes of justice offers a partial and imperfect bridge toward procedural concerns. By running the model under utilitarianism, egalitarianism, and sufficientarianism, the architecture represents the diversity of normative frameworks that different stakeholders might bring to the decision, even if those stakeholders are not literally present in this analysis. This is not procedural equity in any strict sense, since representing value pluralism analytically is not the same as giving communities agency over the decision. However, the architecture is explicitly designed to accommodate direct stakeholder input in future implementations, where weights could be derived from community engagement rather than inferred from academic literature via LLMs. That extension would move the framework closer to procedural equity and represent an important direction for future work.

2.4 Ranking Methodology and Sensitivity Analysis

This section describes how composite scores are translated into phaseout rankings and how the robustness of those rankings is assessed. The ranking methodology and decision tiers are described in Sections 2.4.1. The global sensitivity analysis used to identify indicators that drive ranking variability is described in Sections 2.4.2. Plant rankings and sensitivity results are reported in Chapter 3.

2.4.1 Ranking Methodology

Power plants are ranked by their composite scores under each shape of justice. As described in Section 2.2.10, indicators are normalized so that a higher normalized value always represents a more desirable condition. A higher composite score therefore indicates a plant with a more desirable profile across the indicators. A lower composite score indicates a plant that performs poorly across most or all the indicators. Plants are ranked in ascending order of composite score, so that the plant with the lowest composite score receives Rank 1 and is the highest-priority candidate for phaseout.

Each shape of justice produces a set of weight configurations, and the ranking procedure is repeated for every weight configuration. Thus, each powerplant has a distribution of ranks that reflects how its position changes as weights vary across configurations.

Descriptive statistics are computed to analyze each plant's ranking distribution. The median rank represents the central tendency of a powerplant's ranking distribution and is used as the

primary ranking statistic because it is robust to skewness. The 25th percentile (P25) and 75th percentile (P75) indicate the lower and upper bounds of the central fifty percent of rankings. The interquartile range (P75 minus P25) measures the dispersion of the ranking distribution.

2.4.2 Global Sensitivity Analysis

A global sensitivity analysis was conducted to identify which indicators drive differences in plant rankings when weights vary across configurations. To achieve this, Random Forest regression was utilized. Random Forest regression is an ensemble learning method that constructs a large number of decision trees on bootstrapped samples of the data, with random feature subsets considered at each split, and aggregates predictions across trees by averaging (Breiman, 2001; Schonlau and Zou, 2020). This algorithm was selected over variance-based sensitivity methods, such as Sobol indices, because it can model nonlinear relationships and interactions among indicators without assuming independence among inputs (Antoniadis et al., 2021). This is particularly advantageous given the potential spatial and structural correlations present among the 16 indicators used in this analysis.

Other MCDM robustness methods exist, such as the Stochastic Multicriteria Acceptability Analysis (SMAA) framework. SMAA evaluates the acceptability of alternatives across uncertain or partially specified preference spaces using rank acceptability indices (Lahdelma and Salminen, 2010). This dissertation has a narrower objective. It does not seek "the most" preferred alternative or treat the weight space as a fully stochastic preference distribution. Instead, the weight space is defined by the discrete set of literature-derived weight configurations produced under each shape of justice. The goal is to summarize rank stability within those configurations and identify which indicators are associated with rank variability. Distributional statistics summarize each plant's ranking distribution. Random Forest regression identifies which indicators predict larger IQR values. SMAA and similar methods would be appropriate in future work if stakeholder-derived or probabilistic weight distributions are elicited directly.

For each shape of justice, a Random Forest regressor was fitted with the 16 normalized indicator values as features and the interquartile range of each plant's rank distribution as the target variable. The interquartile range was selected as the target because the analytical question is which indicators drive ranking *instability* across weight configurations rather than which indicators determine a plant's absolute rank. A model targeting the median rank would describe plant performance, while a model targeting the interquartile range would describe how much a plant's rank shifts as weights vary. Plants with large interquartile ranges have rankings that shift substantially across weight configurations, while plants with small interquartile ranges maintain consistent positions.

The model was configured with 500 decision trees and a maximum depth of 6. Five hundred trees is a commonly accepted convention for Random Forest regression. Research demonstrates that performance metrics consistently plateau and converge at or well before this threshold (Payabvash et al., 2020; Probst and Boulesteix, 2018; Zafari et al., 2019). The maximum depth of

6 acts as a regularization constraint. While foundational Random Forest methodology typically grows unpruned trees to their maximum size (Breiman, 2001), defining a maximum depth is an established stopping condition (Zafari et al., 2019) that allows for reasonable interaction modeling among the 16 features while preventing trees from growing deep enough to memorize individual data points, which would produce unstable importance estimates.

Feature importance was quantified using Mean Decrease in Impurity (MDI), defined as the weighted decrease in impurity associated with splits along a given variable; in regression, impurity is the variance of the output, and for Random Forests the MDI is averaged across trees (Scornet, 2020). The resulting importance scores are normalized to sum to one across all indicators within each shape of justice, representing the proportion of explained variance attributable to each indicator. A higher importance score indicates that the indicator has a greater influence on ranking variability.

MDI scores should be interpreted with care because impurity-based importance measures are biased when predictors vary in their scale of measurement or number of categories (Strobl et al., 2007), and random forest tree-building processes inherently exhibit a strong preference for correlated predictor variables (Strobl et al., 2008). When two or more indicators are correlated, the algorithm tends to split on whichever indicator enters the tree first. This can inflate the importance of the indicator that entered the tree first and deflate the importance of the others, even when their underlying contributions to the outcome are similar. This is a relevant concern for the present analysis because several indicators are derived from overlapping spatial units, including county, MSA, and census tract, and may share common variation by construction. Importance scores should therefore be read as descriptive measures of predictive influence on the IQR of the rank distribution rather than as unbiased estimates of an indicator's causal strength. A low importance score does not necessarily indicate that an indicator is unimportant; it may instead indicate that the indicator's variation is shared with another indicator that entered the trees first.

Also note that sensitivity analysis described in this section identifies which indicators drive ranking variability *within* a given shape of justice, when weights vary across the configurations defined in Section 2.2.11. It does not assess sensitivity across shapes of justice. The analysis is also conditional on the indicator set, normalization rule, and plant universe defined in Section 2.2. Different indicators, normalization choices, or a different set of plants would produce different importance rankings, and the importance scores reported here should be read within the bounds of the inputs used in this dissertation.

2.5 Planning Exercise: CPUC Gas Capacity Not Retained Ranking

The California Public Utilities Commission's 24-25 TPP Busbar Mapping Gas Capacity Not Retained ranking serves as the reference baseline in Chapter 3. Produced by CPUC Energy Division Staff (CPUC, 2024), the score identifies specific natural gas units to be modeled as "not retained" in the California Independent System Operator's Transmission Planning Process studies

(California Public Utilities Commission, 2024a; California Public Utilities Commission, 2024b). Plants are scored on six factors organized in three categories: 50 percent on Environmental and Community factors (Disadvantaged Community proximity, NOx emissions rate weighted by capacity factor, and Air Quality Non-Attainment Zone status), 25 percent on Performance factors (Heat Rate and Age), and 25 percent on a Local Reliability factor (Local Effectiveness Factor). Generators were also excluded if they were in the youngest Age Factor quartile or the highest Local Effectiveness Factor (LEF) quartile. CPUC explicitly states that their study is intended for transmission planning purposes and does not direct the retirement of specific gas generators.

Four of CPUC's six factors are the same as the indicators used in this dissertation: Disadvantaged Community status (Section 2.2.9), NOx Emissions (Section 2.2.5), Heat Rate (Section 2.2.8), and Age of Plant (Section 2.2.8). The Air Quality Non-Attainment Zone factor combines ozone and PM2.5 status. However, this dissertation uses NOx and SO2 emissions instead, as NOx is a precursor to ground-level ozone, and NOx and SO2 are precursors to PM2.5 formation via nitrate and sulfate particles. The Local Effectiveness Factor is not included, as it is a CAISO-defined grid-level reliability metric rather than a plant-level attribute relevant to phaseout prioritization. Beyond these six factors, this dissertation includes ten additional, literature-specified indicators that CPUC does not consider.

Note that the four CPUC factors that overlap with this dissertation's indicators all received "high" weight in the LLM-extracted weighting procedure described in Section 2.2.11. Disadvantaged Community, NOx Emissions, Heat Rate, and Age of Plant each received a "high" weight across all three shapes of justice, placing them in the set of indicators identified as universally important for natural gas power plant phase-out decisions.

Chapter 3 uses the CPUC scoring alongside the phase-out rankings produced under utilitarianism, sufficientarianism, and egalitarianism. The comparison is not exact, as the two frameworks differ in purpose, indicator coverage, and weighting methodology. Also, the CPUC scoring covers only CAISO plants (Pacific Gas and Electric, Southern California Edison, and San Diego Gas and Electric). However, this dissertation uses the Global Energy Monitor Global Oil and Gas Plant Tracker (Global Energy Monitor, 2024) to identify active natural gas power plants across all utility territories in California, including plants operated by or located within the service areas of the Los Angeles Department of Water and Power, the Sacramento Municipal Utility District, and the Imperial Irrigation District, which fall outside the CAISO balancing authority and are not covered by the CPUC scoring.

Although the CPUC scores were developed as a planning exercise to support transmission planning rather than to assess equity directly, they provide a useful illustration of how policy priorities and analytical methods can evolve. The CPUC methodology incorporates several indicators with equity relevance, such as Disadvantaged Community proximity and emissions rates, and their weights were derived through a stakeholder workshop process. However, the weights are fixed and do not vary across stakeholder perspectives or normative framings. This means that the methodology, as currently designed, does not make visible how different public

values and aspirations might produce different rankings. The analytical architecture developed in this dissertation complements approaches like the CPUC methodology by offering a structured way to evaluate the same plants under different shapes of justice, each representing different stakeholder values and perspectives. As will be discussed in Chapter 3, the limited convergence between this dissertation's rankings and the CPUC scores suggests that the inclusion of additional equity-focused indicators and the variation of weights across shapes of justice can meaningfully change phaseout prioritization. Incorporating these dimensions into planning exercises could provide additional insights for agencies seeking to align their analytical methods with broader equity commitments.

Chapter 3. Results

This chapter presents the phaseout rankings of 96 California natural gas power plants under three shapes of justice: utilitarianism, sufficientarianism, and egalitarianism.

Section 3.2 reports the dissertation ranking results. It first identifies plants that appear in the top 20 under at least one shape of justice, then examines rank variability across weight configurations using the interquartile range. This section also reports the global sensitivity analysis used to identify which indicators contribute most to rank variability. Section 3.3 compares the dissertation rankings with the CPUC Gas Capacity Not Retained ranking as a planning-based reference point. The comparison is limited by differences in purpose, indicator coverage, fleet scope, and unit of analysis, but it helps show where the dissertation ranking aligns with an existing planning exercise and where the broader indicator set changes phaseout-priority results.

Section 3.2.1 identifies 23 plants that rank in the top 20 under at least one shape of justice, of which 17 appear in the top 20 under all three shapes. Section 3.2.2 shows that ranking variability is concentrated among middle-ranked plants with mixed normalized indicator profiles, and that the indicators driving this variability differ across shapes of justice. Section 3.3 compares the dissertation's rankings with the CPUC baseline and shows that broadening the indicator set to include equity-relevant dimensions may provide decision signals for improved planning. Chapter 4 discusses these findings in relation to the research questions and considers their implications for phaseout planning.

3.2 Plant Rankings Under Three Shapes of Justice

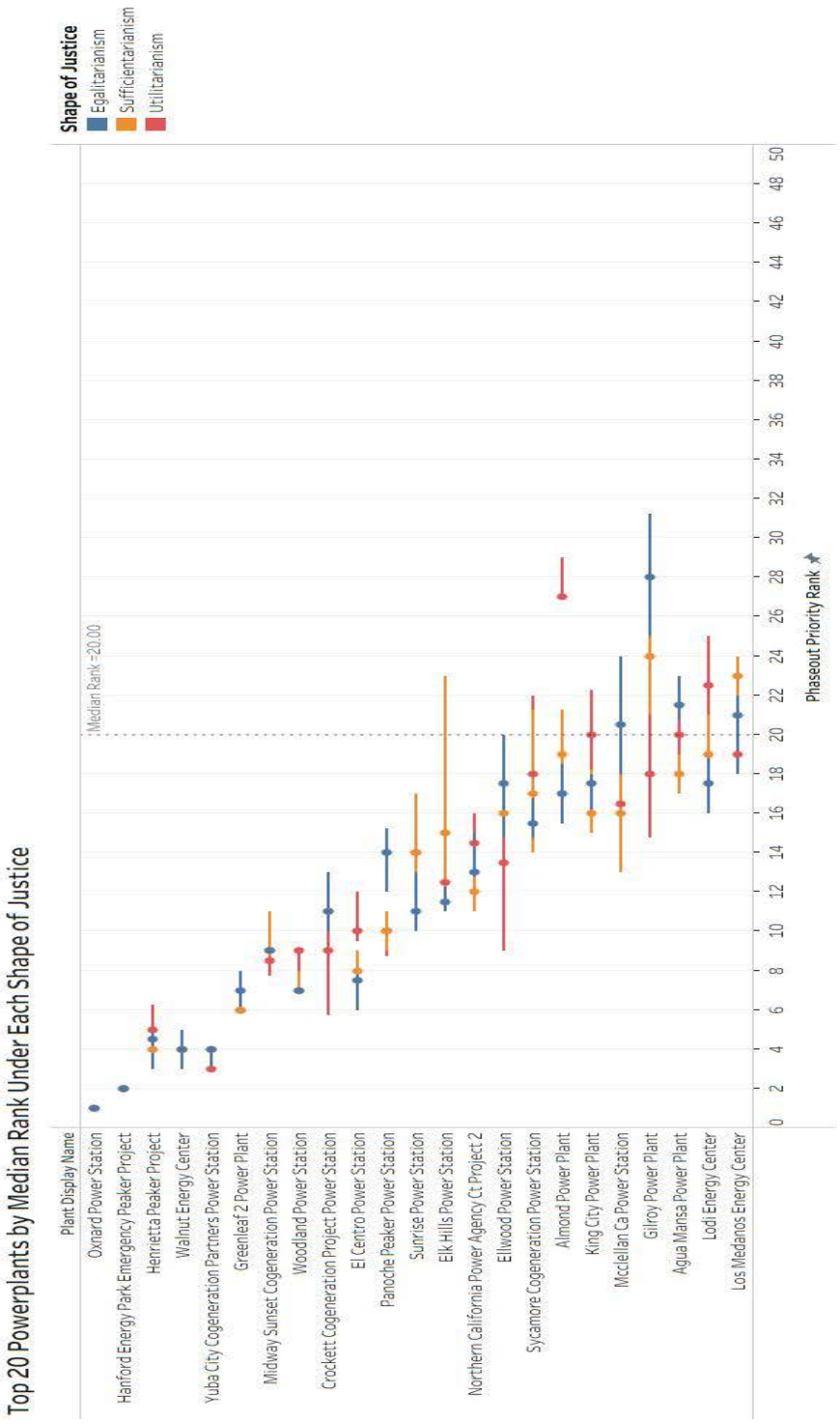
This section presents plant phaseout rankings under the three shapes of justice. Lower median ranks indicate higher phaseout priority. The analysis reports median rank, P25, P75, and IQR for each plant under utilitarianism, sufficientarianism, and egalitarianism. Section 3.2.1 identifies the plants that appear in the top 20 under at least one shape of justice and examines whether their inclusion is stable across shapes. Section 3.2.2 identifies plants whose rankings are most sensitive to indicator and metric weights, then uses global sensitivity analysis to identify which indicators contribute most to that variability. The cutoffs used in this section are for demonstration purposes and are selected to facilitate the presentation of results in an accessible manner.

3.2.1 Top 20 plants by median rank under each shape of justice

Figure 5 presents the 23 powerplants that rank in the top 20 under at least one shape of justice. The figure plots each plant’s median phaseout priority rank across utilitarianism, sufficientarianism, and egalitarianism. Circles indicate the median rank, and horizontal bars show

the interquartile range from the 25th to the 75th percentile across weight configurations. Lower ranks indicate higher priority for phaseout.

Figure 5: Top 20 Powerplants by Median Rank Under Each Shape of Justice



REF: Developed by the Author. Note that a lower median rank means a higher phaseout priority

Of the 23 plants in Figure 5, 17 appear in the top 20 under all three shapes of justice[13]. These 17 plants form a stable early-phaseout core because their inclusion in the top 20 does not depend on the shape of justice.

Oxnard Power Station, Hanford Energy Park Emergency Peaker Project, and Walnut Energy Center share the same median rank across all three shapes, with Oxnard and Hanford also having an IQR of zero, indicating their rankings do not change under any weight configuration. Of the remaining 14 plants in the core, the largest cross-shape median spread is 4 ranks, observed for *Panoche Peaker Power Station* (median rank 10 under utilitarianism and sufficientarianism, 14 under egalitarianism), *Ellwood Power Station* (13.5 under utilitarianism, 16 under sufficientarianism, 17.5 under egalitarianism), and *King City Power Plant* (20 under utilitarianism, 16 under sufficientarianism, 17.5 under egalitarianism). Within-shape ranking variability is also limited for most core plants, though *Elk Hills Power Station* shows an IQR of 12 under sufficientarianism and *Ellwood Power Station shows* an IQR of 10 under utilitarianism, indicating that their position within the top 20 is sensitive to which weight configuration is applied, even though their membership in the top 20 is not. The full median rank, P25, P75, and IQR for the top 20 powerplants under each shape of justice are reported in Appendix D.

There are six plants that appear in the top 20 under one or two shapes of justice, but not all three. This illustrates how a robust analysis at a given cutoff would identify plants whose inclusion depends on the shape of justice applied.

At a top 20 cutoff, Gilroy Power Plant qualifies for early phaseout under utilitarianism, with a median rank of 18.00. It does not qualify under sufficientarianism or egalitarianism, with median ranks of 24.00 and 28.00, respectively. This produces a spread of 10 ranks across shapes of justices. Almond Power Plant shows the opposite pattern. It falls outside the top 20 under utilitarianism, with a median rank of 27.00, but qualifies under sufficientarianism and egalitarianism, with median ranks of 19.00 and 17.00. McClellan CA Power Station and Agua Mansa Power Plant also sit near the cutoff. Both qualify under two shapes of justice but fall just outside the top 20 under egalitarianism, with median ranks of 20.50 and 21.50, respectively.

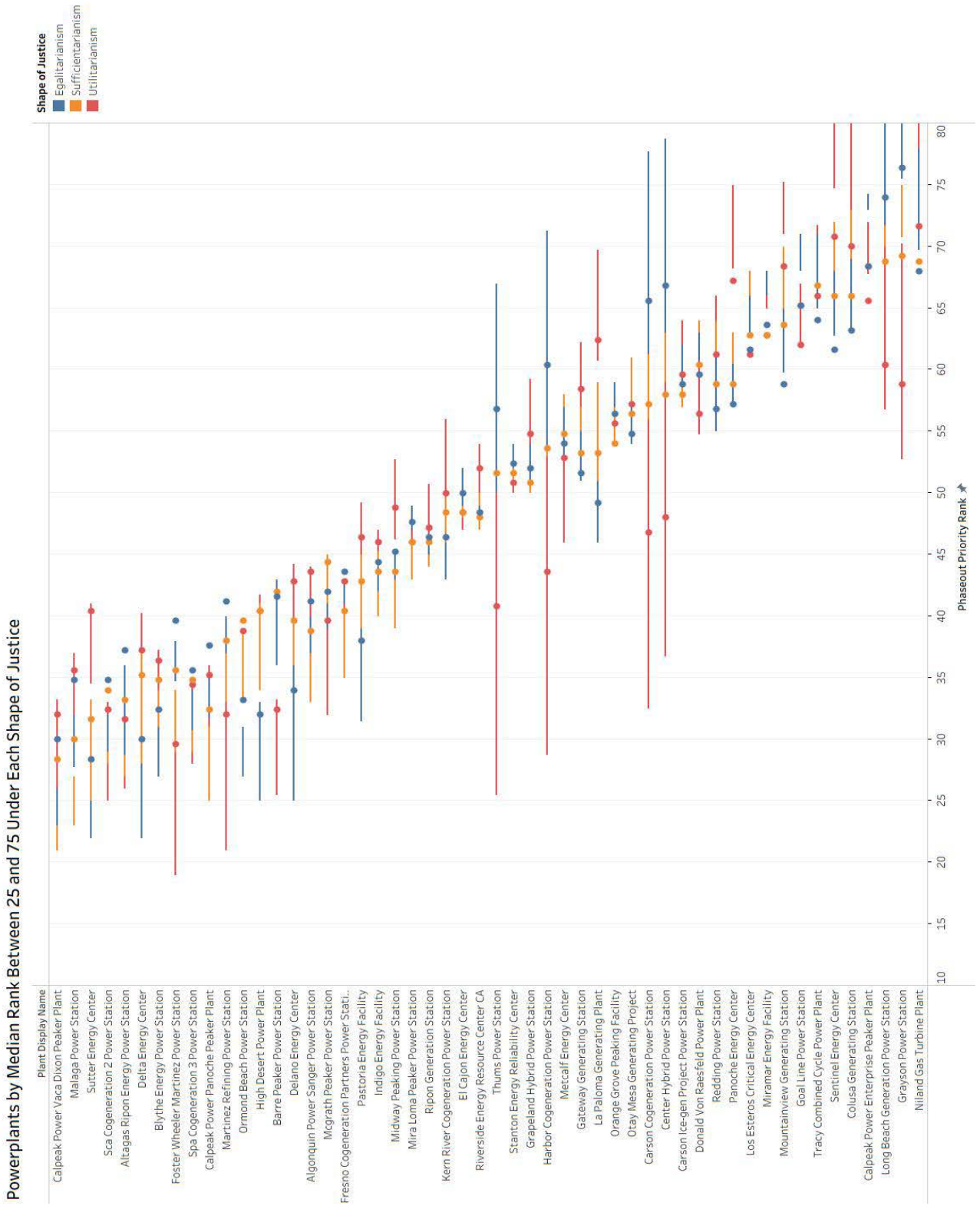
As explained here, the top 20 cutoff is used for demonstration purposes only to show changes in inclusion across shapes of justice. Other cutoffs could be selected based on policy goals or stakeholder priorities. Those cutoffs would likely create similar boundary cases, where a plant's inclusion depends on both the cutoff and the shape of justice used. These cases are important for stakeholder deliberation because their inclusion or exclusion is partly shaped by value judgments about which shape of justice should guide NG phaseout planning.

[13] Oxnard Power Station, Hanford Energy Park Emergency Peaker Project, Yuba City Cogeneration Partners Power Station, Walnut Energy Center, Henrietta Peaker Project, Greenleaf 2 Power Plant, Woodland Power Station, Midway Sunset Cogeneration Power Station, El Centro Power Station, Crockett Cogeneration Project Power Station, Panoche Peaker Power Station, Elk Hills Power Station, Sunrise Power Station, Northern California Power Agency CT Project 2, Ellwood Power Station, Sycamore Cogeneration Power Station, and King City Power Plant.

3.2.2 Causes for Powerplant Phaseout Ranking Variability Under Different Shapes of Justice

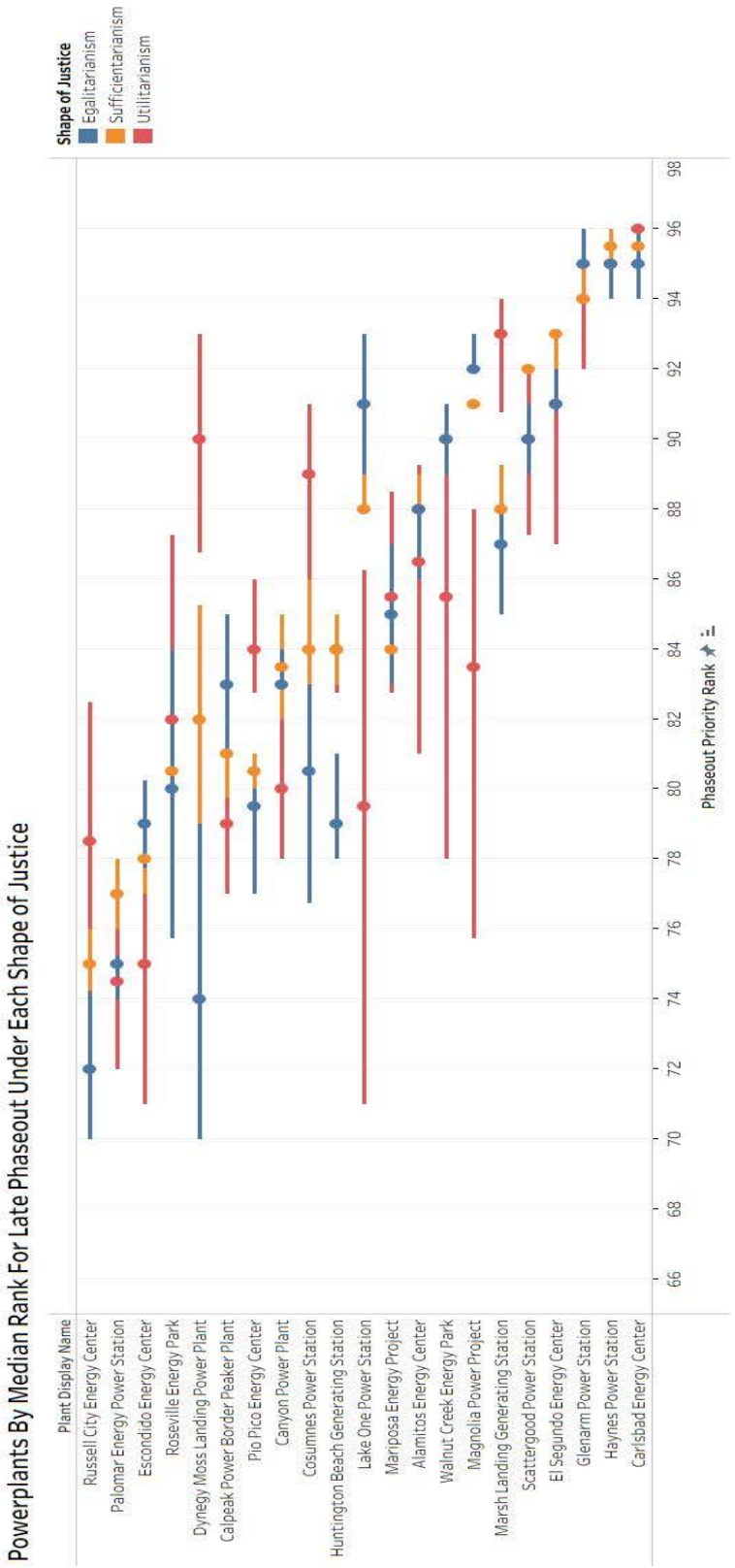
This section discusses the variability in power plant rankings across all weight configurations for each shape of justice, and the causes of this variability. The spread of each plant's rank across the weight configurations (due to uncertainty in the weights of metrics and indicators) is measured using the interquartile range (IQR), as it is robust to outliers. A higher IQR means that a plant's rank changes more across weight configurations within the shape of justice.

Figure 6: Powerplants by Median Rank between 25 to 75 Under Each Shape of Justice



REF: Developed by the Author. Note that a lower median rank means a higher phaseout priority

Figure 7: Powerplants by Median Rank for Late Phaseout Under Each Shape of Justice



REF: Developed by the Author. Note that a lower median rank means a higher phaseout priority

Figure 6 shows plants whose median rank across the three shapes of justice falls between 25 and 75. For each plant, the figure reports the IQR of its rank distribution under utilitarianism, sufficientarianism, and egalitarianism. As shown in the figure, these middle-ranked plants have larger IQRs than those near the top or bottom of the phaseout-priority ranking. For example, *THUMS Power Station* (median ranks of 38.5 under utilitarianism, 52 under sufficientarianism, and 58.5 under egalitarianism) has the largest single IQR in the dataset, at 25.50 under utilitarianism. *Carson Cogeneration Power Station* (median ranks of 46 under utilitarianism, 60 under sufficientarianism, and 69.5 under egalitarianism) follows, with its largest IQR under utilitarianism at 24.75. Harbor Cogeneration Power Station also has its largest IQR under utilitarianism at 24.75.

This pattern differs from that of the plants at the extreme ends (lowest or highest median rank) of the phaseout rankings. This pattern is visible in Figures 5, 6, and 7, where the horizontal bars representing the IQR are narrow for plants at both extremes and wider for plants in the middle of the ranking distribution. The highest-phaseout-priority plants, defined as plants with median ranks of 10 or lower, have an average IQR of 1.48 under utilitarianism, 1.03 under sufficientarianism, and 1 under egalitarianism. For example, *Oxnard Power Station* holds a median rank of 1 under utilitarianism, sufficientarianism, and egalitarianism, with an IQR of 0 under all three shapes of justice. Its normalized indicator values are low across several indicators, including Capacity (0.000), Total Employment (0.000), and Health Care and Social Assistance (0.070). When consistently low indicator values are combined with the indicator and metric weights, the plant receives low composite scores regardless of weight configurations. As a result, Oxnard power station has no variability in its powerplant ranking and remains the highest-priority plant for phaseout across all three shapes of justice.

However, the middle-ranked plants have mixed normalized indicator data profiles, with higher values on some indicators and lower values on others. For example, *THUMS Power Station* illustrates this mixed indicator profile. It has high normalized values for Health Care and Social Assistance, Religious and Civic Organizations, Energy Burden, and Average Annual Energy Cost, all at 1.000. It also has low values for Capacity (0.004), Total Employment (0.020), Job Share by Powerplant (0.101), Disadvantaged Community (0.000), Area Deprivation Index (0.000), and Levelized Cost of Energy (0.229). Its emissions indicators fall between these two groups, with SO_2 Emissions at 0.626, NO_x Emissions at 0.659, and CO_2 Emissions at 0.886. As the plant performs differently across indicators, its composite score is more sensitive to changes in indicator and metric weights. *Harbor Cogeneration Power Station* shows a similar mixed indicator profile. It has high normalized values for Health Care and Social Assistance and Religious and Civic Organizations, both 1.000. It also has high normalized values for SO_2 Emissions (0.980), NO_x Emissions (0.992), and CO_2 Emissions (0.960). Its Energy Burden and Average Annual Energy Cost values are also higher, at 0.714 and 0.686. At the same time, it has lower values for Capacity (0.036), Total Employment (0.047), Job Share by Powerplant (0.101), Area Deprivation Index (0.000), and Levelized Cost of Energy (0.229). This mix of higher and lower indicator values

makes its rank responsive to changes in the indicator and metric weights. There are 22 plants whose phaseout rankings are sensitive to the shape of justice.[14]

Carlsbad Energy Center illustrates the late-phaseout end of the ranking (Figure 7). It has median ranks of 96.00 under utilitarianism, 95.50 under sufficientarianism, and 95.00 under egalitarianism. Its IQR values are 0.25, 1.00, and 2.00, respectively. Its normalized indicator values are high on several indicators, including Disadvantaged Community at 0.900, Area Deprivation Index at 0.800, SO_2 Emissions at 0.986, NO_x Emissions at 0.998, CO_2 Emissions at 0.977, Age of Plant at 0.958, and Heat Rate at 0.940. Because higher normalized values indicate better performance, they yield high composite scores across weight configurations. Thus, Carlsbad Energy Center remains near the lowest phaseout-priority end under all three shapes of justice. There are 16 plants that have low phaseout priority under all three shapes of justice.[15]

The phaseout rankings of powerplants with high IQR values are sensitive to the choice of weights. Since the normalized indicator values are fixed in the model, this variation originates from the weights assigned to indicators and metrics under each shape of justice. Stakeholder engagement in weight assignments would therefore have the greatest effect in reducing rank variability for these plants, as it can help narrow or justify the weight choices that drive their rankings. The global sensitivity analysis described in Section 2.4.2 identifies which indicators contribute most to this variability under each shape of justice, allowing policymakers to direct deliberation toward the specific weight assumptions that most affect phaseout prioritization rather than revisiting the entire model. Tables 4, 5, and 6 report the indicator importance scores under utilitarianism, sufficientarianism, and egalitarianism, respectively.

Table 4: Global Sensitivity Analysis: Indicator Importance Scores Under Utilitarianism

Shape of Justice	Indicator	Importance
utilitarianism	Religious and Civic Organizations	0.167
utilitarianism	Total Employment	0.167
utilitarianism	Health Care and Social Assistance	0.151

[14] Some examples include Sentinel Energy Center, SPA Cogeneration 3 Power Station, Panoche Energy Center, Delta Energy Center, Sutter Energy Center, and Blythe Energy Power Station. These plants have mixed, normalized indicator data profiles, making their composite score sensitive to the choice of weights.

[15] Carlsbad Energy Center, Haynes Power Station, Glenarm Power Station, El Segundo Energy Center, Marsh Landing Generating Station, Magnolia Power Project, Walnut Creek Energy Park, Alamitos Energy Center, Lake One Power Station, Mariposa Energy Project, Cosumnes Power Station, Huntington Beach Generating Station, Canyon Power Plant, Pio Pico Energy Center, CalPeak Power Border Peaker Plant, and Roseville Energy Park. These 16 plants are bottom 20 under all three shapes of justice. They have high normalized values across most indicators and are candidates for later phaseout under any shape of justice.

utilitarianism	Heat Rate	0.120
utilitarianism	Capacity	0.073
utilitarianism	NO_x Emissions	0.047
utilitarianism	Age of Plant	0.043
utilitarianism	Area Deprivation Index	0.043
utilitarianism	Relative Compensation	0.032
utilitarianism	Energy Burden	0.030
utilitarianism	Disadvantaged Community	0.029
utilitarianism	Average Annual Energy Cost	0.024
utilitarianism	SO_2 Emissions	0.022
utilitarianism	CO_2 Emissions	0.020
utilitarianism	LCOE	0.019
utilitarianism	Job Share by Powerplant (Location Quotient)	0.010

Note: Importance scores are derived from a Random Forest regression model fitted with the 16 normalized indicator values as features and the interquartile range (IQR) of each plant's rank distribution as the target variable. Scores are normalized to sum to 1.0 across all indicators. A higher score indicates that the indicator has a greater influence on rank variability across the 64 weight configurations.

Table 5: Global Sensitivity Analysis: Indicator Importance Scores Under Sufficientarianism

Shape of Justice	Indicator	Importance
sufficientarianism	Job Share by Powerplant (Location Quotient)	0.124
sufficientarianism	CO_2 Emissions	0.106
sufficientarianism	LCOE	0.099
sufficientarianism	Average Annual Energy Cost	0.086
sufficientarianism	Capacity	0.071
sufficientarianism	Heat Rate	0.071
sufficientarianism	Total Employment	0.067
sufficientarianism	Age of Plant	0.059
sufficientarianism	NO_x Emissions	0.058
sufficientarianism	SO_2 Emissions	0.052
sufficientarianism	Religious and Civic Organizations	0.042
sufficientarianism	Disadvantaged Community	0.040

sufficientarianism	Relative Compensation	0.035
sufficientarianism	Health Care and Social Assistance	0.035
sufficientarianism	Area Deprivation Index	0.032
sufficientarianism	Energy Burden	0.022

Note: Importance scores are derived from a Random Forest regression model fitted with the 16 normalized indicator values as features and the IQR of each plant's rank distribution as the target variable. Scores are normalized to sum to 1.0. A higher score indicates greater influence on rank variability across the 384 weight configurations.

Table 6: Global Sensitivity Analysis: Indicator Importance Scores Under Egalitarianism

Shape of Justice	Indicator	Importance
egalitarianism	Total Employment	0.167
egalitarianism	Heat Rate	0.137
egalitarianism	Capacity	0.112
egalitarianism	Disadvantaged Community	0.093
egalitarianism	Age of Plant	0.077
egalitarianism	Religious and Civic Organizations	0.072
egalitarianism	Health Care and Social Assistance	0.069
egalitarianism	LCOE	0.051
egalitarianism	Average Annual Energy Cost	0.046
egalitarianism	NO_x Emissions	0.040
egalitarianism	Relative Compensation	0.032
egalitarianism	CO_2 Emissions	0.031
egalitarianism	Energy Burden	0.025
egalitarianism	Area Deprivation Index	0.021
egalitarianism	SO_2 Emissions	0.018
egalitarianism	Job Share by Powerplant (Location Quotient)	0.0120

Note: Importance scores are derived from a Random Forest regression model fitted with the 16 normalized indicator values as features and the IQR of each plant's rank distribution as the target variable. Scores are normalized to sum to 1.0. A higher score indicates greater influence on rank variability across the 32 weight configurations.

Table 7: Top Four Indicators by Importance Under Each Shape of Justice

Rank	Indicator	Metric	Importance	Equity Components
Egalitarianism				

1	Total Employment	Economic Impact	0.167	Distributional
2	Heat Rate	Powerplant	0.137	Distributional
3	Capacity	Powerplant	0.112	Distributional
4	Disadvantaged Community	Living Standards	0.093	Contextual
Sufficientarianism				
1	Job Share by Powerplant (Location Quotient)	Economic Impact	0.124	Distributional | Contextual
2	CO_2 Emissions	Environmental Impact	0.106	Distributional
3	LCOE	Affordability	0.099	Distributional
4	Average Annual Energy Cost	Affordability	0.086	Distributional
Utilitarianism				
1	Religious and Civic Organizations	Community Impact	0.167	Contextual
2	Total Employment	Economic Impact	0.167	Distributional
3	Health Care and Social Assistance	Community Impact	0.151	Contextual
4	Heat Rate	Powerplant	0.120	Distributional

Note: Indicators are ranked by Global Sensitivity Analysis importance scores within each shape of justice. The top four represent the top quartile of the 16 indicators. Metric category and equity component classification (as determined in Section 2.3) are included to show how the drivers of rank variability relate to equity dimensions and thematic groupings.

The global sensitivity analysis (GSA) identifies which indicators are most associated with rank variability. Please note that GSA does not measure sensitivity across shapes of justice. A different indicator data source or a different set of plants would affect the results. These are the scope limitations of the analysis.

Under utilitarianism, the largest contributors are Religious and Civic Organizations (0.167), Total Employment (0.167), and Health Care and Social Assistance (0.151). These indicators together account for 48.6% of total importance. This is an important observation, as several middle-ranked plants have mixed data profiles across these specific indicators. For example, THUMS Power Station and Harbor Cogeneration Power Station have high values for Religious and Civic Organizations and Health Care and Social Assistance, but low value for Total Employment and Capacity. Their rank variability is therefore linked to mixed indicator data profiles, not only to uncertainty in the weights. This observation is also supported by the share of importance assigned to fixed-weight indicators. Fixed-weight indicators account for 70.6% of total importance under utilitarianism, 79.0% under egalitarianism, and 47.4% under sufficientarianism.

This finding helps identify where stakeholder engagement is most useful. If rank variability is concentrated in a small set of indicators, then uncertainty in those indicators' weights has a greater influence on the ranking distribution of each power plant. These indicators therefore identify where stakeholder deliberation and further data validation would be most valuable. Also, the finding that dominant indicators differ across shapes of justice supports the methodological claim that equity-informed phaseout planning depends not only on the indicator set but also on the justice principle used to weight and interpret those indicators. Table 7 identifies the top four indicators by importance for each shape of justice and reports their equity-component classification from Section 2.3. This allows a direct comparison of which indicators, and which equity dimensions drive ranking variability under each shape of justice.

No indicator appears in the top quartile under all three shapes of justice. Total Employment appears in the top quartile under egalitarianism and utilitarianism but not under sufficientarianism. Heat Rate appears in the top quartile under egalitarianism and utilitarianism but not under sufficientarianism. Under sufficientarianism, the top quartile is dominated by affordability and environmental indicators (LCOE, Average Annual Energy Cost, CO2 Emissions), none of which appear in the top quartile under egalitarianism or utilitarianism. The choice of the shape of justice determines which indicators drive ranking uncertainty.

The equity component classification of the four indicators that contribute most to ranking variability varies by shape of justice. Under egalitarianism, three are distributional and one is contextual (Disadvantaged Community). Under sufficientarianism, all four are distributional. Under utilitarianism, two are contextual (Religious and Civic Organizations, Healthcare and Social Assistance) and two are distributional (Total Employment, Heat Rate). As established in Chapter 1, existing equity frameworks primarily focus on distributional equity only and thus would have missed the contextual indicators that are important drivers of variability under utilitarianism and egalitarianism.

Another important consideration is whether the ranking is stable when a plant is added to or removed from the set of plants considered. The weights and the aggregation are independent across plants, so removing a plant does not change how any other plant's composite score is computed, and the relative order of the remaining plants is preserved through these steps. The channel through which removal can change the ranking is normalization. For any indicator normalized using the minimum and maximum across the plant set, removing a plant that holds an extreme value on that indicator changes the minimum or maximum and rescales the normalized values of the remaining plants, which can alter their composite scores and relative ranks. This is most relevant to the practical case in which the highest-priority plants are phased out and removed from the set, because these plants tend to have low normalized values across several indicators, as shown in Section 3.2.1, and are therefore more likely to hold the extreme values that anchor the normalization. Quantifying how much the ranking of the remaining plants shifts under sequential removal, and how this interacts with the sensitivity of the ranking to different shapes of justice, was beyond the scope of this dissertation.

3.3 Referencing Against CPUC Gas Capacity Non-Retention Methodology

Section 2.5 introduced the CPUC Gas Capacity Not Retained scores as a reference for comparison. The comparison is limited because the two methods differ in purpose, indicator coverage, fleet scope, and unit of analysis. CPUC's ranking was developed for transmission planning and uses *Disadvantaged Communities*, *NOx emissions weighted by capacity factor*, *air quality nonattainment zones*, *heat rate*, *age*, and *Local Effectivenesss Factor*. This dissertation ranks 96 active natural gas power plants using 16 indicators across six metrics, with weights that vary by shape of justice. CPUC covers plants within CAISO, while this dissertation also includes plants in the other territories, such as LADWP and SMUD/BANC. CPUC also scores some generating units separately, whereas this dissertation's unit of analysis is the power plant. These differences limit a direct one-to-one comparison. The CPUC ranking is used here as a planning-based reference point to examine how the dissertation results align with or differ from existing gas non-retention exercise.

Several plants that rank early in the dissertation also receive high CPUC scores, although they do not all fall within CPUC's strict top 20 group. *Elk Hills Power Station* has a CPUC score of 83.33 and median ranks of 12.50 under utilitarianism, 15.00 under sufficientarianism, and 11.50 under egalitarianism. *Sunrise Power Station* has a CPUC score of 82.29 and median ranks of 14.00, 14.00, and 11.00, respectively. *Henrietta Peaker Project* has a CPUC score of 81.25 and median ranks of 5.00, 4.00, and 4.50. These cases show partial alignment between the two rankings because they perform poorly on the four shared indicators: Disadvantaged Community, NOx Emissions, Heat Rate, and Age of Plant. For example, both methods identify *the Henrietta Peaker Project* as being in a disadvantaged community and assign it moderate-to-high scores for heat rate and age.

However, the comparison should be read at the ranking level rather than as an exact indicator-level match, because CPUC and the dissertation operationalize shared indicators differently. For example, the CPUC scores NOx emissions as a rate weighted by capacity factor using EIA 2018 data, whereas this dissertation uses total NOx emissions in tons from the EPA Clean Air Markets Division for 2022. Similarly, CPUC measures DAC status on a four-tier proximity scale, while this dissertation uses a binary indicator. These differences mean that plant-level indicator comparisons between the two methods can't be executed with confidence

Greenleaf 2 Power Plant are ranked for early phaseout in the dissertation, but middle-tier priority for non-retention in the CPUC scores. Greenleaf 2 Power Plant has median ranks of 6.00 under utilitarianism, 6.00 under sufficientarianism, and 7.00 under egalitarianism, but has a CPUC score of 52.08. The dissertation ranking is influenced by low normalized values on indicators not included in the CPUC score. For example, *Greenleaf 2 Power Plant* has low values for Capacity (0.000), Total Employment (0.017), Job Share by Powerplant (0.000), Relative Compensation (0.000), Health Care and Social Assistance (0.007), and Religious and Civic Organizations (0.007). These values lower the dissertation composite scores, moving the plant toward the early

phaseout-priority end. They also show how indicators outside the CPUC score can change the ranking outcome.

Oxnard Power Station is the strongest divergence case. It ranks 1st for early phaseout among all three shapes of justice but has a CPUC score of 72.92, indicating a moderate priority for non-retention. Its normalized values on the overlapping indicators are low: Disadvantaged Community at 0.100, NOx Emissions at 0.465, and Age of Plant at 0.354, with only Heat Rate at 0.520 above 0.5. These values might explain why Oxnard does not appear in CPUC's highest-priority non-retention group. The non-overlapping indicators reinforce the case for early phaseout: Capacity (0.000), Total Employment (0.000), Healthcare and Social Assistance (0.070), Religious and Civic Organizations (0.077), LCOE (0.235), and Area Deprivation Index (0.100) are all low. Oxnard ranks first under the dissertation's framework because it performs poorly on both the overlapping and non-overlapping indicators. Including equity-relevant indicators can change the prioritization of plants that would otherwise be assessed using the CPUC's more limited set of scoring factors. A similar observation can be made broadly for *Kern River Cogeneration Power Station*. This powerplant's non-overlapping indicator values include Job Share by Powerplant (1.000), Energy Burden (1.000), and Average Annual Energy Cost (1.000). These high normalized values raise the plant's composite score in the dissertation framework and move it away from the early phaseout group.

The CPUC scoring is based on environmental, performance, and local reliability factors, including a disadvantaged community indicator. When additional equity-relevant indicators are included, such as affordability and community support, the ranking changes, as the divergence cases above show. This reflects that equity is one consideration among several in phaseout decisions.

This is relevant to current phaseout activity in California. There is increasing attention to the early phaseout of Oxnard Power Station (Becker 2023), and this dissertation's rankings support that prioritization, as Oxnard performs poorly across nearly all indicators and is ranked first for early phaseout under all three shapes of justice. Scattergood Power Station in El Segundo (Cota-Robles 2019) and Alamitos Energy Center (Becker 2023) have also received attention for phaseout but receive a lower and less consistent priority in this dissertation's rankings. This is because both plants have mixed indicator profiles, performing relatively better on indicators such as emissions, affordability, and community support, and poorly on others such as disadvantaged community status and area deprivation.

The comparison shows partial alignment between the dissertation ranking and CPUC's planning-based ranking. Some plants receive high priority under both approaches, but many move earlier or later when the dissertation's additional 12 indicators are included. A more granular indicator-level comparison is constrained by differences in data source, measurement year, and operationalization between the two frameworks, as illustrated by the NOx and DAC examples discussed above. The comparison supports the use of a broader, equity-relevant indicator set in

"expanding the set of information that can be used in planning and potential procurement in the future [by CAISO]." (California Public Utilities Commission, 2024a).

Chapter 4. Discussion and Implication

This chapter answers the three research questions that guided the dissertation. It explains what the answers mean for California's natural gas phaseout and how equity-centered decision-support models can enable more transparent, credible policymaking.

The dissertation develops a novel analytical architecture for incorporating equity into decision-support models. The architecture integrates two equity dimensions (distributional and contextual) and three shapes of justice (utilitarianism, sufficientarianism, and egalitarianism) into a multicriteria decision-making model. It also promotes transparency by preserving the uncertainty inherent in normative choices about the importance of indicators and metrics (or weights), rather than collapsing them into a single set of weights. The scholarly literature does not provide a ready-made set of weights or equity classifications, and direct stakeholder elicitation across many indicators, metrics, and shapes of justice proved infeasible. The dissertation, therefore, develops an LLM-RAG method for extracting indicator and metric weights from a defined corpus of academic literature, retaining disagreement across models as uncertainty in the weights. The application to California's 96 natural gas power plants shows that this architecture changes which plants are prioritized for phaseout, identifies where prioritization is robust to normative disagreement, and isolates where stakeholder deliberation would have the greatest effect.

The remainder of this chapter is organized as follows. Section 4.1 answers RQ1 on whether distributional, contextual, and procedural equity can be operationalized in a quantitative decision-support model. Section 4.2 answers RQ2 on whether incorporating those equity indicators produces meaningfully different phaseout rankings than a distributional-only model. Section 4.3 answers RQ3 on which plants converge across shapes of justice, and which diverge. Section 4.4 evaluates the LLM-based method used for weight extraction and equity classification. Section 4.5 identifies limitations and directions for future work.

4.1 Can distributional, contextual, and procedural equity be operationalized in a quantitative decision-support model, and what design requirements does doing so impose?

Distributional and contextual equity can be operationalized into a quantitative decision-support model, but procedural equity cannot be captured by indicators alone. The architecture presented in Chapter 2 describes the design requirements imposed by operationalizing equity. The architecture is valuable to both policy and the broader modeling literature. California has committed to multidimensional equity through policies such as Executive Order N-16-22 and related legislative efforts, such as CA Senate Bill 17. Yet the analytical tools that support phaseout and related energy planning decisions continue to operationalize equity primarily through distributional indicators.

This dissertation addresses that gap by introducing a novel analytical architecture that translates equity principles into actionable indicators and metrics. The architecture moves beyond the limited focus on distributional equity that characterizes much of the existing modeling literature by incorporating contextual equity through measurable indicators and by creating structured entry points for procedural equity through the multiple shapes of justice (analytical representations of different stakeholder value positions) and the capacity to incorporate stakeholder-derived weights (via participatory deliberations).

The design and development of the architecture revealed several specific insights about how equity can be integrated. For one, the LLM-based classification in Section 2.3 found that 12 of 16 indicators and 5 of 6 metrics were classified as distributional, 4 indicators and 1 metric were classified as contextual, and no indicator or metric was classified as procedural. This means that the indicators used in this dissertation capture distributional and contextual equity but do not directly operationalize procedural equity. The predominance of distributional equity among the indicators is consistent with the observation established in Chapter 1 that, where equity has been operationalized in energy planning, it has been primarily through distributional indicators. The analytical architecture makes this imbalance visible rather than leaving it implicit in the model design.

Procedural equity concerns the process through which decisions about siting, retirement, replacement, and benefit-sharing are made, and the agency that affected communities have within that process. This is a property of the decision process rather than an outcome or condition that can be measured at the plant or community level. As indicators measure outcomes and conditions, procedural equity falls outside what an indicator set can represent, which is why none of the 16 indicators was classified as procedural. This finding is itself a contribution of the architecture as it makes visible a limitation that other decision-support models may leave implicit. In the analytical architecture, procedural equity enters through two channels: the use of multiple shapes of justice, which represent different stakeholder value positions analytically, and the capacity to incorporate stakeholder-derived weights in future implementations. The current implementation uses LLM-extracted weights from academic literature rather than direct stakeholder input, thereby limiting its contribution to procedural equity. This limitation is discussed further in Section 4.4.

The classification process itself produced a notable discrepancy with implications for how equity components are operationalized. The *Living Standards* metric was classified as distributional by all three models, but its two constituent indicators, *Disadvantaged Community* and *Area Deprivation Index*, were classified as contextual. The metric was interpreted as describing outcomes distributed across populations, while the indicators were interpreted as describing place-specific conditions and neighborhood-level disadvantage. This discrepancy illustrates a practical challenge in operationalizing equity dimensions as the same variable can be read as a distributional outcome or a contextual condition depending on the level of aggregation at which it is defined. For analysts applying this approach, this means that equity component classification should be conducted at both the metric and indicator levels, and discrepancies

between the two should be reported. They are important candidates for stakeholder deliberation, since affected communities may be better positioned to determine whether a variable such as neighborhood disadvantage is an outcome or a condition. The analytical architecture is designed to support iterative refinement, allowing results to be updated as new stakeholder input or data becomes available without rebuilding the entire model.

4.2 Does incorporating those equity indicators into a decision-support model produce meaningfully different phaseout rankings than a model relying on distributional equity indicators alone?

Yes, for a given shape of justice, incorporating contextual equity indicators into the decision-support model produces meaningfully different phaseout rankings than a model that relies on distributional equity indicators alone.

As Section 3.2.2 demonstrates, under utilitarianism, two of the four most important drivers of ranking variability are contextual indicators, and together they account for approximately 32% of the total explained variability in plant rankings across weight configurations.[16] Under egalitarianism, one of the top four indicators is contextual.[17] Under sufficientarianism, all four top indicators are distributional. A model limited to distributional indicators would not have included the contextual indicators that drive variability in rankings under utilitarianism. It also would not have identified the *Community Impact* metric as the single largest source of ranking uncertainty under that shape of justice. The results confirm that, depending on the shape of justice, including contextual equity indicators changes which plants are identified as contested and needs further stakeholder deliberation.

The comparison with the CPUC Gas Capacity Not Retained scoring in Section 3.3 also shows that broadening the indicator set to include contextual equity indicators and varying the weights across shapes of justice leads to different phaseout rankings. Plants such as Oxnard Power Station and Greenleaf 2 Power Plant shift in phaseout priority when the broader indicator set is applied. As discussed in Section 3.3, Oxnard Power Station, Alamitos Energy Center, and Scattergood Power Station have all received attention for early phaseout in California. The model supports the prioritization of Oxnard Power Station, which ranks first for early phaseout under all three shapes of justice. However, this dissertation assigns a lower phaseout priority to Alamitos Energy Center and Scattergood Power Station because they have mixed indicator profiles, performing well on some indicators and poorly on others. As a result, their rankings are sensitive to how the indicators are weighted and shift across the three shapes of justice.

---

[16] Religious and Civic Organizations (importance 0.167, meaning this indicator explains 16.7% of the variation in a plant's ranking distribution caused by weight configurations), and Healthcare and Social Assistance (importance 0.151 meaning this indicator explains 15.1% of the variation in a plant's ranking distribution caused by weight configurations

[17] Disadvantaged Community (importance 0.093, meaning this indicator explains 9.3% of the variation in a plant's ranking distribution caused by weight configurations)

This suggests that incorporating different equity components into phaseout planning can change which plants are prioritized and where deliberation effort is directed. The RAG-LLM pipeline further supports this contribution by promoting transparency in how weights are assigned to indicators and metrics. In planning exercises (such as CPUC Gas Capacity Not Retained) where weights are fixed and the rationale for their selection is not documented, it is difficult for stakeholders to assess whether the resulting rankings reflect their priorities. The pipeline created in this dissertation addresses that concern. The weight assignments are traceable to a defined literature corpus, the priority tagging protocol distinguishes between literature-grounded and reasoning-based assignments, and the use of multiple shapes of justice makes explicit how different normative positions produce different rankings. This traceability allows stakeholders and policymakers to examine the basis for each weight assignment rather than accepting rankings as given.

This result has implications beyond California's NG Powerplant phaseout planning. Decision-support models in energy planning and environmental policy frequently rely on a fixed set of indicators (mostly distributive equity) evaluated within a single normative framework. The results of this dissertation show that such an approach cannot reveal whether the equity dimension driving the ranking outcome is distributional or contextual, because that answer changes depending on the shape of justice applied. By evaluating the same indicator data under three shapes of justice and retaining weight uncertainty across models, the analytical architecture developed in this dissertation makes that dependency visible and available for scrutiny. This is a novel capability that existing equity-aware decision-support models in the energy planning literature do not provide.

4.3 Which plants are ranked similarly across different shapes of justice, making them stronger candidates for robust policy action, and which plants are ranked differently, suggesting a need for further stakeholder deliberation?

4.3.1 Convergence and Divergence in Phaseout Rankings

Phaseout rankings are most stable among plants at the top and bottom of the phaseout priority and most variable among plants in the middle. The results reported in Section 3.2.1 show that 17 plants appear in the top 20 under all three shapes of justice, forming a stable early-phaseout group whose prioritization does not depend on the shape of justice. Further, a group of plants ranks consistently lower for phaseout under all three shapes of justice.

Between these two groups, plants with moderate median ranks exhibit the greatest variability in phase-out rankings, both across shapes of justice and within a given shape of justice across weight configurations. These three groups carry distinct implications for phaseout planning, and

the ability to distinguish between them is a novel output of the analytical architecture developed in this dissertation.

The convergence and divergence in rankings can be explained by the structure of each plant's underlying indicator data. Plants at the top of the phaseout priority have uniformly low normalized indicator values across most indicators, resulting in low composite scores regardless of the weight configuration applied.[18] Plants at the bottom exhibit uniformly high normalized indicator values, yielding high composite scores across all weight configurations.[19] In both cases, the ranking is stable because a plant whose indicator values point consistently in one direction will receive a similar composite score regardless of how the weights are configured. Ranking divergence is concentrated among plants with mixed indicator data profiles, where some indicators score high, and others score low for the same plant.[20] For these plants, the composite score is sensitive to the weights assigned to the metric and indicator, and those weights vary across shapes of justice.

---

[18] For example, *Oxnard Power Station* has a median rank of 1 under utilitarianism, sufficientarianism, and egalitarianism, with an IQR of 0 under all three shapes. *Hanford Energy Park Emergency Peaker Project* has a median rank of 2 and an IQR of 0 under all three shapes. Their normalized indicator values are consistently low across most indicators. For example, *Oxnard Power Station* has values of 0.000 for Capacity, 0.000 for Total Employment, 0.070 for Healthcare and Social Assistance, 0.077 for Religious and Civic Organizations, 0.235 for LCOE, 0.465 for NOx Emissions, 0.474 for SO2 Emissions, and 0.552 for CO2 Emissions. Since higher normalized values represent more desirable conditions, these low values produce low composite scores across all weight configurations. As a result, Oxnard Power station remains the highest phaseout-priority plant under all three shapes of justice

[19] For example, *Carlsbad Energy Center* has median ranks of 96.00 under utilitarianism, 95.50 under sufficientarianism, and 95.00 under egalitarianism, with IQR values of 0.25 under utilitarianism, 1.00 under sufficientarianism, and 2.00 under egalitarianism. Its normalized indicator values are high across most indicators, including Disadvantaged Community (0.900), Area Deprivation Index (0.800), SO2 Emissions (0.986), NOx Emissions (0.998), CO2 Emissions (0.977), Age of Plant (0.958), and Heat Rate (0.940). These high values produce high composite scores across weight configurations, keeping *Carlsbad Energy Center* near the lowest phaseout priority under all three shapes of justice.

[20] For example, THUMS Power Station has median ranks of 38.50 under utilitarianism, 52.00 under sufficientarianism, and 58.50 under egalitarianism, with IQR values of 25.50 under utilitarianism, 8.00 under sufficientarianism, and 17.00 under egalitarianism. The IQR of 25.50 under utilitarianism is the largest in the dataset. THUMS has normalized values of 1.000 for Healthcare and Social Assistance, Religious and Civic Organizations, Energy Burden, and Average Annual Energy Cost, but near-zero values for Capacity (0.004), Total Employment (0.020), Disadvantaged Community (0.000), and Area Deprivation Index (0.000). Its emissions indicators fall between these two groups, with SO2 Emissions at 0.626, NOx Emissions at 0.659, and CO2 Emissions at 0.886. This means that THUMS's phaseout priority depends on whether the weight configuration emphasizes community and affordability indicators, which produce a higher composite score and lower phaseout priority, or economic and living standards indicators, which produce a lower composite score and higher phaseout priority. Harbor Cogeneration Power Station shows a similar pattern, with median ranks of 42.00 under utilitarianism, 54.50 under sufficientarianism, and 63.00 under egalitarianism, and IQR values of 24.75 under utilitarianism, 4.00 under sufficientarianism, and 17.75 under egalitarianism. Its IQR of 24.75 under utilitarianism is the second largest in the dataset. Its indicator profile is similarly mixed, with values of 1.000 for Healthcare and Social Assistance and Religious and Civic Organizations alongside low values for Capacity (0.036), Total Employment (0.047), and Area Deprivation Index (0.000). Center Hybrid Power Station (median ranks of 47.50 under utilitarianism, 60.00 under sufficientarianism, and 71.00 under egalitarianism) and Long Beach Generation Power Station (median ranks of 63.00 under utilitarianism, 73.50 under sufficientarianism, and 80.00 under egalitarianism) show the same pattern of mixed indicator data profiles producing high ranking variability.

Thus, the rankings of powerplants converge where the indicator data leave little room for normative disagreement to affect the outcome. They diverge where normative choices about weights interact with mixed indicator profiles, and where these interactions have the greatest effect on phaseout prioritization. Note that the degree of this divergence also differs by the shape of justice, with utilitarianism producing the most ranking variability.[21]

This convergence-divergence pattern is itself the novel analytical product of the architecture. Decision-support models that produce a single ranking under a fixed set of weights cannot distinguish between plants whose prioritization is robust to normative disagreement and plants whose prioritization depends on it. The architecture developed in this dissertation produces that distinction by evaluating the same indicator data under multiple shapes of justice. The result is not a single recommended phaseout order but a structured map of where the ranking is stable, where it is contested, and what drives the contestation. This is a more useful output for policymakers than a single ranking because it identifies where decisions can proceed based on the model and where they require deliberation beyond it.

The three groups that emerge from this analysis give policymakers a structured basis for sequencing phaseout decisions. The first group consists of the 17 plants that rank in the top 20 under all three shapes of justice. These plants can be prioritized for early phaseout without needing to resolve disagreements about which shape of justice should guide the decision, because their prioritization is robust to that choice. This observation also hints toward a broader methodological contribution. By identifying plants that are prioritized consistently across utilitarianism, sufficientarianism, and egalitarianism, the method outlines a space for consensus among stakeholder groups that can be reached without requiring everyone to subscribe to a commonly accepted shape of justice. The method accommodates stakeholders who hold different perspectives and may have different reasons to agree on the same decision, so common understanding can be reached without agreement on underlying principles.

The second group consists of plants whose ranking shifts depending on the shape of justice and the weight configuration applied. For these plants, the model does not prescribe a ranking but identifies the source of disagreement, providing a structured basis for stakeholder deliberation. The global sensitivity analysis reported in Section 3.2.2 supports this by identifying which indicators drive variability in rankings under each shape of justice. Identifying which indicators drives ranking variability under each shape of justice helps direct stakeholder deliberation toward the weight assumptions that most affect phaseout prioritization. As reported in Section 3.2.2, the indicators that contribute most to ranking variability differ across shapes of justice.[22] In this study,

---

[21] Under utilitarianism, the average IQR across all 96 plants is 6.40, with the largest single IQR being 25.50 for THUMS Power Station. Under sufficientarianism, the average IQR drops to 3.88, with the largest being 12.00 for Elk Hills Power Station. Under egalitarianism, the average IQR is 4.11, with the largest being 17.75 for Harbor Cogeneration Power Station

[22] Under utilitarianism, Religious and Civic Organizations (0.167, meaning this single indicator explains 16.7% of the variation in how plants are ranked across weight configurations) and Healthcare and Social Assistance (0.151) are

for example, a policymaker adopting a utilitarian framing would benefit most from refining the weights assigned to community indicators, while one adopting an egalitarian framing would benefit most from refining assumptions about plant characteristics and employment indicators. Under sufficientarianism, importance is distributed more evenly and no targeted refinement on a small set of indicators would substantially reduce ranking variability.

The third group consists of plants that are late phaseout across all three shapes of justice. These plants can be deferred under any normative framing. Similar to the early-phaseout group, this group reflects agreement that holds across all three shapes of justice. Taken together, the plants at both ends of the priority ranking show where the analytical architecture outlines a space for consensus among stakeholder groups that can be reached without requiring everyone to harmonize different values and perspectives.

The results support the third contribution of the dissertation: equity-informed rankings can identify both robust phaseout candidates and plants whose prioritization depends on normative assumptions and weight uncertainty. The convergence-divergence pattern provides a structured way to allocate deliberative attention across a large fleet, directing stakeholder engagement where it would have the greatest effect on phaseout outcomes. The results also show that no indicator appears in the top quartile of importance under all three shapes of justice simultaneously, as reported in Table 5 in Section 3.2.2. This means that the targets of stakeholder engagement themselves depend on the shape of justice applied. A model that does not account for different stakeholder perspectives cannot surface this dependency. Section 4.4 evaluates the LLM-based method used in this dissertation and discusses this extension.

4.4 Commentary on the LLM-RAG-based Weight Extraction Method

The LLM-RAG pipeline developed in this dissertation provides a novel systematic and reproducible method for extracting structured normative information from academic literature. This innovation merits a dedicated discussion of its implications and limitations.

4.4.1 Novelty of the LLM-Based Weight Extraction Pipeline

Three open-source LLMs (Mistral-7B-Instruct-v0.3, Gemma-2-9b-it, Qwen2.5-7B-Instruct) each received the same prompts, definitions, and retrieved literature, ensuring that differences in outputs are attributable to model behavior rather than differences in input. Disagreements across models were retained as uncertainty rather than resolved into a single estimate, preserving information about where the literature provides clear guidance and where it does not. Priority 1 (Literature-based) and Priority 2 (LLM-Reasoning-based) tagging protocol made this distinction

---

the largest contributors. Under egalitarianism, Total Employment (0.167), Heat Rate (0.137), and Capacity (0.112) contribute the most. Under sufficientarianism, importance is distributed more evenly, with no single indicator exceeding 12.4%.

explicit.[23] The resulting pattern may also reflect the current state of the normative literature. Utilitarianism appears to have a more established empirical grounding in the energy justice literature, while sufficientarianism may have less. This gap could help identify where future normative research and stakeholder engagement might be most productive.

However, model selection introduces a source of variation that is independent of the literature and should be acknowledged as a limitation of the method. The three models differ in their ability to follow structured extraction prompts, as documented by the IFEval benchmark (Zhou et al., 2023). A model that has difficulty following the extraction prompt may default to reasoning-based assignments more frequently, not because the literature lacks support, but because the model fails to identify and report it.[24] Beyond instruction-following, LLMs carry inductive biases from their pretraining data that influence how they interpret and summarize text. The multi-model approach and the RAG constraint mitigate this by grounding outputs in a defined corpus and introducing diversity across model families, but they do not eliminate model-level bias entirely. The Priority 2 tagging is designed to surface rather than resolve this limitation. Documented sources of prompt sensitivity (Chatterjee et al., 2024) and inconsistent context utilization in retrieval-augmented generation (Hagström et al., 2024; Zhang et al., 2025) mean that the extracted weights carry model-specific noise that cannot be fully separated from the signal in the underlying literature. These limitations reinforce the position stated in Section 4.1 that the LLM-extracted weights are a proof of concept and should be supplemented or replaced by direct stakeholder input in policy applications.

Direct stakeholder engagement is the natural approach to addressing these limitations, in which practitioners, policymakers, or affected community members articulate what matters most to them and how they understand fairness. This is the gold standard for participatory, procedurally equitable modeling. However, this approach faces a practical scaling problem in the current context. With 16 indicators, 6 metrics, and 3 shapes of justice, the elicitation task requires respondents to make coherent, internally consistent normative judgments across a large and cognitively demanding combination of criteria. This dissertation attempted structured expert elicitation and found that the complexity of the task produced low response rates and unreliable, inconsistent responses. LLMs are therefore used not as a substitute for human judgment, but as a proof-of-concept method for extracting implicit normative weights from a defined body of academic literature, standing in where direct elicitation proved impractical.

This approach also points toward a broader methodological application that extends beyond the current implementation. Consider a series of community meetings in which residents discuss

[23] Across all assignments, 42% were tagged Priority 1 and 58% were tagged Priority 2. Utilitarianism had the highest share of Priority 1 assignments, consistent with a more established presence in literature. Sufficientarianism had the lowest, consistent with the limited empirical literature applying this principle to energy transition planning

[24] Qwen consistently produced the fewest Priority 1 assignments across all shapes of justice, as reported in Section 2.2.11. This pattern may reflect instruction-following limitations rather than a substantive difference in the underlying literature

what equity means to them, what they fear losing when a power plant closes, and what they believe fairness requires of the transition. Those conversations would be rich, context-specific, and genuinely participatory, but they would be expressed in natural, unstructured language rather than as numerical weights. An LLM pipeline like the one used in this study could be applied to such transcripts to map community voice onto shapes of justice and derive weights from them, translating qualitative stakeholder input into quantitative model parameters. The current implementation, which uses academic literature as its input corpus, is a proof of concept for that pipeline. It demonstrates that the extraction architecture is viable before applying it to richer, more participatory inputs. The academic literature serves as a reasonable first approximation of normative guidance, but it is explicitly understood as a placeholder for the stakeholder voice it is designed to eventually incorporate.

4.4.2 Limitation of the LLM-Based Weight Extraction Pipeline

However, several limitations of this approach should be acknowledged. First, academic literature is not a neutral or representative source of normative weights. It reflects the perspectives of researchers who skew toward particular institutions, disciplines, and countries, and who may not share the values or priorities of the communities most affected by natural gas phaseout. Weights extracted from this corpus inherit those biases and should be interpreted accordingly.

Second, the Priority 1 and Priority 2 distinction reveals something substantively important about the state of the literature. For several indicator-justice combinations, existing scholarship does not speak directly to the appropriate weight. Those gaps are themselves findings about where normative guidance is underdeveloped. However, they also imply that a portion of the weights in this model rests on model reasoning rather than on empirical or theoretical evidence, which is a weaker epistemic foundation. This should reduce confidence in those specific outputs.

Third, and most fundamentally, this approach approximates stakeholder input from a distance. Weights derived from academic literature may not reflect what affected communities value, fear, or prioritize. But as explained in preceding sections, the analytical architecture developed in this dissertation is explicitly designed to accommodate direct stakeholder input through participatory workshops or other deliberative processes. That extension is essential for any real-world policy application. The LLM extraction pipeline presented here should be understood as proof of concept and a methodological foundation, not as a finished substitute for genuine community voice.

4.5 Limitations and Future Work

The results of this dissertation are subject to several limitations. First, the rankings are conditional on the 16 indicators selected through the scoping literature review in Section 2.2.2; a different set of indicators would produce different rankings, so the results should be read as conditional on this indicator set rather than as a complete ranking of phaseout priority. Second, the rankings depend on the normalization rule described in Section 2.2.10, which rescales each

indicator to a [0,1] range using min–max normalization. An alternative method, such as z-score standardization, would rescale indicators by their mean and standard deviation rather than their minimum and maximum, which would change the normalized values and could change the rankings. Third, the findings are based on a single case study of California natural gas power plants; the architecture can inform the California case, but its performance in other policy settings is not established by this analysis. Fourth, as discussed in Section 2.1, the shapes of justice are represented through differences in weights rather than differences in aggregation form, so the rankings reflect how each shape of justice prioritizes indicators rather than the distinct aggregation forms associated with each principle. Fifth, as described in Section 2.2.1, the weighted sum model treats indicators as mutually independent, so interactions among indicators, such as between Energy Burden and the Area Deprivation Index, are not modeled and are not reflected in the composite scores. Sixth, the indicator data was collected at the powerplant level. As such they serve as a proxy for the equity outcomes experienced by the surrounding communities rather than a direct measure of those outcomes. As discussed in Section 4.1, the model also does not capture procedural equity, and as discussed in Section 4.4, the extracted weights depend on the reliability of the LLM pipeline and should be treated as proof of concept rather than as final weights for policy decisions.

Several directions for future work follow these limitations. The architecture can be applied to other geographies facing fossil-fuel phaseout decisions or to other types of fossil-fuel infrastructure, such as coal plants or refineries. The method would remain the same, but the indicators and metrics, their weights, and the equity classifications would need to be extracted again for the new application and will have the same limitations discussed previously. The indicator set can also be expanded to include dimensions not covered in this study. One approach is to recover indicators that were dropped for missing data in Section 2.2.2 using imputation techniques such as Bayesian hierarchical imputation, which estimates missing values by borrowing strength from similar counties, regions, plant types, or capacity classes and provides uncertainty intervals around those estimates. Stakeholder engagement can also be used to identify additional indicators for inclusion.

This also connects to the McDermott et al. (2013) equity framework as mentioned in Chapter 1. McDermott framework has four parameters: what counts as equity, who counts, why equity is sought, and how the equity parameters are determined. This thesis operationalizes the "what counts" parameter through distributional and contextual indicators, and it determines the "how" parameter through the literature-driven weight extraction and using them with WSM to create phaseout planning. The "who" and "why" parameters are addressed only implicitly. The "who" is treated as the communities affected by plant closure, and the "why" is treated as the disproportionate distribution of costs and benefits from the energy transition. These two parameters were not set directly because the unit of analysis in this thesis is the power plant rather than the community. Future work can address them directly by shifting the unit of analysis to the

community level and using stakeholder engagement to define which populations count and why equity is being sought.

Future work can also test whether the relative ordering of plants is robust to alternative normalization methods. This study uses min–max normalization, which rescales each indicator by its minimum and maximum, while an alternative such as z-score standardization rescales each indicator by its mean and standard deviation, which would change the normalized values and could alter the ranking of plants with close composite scores. Future work can also test the stability of the ranking to changes in the plant set, including how the ranking of the remaining plants shifts as the highest-priority or lowest priority plants are phased out and removed, and how this interacts with the sensitivity of the ranking to different shapes of justice.

The LLM-extracted weights can also be supplemented or replaced with weights derived from direct stakeholder engagement. Community meetings in which residents discuss what equity means to them, what they fear losing when a power plant closes, and what they believe fairness requires would produce input in natural, unstructured language rather than as numerical weights, and a pipeline similar to the one used in this study could map that input onto the shapes of justice and derive weights from it. Finally, future work can test alternative aggregation rules that reflect the definitions of the shapes of justice rather than representing them through weights alone. The random forest used in this study is a sensitivity analysis that identifies which indicators drive rank variability, not an aggregation rule.

As discussed in Section 2.1, this dissertation represents the shapes of justice through differences in weights within a weighted sum, while a strict egalitarian rule corresponds to a maximin or leximin function and sufficientarianism corresponds to a threshold rule. Testing these aggregation rules would require specifying a utility function for each affected individual, which was not feasible at the scale of this analysis, so future work would need a method for approximating these functions before such rules could be applied.

4.6 Conclusion

This chapter interpreted the results of the phaseout rankings under three shapes of justice, evaluated the analytical architecture and the LLM-based weight extraction method, and identified broader limitations and directions for future work. The findings demonstrate that distributional and contextual equity can be operationalized in a quantitative decision-support model that incorporates contextual indicators and produces meaningfully different rankings than a distributional-only approach, and that the rankings converge for plants with consistently high or low indicator profiles while diverging for plants with mixed profiles. The analytical architecture is designed to accommodate direct stakeholder input in future implementations, and the LLM-based extraction pipeline serves as a proof of concept for that extension. As California and other jurisdictions confront the equity implications of fossil-fuel phaseout, the analytical tools used to inform those decisions will shape whose interests are represented and how trade-offs are resolved. This

dissertation offers one approach to ensuring that those tools reflect the full complexity of equity rather than reducing it to a single dimension.

Appendix A. Selected List of Indicators Not Included in the Model

Table 8: Potential Indicator and Their Exclusion Criteria

Potential Indicator	Exclusion Criteria	Anticipated Effect on Metric
Reemployment Needs (number of workers under 45 requiring reemployment)	Lack Of Data - Age-stratified plant-level workforce data not available for California NG plants	Affects the Economic Impact metric, which may understate the transition burden at plants whose workers face lower chance of reemployment.
Loss of Tax Revenue and City Services	Lack of Data - Plant-level property tax contributions and service-specific fiscal impacts not reported at the level of analysis	Affects the Economic Impact metric, which does not capture fiscal losses to local governments.
Lower salary of a non-energy job	Overlapping - Captured by Relative Compensation, which measures the wage premium of power-plant employment relative to California per capita income	Not Applicable
Reliability (the ability of the power system to withstand uncontrolled events)	Not Applicable - A system-level attribute determined by grid composition and operations, not a plant-level indicator for phaseout prioritization. Also, Lack of Data at the unit of analysis.	Not Applicable
Loss of pensions or benefits due to Economic Decline	Lack of Data - Plant-level pension and benefit data not available	Affects the Economic Impact metric, which may understate the loss of non-wage benefits for displaced workers.

Loss of Small Business	Lack of Data - Plant-level or localized small-business impact data tied to individual powerplant closures not available	Affects the Economic Impact metric through lost local business activity, and may also affect the Community Impact metric, as business decline can weaken local support institutions.
Resource Curse/Lack of diversification	Lack of Data and construct is partially captured by Location Quotient, which measures local economic concentration in power-plant employment	Affects the Economic Impact metric, as it doesn't capture the long-term risk of economic overdependence on a single industry.
PM 2.5	Overlapping - PM2.5 is largely formed from SO2 and NOx precursor emissions (Sergi et al. 2020), already captured as indicators	Not Applicable
Increase in commute to non-aligned non-energy jobs	Lack of Data - Plant-level commute-pattern data under phaseout weight configurations is not available	Affects the Economic Impact metric, as it doesn't include the potential commuting costs for workers who move to non-energy jobs after closure.
Loss of Retail or commercial employment	Lack of Data - Localized plant-specific retail and commercial employment impact data not available	Affects the Economic Impact metric, as it doesn't include the indirect job losses in retail and commercial sectors linked to plant activity.
Adult education and training program per 1000 population	Lack of Data - County-level adult education programs tied to energy transition are not consistently available	Affects the Community Impact metric, as it impacts the local retraining and education capacity available to support workers through the transition.
Ratio of % college degree to % no high school diploma	Overlapping – Partially captured by Area Deprivation Index	Not Applicable

% that doesn't speak English as a second language	Overlapping - Captured by CalEnviroScreen 4.0 as *linguistic isolation*, defined as limited-English-speaking households where no one over age 14 speaks English well.	Not Applicable
ratio of % minority to nonminority population	Not directly applicable as a standalone indicator.	Not Applicable
% without disability	Not directly applicable as a standalone indicator.	Not Applicable
Job density (the ratio between the number of jobs in the area and the number of working-age residents)	Not directly applicable as a standalone indicator.	Not Applicable
Racial distribution of the workforce	Lack of Data – Powerplant level racial composition of the power-plant workforce not reported	Affects the Economic Impact metric, as it impacts how the job losses are distributed across racial groups.
Energy Insecurity	Overlapping - Closely related to Energy Burden	Not Applicable
Loss of Culture	No standardized, quantitative plant-level measure exists	Not Applicable
Decent living standard	Overlapping - Captured by the Area Deprivation Index, which ranks neighborhoods by socioeconomic disadvantage using factors related to income, education, employment, and housing quality.	Not Applicable

Job Retraining	Plant-level or community-specific job retraining program capacity data not available	Affects the Community Impact metric, as it impacts the local retraining capacity available to help displaced workers.

Note: This table lists candidate indicators identified during the scoping literature review but not included in the final model, the criterion used to exclude each, and, for indicators dropped for lack of data, the metric affected and the anticipated effect on that metric.

Appendix B. Cross-Model Consistency of LLM-Based Weight Assigned to indicators and Metrics

This appendix reports the model-by-model Priority 1 (Literature-based) and Priority 2 (LLM-Reasoning-based) weight assignment counts for the three LLMs used in Section 2.2.11, along with cross-model consistency statistics for indicator and metric weights under each shape of justice.

Table 9: Summary of Priority 1 vs Priority 2 patterns for indicators by model and justice type

Row Labels	Priority 1	Priority 2	Grand Total
Egalitarianism	**17**	**31**	**48**
Gemma	5	11	16
Mistral	10	6	16
Qwen	2	14	16
Sufficientarianism	**14**	**34**	**48**
Gemma	4	12	16
Mistral	8	8	16
Qwen	2	14	16
Utilitarianism	**21**	**27**	**48**
Gemma	5	11	16
Mistral	11	5	16
Qwen	5	11	16
Grand Total	**52**	**92**	**144**

Note: Each cell represents the number of weight assignments tagged as Priority 1 (Literature-based) or Priority 2 (LLM-Reasoning-based) by each model under each shape of justice. Each model assigns one weight per indicator per shape of justice. There are 16 indicators, producing 48 indicator-level assignments per shape of justice across the three models.

Table 9 reports the Priority 1 (Literature-based) and Priority 2 (LLM-Reasoning-based) tagging counts for indicator weights, broken down by model and shape of justice.

Across all three models, all three shapes of justice, and the indicator-level assignments shown in Table 1, Priority 1 and Priority 2 tagging is unevenly distributed across both models and shapes of justice. Mistral consistently produces the highest frequency of Priority 1 assignments, followed by Gemma, then Qwen, and this ordering holds across all three shapes of justice. Utilitarianism exhibits higher Priority 1 counts than the other two shapes of justice, consistent with utilitarianism having a more established literature base, which leads the models to identify sufficient textual support for evidence-based tagging more frequently.

These differences across models are likely attributable to two documented sources of variation in LLM behavior. First, LLM outputs can vary under minor, intent-preserving prompt changes, such as variations in wording or templates, a phenomenon that Chatterjee et al. (2024) characterize as prompt sensitivity rather than a task-level capability failure. Second, models differ in how faithfully they incorporate retrieved passages during retrieval-augmented generation. Hagström et al. (2024) show that real-world context utilization is more difficult than synthetic evaluations suggest, and Zhang et al. (2025) find that RAG models can generate outputs that either ignore retrieved context or inconsistently blend it with parametric knowledge, particularly when knowledge conflicts arise. This is why the study employs three models and retains disagreements as uncertainty rather than resolving them into a single estimate.

Table 10: Summary of Priority 1 vs Priority 2 patterns for metrics by model and justice type

Row Labels	Priority 1	Priority 2	Grand Total
Egalitarianism	**10**	**8**	**18**
Gemma	3	3	6
Mistral	5	1	6
Qwen	2	4	6
Sufficientarianism	**9**	**9**	**18**
Gemma	3	3	6
Mistral	5	1	6
Qwen	1	5	6
Utilitarianism	**13**	**5**	**18**
Gemma	4	2	6
Mistral	5	1	6
Qwen	4	2	6
Grand Total	**32**	**22**	**54**

Note: Each cell represents the number of weight assignments tagged as Priority 1 (Literature-based) or Priority 2 (LLM-Reasoning-based) by each model under each shape of justice. Each model assigns one weight per item (metric) per shape of justice. There 6 metrics, producing 18 metric-level assignments per shape of justice across the three models.

Table 10 reports the Priority 1 and Priority 2 tagging counts for metric weights, broken down by model and shape of justice.

The same model ordering observed at the indicator level holds for metric weights: Mistral most frequently identifies retrieved support for its assignments, followed by Gemma, then Qwen. Priority 1 tagging is more frequent at the metric level than at the indicator level across all models, and utilitarianism again exhibits the highest Priority 1 counts among the three shapes of justice. Combining the indicator and metric assignments in Tables 1 and 2, 84 of 198 total assignments (42%) were tagged as Priority 1 and 114 (58%) as Priority 2.

Table 11: Fleiss Kappa and ICC(2,k) by Shape of Justice for indicators and metrics

Shape of Justice	Fleiss Kappa (Indicators)	Fleiss Kappa (Metrics)	ICC(2,k) Indicators [95% CI]	ICC(2,k) Metrics [95% CI]
Egalitarianism	0.260s	0.478	0.789 [0.34, 0.93]	0.474 [−0.26, 0.90]
Utilitarianism	0.149	0.532	0.802 [0.54, 0.92]	0.882 [0.55, 0.98]
Sufficientarianism	0.324	0.065	0.605 [0.14, 0.85]	0.818 [0.34, 0.97]

Note: Table 10 reports two cross-model consistency statistics, Fleiss' Kappa and ICC(2,k), for indicator and metric weights under each shape of justice. Fleiss' Kappa (Fleiss 1971) measures the degree to which the three LLMs assign the same categorical weight (low, medium, or high) to the same indicator or metric, corrected for the level of agreement expected by chance alone. Values are interpreted following Landis and Koch (1977): below 0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement. ICC(2,k) measures whether averaging the three models' weight assignments produces a stable result; a high value indicates that if the same three models were run again on the same task, the averaged weights would be similar.

Table 11 reports two cross-model consistency statistics, Fleiss Kappa and ICC(2,k), for indicator and metric weights under each shape of justice (Fleiss, 1971; Shrout and Fleiss, 1979; Koo and Li, 2016).[25] ICC(2,k) was selected based on the assumption that the three LLMs represent a random sample from a broader population of open-source instruction-tuned models, and the results are intended to generalize beyond these specific models.

Fleiss' Kappa measures the degree to which the three LLMs assign the same categorical weight (low, medium, or high) to the same item, corrected for chance agreement. ICC(2,k) measures whether averaging the three models' weight assignments produces a stable result. A high ICC value indicates that if the same three models were run again on the same task, the averaged weights would be similar.

For indicator weights, Fleiss' Kappa values indicate fair agreement under sufficientarianism ($\kappa = 0.324$) and egalitarianism ($\kappa = 0.26$), and slight agreement under utilitarianism ($\kappa = 0.149$). ICC values are higher across all three shapes of justice, with egalitarianism (ICC = 0.789) and utilitarianism (ICC = 0.802) showing high consistency and sufficientarianism showing moderate consistency (ICC = 0.605). The gap between Fleiss' Kappa and ICC is expected. Kappa measures exact categorical agreement corrected for chance, while ICC captures whether the models rank items in a consistent relative order even when they assign different categorical labels. For example, under utilitarianism, the models frequently assign adjacent but not identical categories (e.g., medium vs. high), producing low exact agreement but high consistency.

For metric weights, Fleiss' Kappa indicates moderate agreement under utilitarianism ($\kappa = 0.532$) and egalitarianism ($\kappa = 0.478$), but slight agreement under sufficientarianism ($\kappa = 0.065$).

[25] Note that ICC(2,k) assumes that the numerical distances between weight categories are meaningful (the distance from low to medium equals the distance from medium to high). Although the categorical labels are converted into ordinal data with only three levels, the equal-spacing assumption is difficult to verify. Fleiss' Kappa is therefore treated as the primary agreement statistic with ICC as secondary.

ICC values are high under utilitarianism (ICC = 0.882) and sufficientarianism (ICC = 0.818), but lower under egalitarianism (ICC = 0.474). The wide confidence interval for egalitarianism metric weights (ICC 95% CI [−0.26, 0.90]) indicates that the aggregated metric-level profile under egalitarianism is less stable.

Fleiss' Kappa and ICC(2,k) together tell how much confidence to place in the weights used by the MCDM model. Where both statistics are high, the weights and the rankings they produce are driven primarily by the shape of justice and the underlying evidence in the literature. Where consistency is lower (indicator weights under sufficientarianism, metric weights under egalitarianism), the rankings are also sensitive to which model produced the weights. The MCDM model accounts for this by preserving weight uncertainty across multiple weight configurations rather than collapsing it into a single estimate, yielding 384 weight configurations under sufficientarianism, 64 under utilitarianism, and 32 under egalitarianism.

Note that ICC(2,k) assumes that numerical distances between weight categories are meaningful, i.e., that the distance from low to medium equals the distance from medium to high. Although in this case, the categorical label are converted into numerical ordinal data however with With only three ordinal levels, this is still assumption is difficult to verify. Therefore, Fleiss' Kappa is therefore reported as the primary agreement statistic.

Appendix C. Cross-Model Consistency of Equity Component Classification of Indicators and Metrics

This appendix reports the model-level outputs of the LLM-based equity classification described in Section 2.3.

Table 12: Equity Component Classification of Indicator by Model

Indicator	Gemma	Mistral	Qwen
Total Employment	Distributional	Distributional	Distributional
Location Quotient	Contextual	Distributional	Distributional
Relative Compensation	Distributional	Distributional	Distributional
SO2 Emissions	Distributional	Distributional	Distributional
NOx Emissions	Distributional	Distributional	Distributional
CO2 Emissions	Distributional	Distributional	Distributional
Healthcare & Social Assistance	Contextual	Distributional	Distributional
Religious & Civic Organizations	Contextual	Contextual	Contextual
Energy Burden	Distributional	Distributional	Distributional
Average Annual Energy Cost	Distributional	Distributional	Distributional
LCOE	Distributional	Distributional	Distributional
Capacity	Distributional	Distributional	Distributional
Age of Plant	Distributional	Distributional	Distributional
Heat Rate	Distributional	Distributional	Distributional
Disadvantaged Community	Contextual	Distributional	Distributional
Area Deprivation Index	Contextual	Distributional	Contextual

Note: Equity classifications are independent of the shape of justice. None of the models classified any indicator or metric as procedural equity.

Table 12 reports the equity component classifications assigned to each of the 16 indicators by Gemma, Qwen, and Mistral.

All three models agreed on the classification of 12 of 16 indicators. Disagreement occurred for four indicators: Location Quotient, Healthcare and Social Assistance, and Disadvantaged Community (Gemma assigned a contextual classification while Qwen and Mistral assigned a distributional classification), and Area Deprivation Index (Gemma and Qwen assigned a contextual classification and Mistral assigned a distributional classification).

Table 13: Equity Component Classification of Metric by Model

Metric	Gemma	Mistral	Qwen
Economic Impact	Distributional	Distributional	Distributional
Environmental Impact	Distributional	Distributional	Distributional
Community Impact	Contextual	Contextual	Contextual
Affordability	Distributional	Distributional	Distributional
Plant Characteristics	Contextual	Distributional	Distributional
Living Standard	Distributional	Distributional	Distributional

Note: Equity classifications are independent of the shape of justice. None of the models classified any indicator or metric as procedural equity.

Table 13 reports the equity component classifications assigned to each of the 6 metrics by Gemma, Qwen, and Mistral. All three models agreed on the classification of 5 of 6 metrics. The single disagreement occurred for Plant Characteristics, where Gemma assigned a contextual classification while Qwen and Mistral assigned a distributional classification.

Table 14: Summary of Priority 1 vs Priority 2 patterns for indicator equity classification by model

Model	Priority 1	Priority 2	Total
Gemma	14	2	16
Qwen	0	16	16
Mistrak	8	8	16
Total	22	26	48

Note: Each cell reports the number of equity classification assignments tagged as Priority 1 (Literature-based) or Priority 2 (LLM-Reasoning-based) by each model. Each model assigns one classification per indicator, producing 48 indicator-level assignments across the three models.

Table 14 reports the Priority 1 (Literature-based) and Priority 2 (LLM-Reasoning-based) tagging counts for the indicator classification task, broken down by model. Gemma produced the highest frequency of Priority 1 classifications, followed by Mistral, and then Qwen, which assigned all 16 indicator classifications to Priority 2. This model ordering matches the pattern observed for the weight extraction task in Appendix B

Table 15: Summary of Priority 1 vs Priority 2 patterns for metric equity classification by model

Model	Priority 1	Priority 2	Total
Gemma	5	1	6
Qwen	3	3	6
Mistral	4	2	6
Total	12	26	18

Note: Each cell reports the number of equity classification assignments tagged as Priority 1 (Literature-based) or Priority 2 (LLM-Reasoning-based) by each model. Each model assigns one classification per metric, producing 18 metric-level assignments across the three models.

Table 15 reports the Priority 1 and Priority 2 tagging counts for the metric classification task, broken down by model. Priority 1 tagging is more frequent at the metric level than at the indicator level across all three models, consistent with the pattern observed for the weight extraction task in Appendix B. Across both classification tasks, 34 of 66 total assignments (52%) were tagged as Priority 1 and 32 (48%) as Priority 2.

Appendix D. Top 20 Powerplants by Median for Early Phaseout Under Each Shape of Justice

Table 16: Top 20 Powerplants by Median for Early Phaseout Under Utilitarianism

Justice Type	Plant Name	Median	p25	p75	IQR
utilitarianism	Oxnard power station	1	1	1	0
utilitarianism	Hanford Energy Park Emergency Peaker Project	2	2	2	0
utilitarianism	Yuba City cogeneration Partners power station	3	3	4	1
utilitarianism	Walnut energy center	4	3	4	1
utilitarianism	Henrietta Peaker Project	5	5	6.25	1.25
utilitarianism	Greenleaf 2 power plant	6	6	6.25	0.25
utilitarianism	Midway Sunset cogeneration power station	8.5	7.75	9.5	1.75
utilitarianism	Crockett cogeneration Project power station	9	5.75	11.25	5.5
utilitarianism	Woodland power station	9	7.75	9	1.25
utilitarianism	El Centro power station	10	9.5	12	2.5
utilitarianism	Panoche Peaker power station	10	8.75	10.5	1.75
utilitarianism	Elk Hills power station	12.5	11	14.25	3.25
utilitarianism	Ellwood power station	13.5	9	19	10
utilitarianism	Sunrise power station	14	13	15	2
utilitarianism	Northern California Power Agency CT Project 2	14.5	13	16	3
utilitarianism	McClellan (CA) power station	16.5	15	19	4
utilitarianism	Gilroy power plant	18	14.75	21	6.25
utilitarianism	Sycamore cogeneration power station	18	16.75	22	5.25
utilitarianism	Los Medanos energy center	19	18	21	3
utilitarianism	Agua Mansa power plant	20	17	23	6
utilitarianism	King City power plant	20	16.5	22.25	5.75

Note: Plants are ranked by median rank across the 64 utilitarianism weight configurations described in Section 2.2.11. A lower median rank indicates a higher phaseout priority. P25 and P75 are the 25th and 75th percentiles of each plant's rank distribution, and IQR (P75 minus P25) measures the dispersion of that distribution across weight configurations.

Table 17: Top 20 Powerplants by Median for Early Phaseout Under Sufficientarianism

Justice Type	Plant Name	Median	p25	p75	IQR
sufficientarianism	Oxnard power station	1	1	1	0
sufficientarianism	Hanford Energy Park Emergency Peaker Project	2	2	2	0
sufficientarianism	Henrietta Peaker Project	4	3	5	2
sufficientarianism	Walnut energy center	4	4	5	1
sufficientarianism	Yuba City cogeneration Partners power station	4	3	4	1

sufficientarianism	Greenleaf 2 power plant	6	6	6	0
sufficientarianism	Woodland power station	7	7	8	1
sufficientarianism	El Centro power station	8	8	9	1
sufficientarianism	Midway Sunset cogeneration power station	9	8.75	11	2.25
sufficientarianism	Panoche Peaker power station	10	9	11	2
sufficientarianism	Crockett cogeneration Project power station	11	10	13	3
sufficientarianism	Northern California Power Agency CT Project 2	12	11	15	4
sufficientarianism	Sunrise power station	14	11	17	6
sufficientarianism	Elk Hills power station	15	11	23	12
sufficientarianism	Ellwood power station	16	15	19	4
sufficientarianism	King City power plant	16	15	18.25	3.25
sufficientarianism	McClellan (CA) power station	16	13	20	7
sufficientarianism	Sycamore cogeneration power station	17	14	21.25	7.25
sufficientarianism	Agua Mansa power plant	18	17	19	2
sufficientarianism	Almond power plant	19	17	21.25	4.25
sufficientarianism	Lodi energy center	19	16.75	21	4.25

Note: Plants are ranked by median rank across the 384 sufficientarianism weight configurations described in Section 2.2.11. A lower median rank indicates a higher phaseout priority. P25 and P75 are the 25th and 75th percentiles of each plant's rank distribution, and IQR (P75 minus P25) measures the dispersion of that distribution across weight configurations.

Table 18: Top 20 Powerplants by Median for Early Phaseout Under Egalitarianism

Justice Type	Plant Name	Median	p25	p75	IQR
egalitarianism	Oxnard power station	1	1	1	0
egalitarianism	Hanford Energy Park Emergency Peaker Project	2	2	2	0
egalitarianism	Walnut energy center	4	3	5	2
egalitarianism	Yuba City cogeneration Partners power station	4	3	4	1
egalitarianism	Henrietta Peaker Project	4.5	3	5	2
egalitarianism	Greenleaf 2 power plant	7	6	8	2
egalitarianism	Woodland power station	7	7	7	0
egalitarianism	El Centro power station	7.5	6	8	2
egalitarianism	Midway Sunset cogeneration power station	9	9	9	0
egalitarianism	Crockett cogeneration Project power station	11	10	13	3
egalitarianism	Sunrise power station	11	10	13	3
egalitarianism	Elk Hills power station	11.5	11	12.25	1.25
egalitarianism	Northern California Power Agency CT Project 2	13	12.75	15	2.25
egalitarianism	Panoche Peaker power station	14	12	15.25	3.25
egalitarianism	Sycamore cogeneration power station	15.5	14.75	16.75	2
egalitarianism	Almond power plant	17	15.5	18.5	3

egalitarianism	Ellwood power station	17.5	14.75	20	5.25
egalitarianism	King City power plant	17.5	16	18	2
egalitarianism	Lodi energy center	17.5	16	19	3
egalitarianism	McClellan (CA) power station	20.5	18	24	6

Note: Note: Plants are ranked by median rank across the 32 egalitarianism weight configurations described in Section 2.2.11. A lower median rank indicates a higher phaseout priority. P25 and P75 are the 25th and 75th percentiles of each plant's rank distribution, and IQR (P75 minus P25) measures the dispersion of that distribution across weight configurations.

Appendix E. Indicator Direction and Normalization Treatment

Table 19: Indicator and their normalization direction

Indicator	Higher raw value means	Normalization direction	Rationale
Total Employment	More desirable	Direct	Higher employment indicates greater local economic contribution from the plant.
Job Share by Powerplant	More desirable	Direct	Higher job share indicates greater local employment dependence on the plant.
Relative Compensation	More desirable	Direct	Higher relative compensation indicates stronger wage benefits from plant employment.
SO_2 Emissions	Less desirable	Reversed	Higher SO_2 emissions indicate greater local air pollution burden.
NO_x Emissions	Less desirable	Reversed	Higher NO_x emissions indicate greater local air pollution burden.
CO_2 Emissions	Less desirable	Reversed	Higher CO_2 emissions indicate greater greenhouse gas contribution.
Health Care and Social Assistance	More desirable	Direct	Higher values indicate greater local support capacity.
Religious and Civic Organizations	More desirable	Direct	Higher values indicate greater local civic and organizational capacity.
Energy Burden	Less desirable	Reversed	Higher energy burden indicates greater household energy cost vulnerability.

Average Annual Energy Cost	Less desirable	Reversed	Higher annual energy cost indicates greater household cost exposure.
Levelized Cost of Energy	Less desirable	Reversed	Higher LCOE indicates higher replacement or generation cost.
Capacity	More desirable	Direct	Higher capacity indicates greater generation contribution and larger reliability role.
Age of Plant	Less desirable	Reversed	Higher age indicates older infrastructure and greater retirement relevance.
Heat Rate	Less desirable	Reversed	Higher heat rate indicates lower thermal efficiency.
Disadvantaged Community	Less desirable	Reversed	Higher disadvantage indicates greater existing community vulnerability.
Area Deprivation Index	Less desirable	Reversed	Higher deprivation indicates greater socioeconomic disadvantage.

Note: This table documents the direction assigned to each of the 16 indicators before normalization and the corresponding normalization treatment, as described in Section 2.2.10. Direction reflects whether a higher raw value represents a more desirable or less desirable plant-level condition for phaseout planning. Indicators for which a higher raw value is more desirable were normalized directly, while indicators for which a higher raw value is less desirable were reverse-coded. After this transformation, all indicators are on a common [0,1] scale on which higher normalized values represent more desirable plant-level conditions, so that a higher composite score corresponds to a lower phaseout priority.

Appendix F. Prompts

Prompt for Weight Extraction:

```
header = f"""
You are an evidence-bound climate change and energy expert tasked with ASSESSING the importance of specific METRICS or INDICATORS by assigning exactly ONE weight from the allowed set {json.dumps(ALLOWED_WEIGHTS)}.

IMPORTANT CONTEXT:
- Assessing means selecting a single weight label (High, Medium, Low, or None) that represents how strongly this metric or indicator SHOULD influence decisions about WHICH NATURAL GAS POWER PLANTS ARE PRIORITIZED FOR PHASEOUT.
- You are NOT evaluating whether a metric or indicator is generally valuable, interesting, or morally relevant.
- You ARE evaluating whether excluding this metric or indicator would plausibly change the RANKING or PRIORITIZATION of plants under the given justice definition.

DEFINITION LOCK:
- Use ONLY the exact definitions provided below.
- If a chunk contains different terminology, ignore it and use the provided definitions only.
- Justice type in this run: {justice_type}
- Allowed weight labels: {json.dumps(ALLOWED_WEIGHTS)}
- Priority levels allowed: {json.dumps(ALLOWED_PRIORITIES)}

Authoritative justice definition:
{justice_type}: {justice_defs[justice_type]}

Item to evaluate:
- item_type: {item_type}
- item_name: {item_name}
- item_definition (authoritative): {item_definition}

Retrieved evidence (K={len(retrieved_chunks)}):
""".strip()

    parts = [textwrap.dedent(header)] # VIMP: It removes common leading whitespace from every line in a multi-line string.

    for ch in retrieved_chunks:
        parts.append("--- CHUNK BEGIN ---")
        parts.append(f"chunk_id: {ch['chunk_id']}")
        parts.append("text:")
        parts.append(ch["text"])
```

```
        parts.append("--- CHUNK END ---")

    task = f"""
---
TASK:

WEIGHT DECISION RULE:
You are assessing the importance of the following item_type: {item_type}

- item_name: {item_name}
- item_definition (authoritative): {item_definition}

For the specified justice type (Utilitarianism, Egalitarianism, or Sufficientarianism), assign
exactly ONE weight ∈ ["High","Medium","Low"] to the item_name: {item_name}.

The weight should reflect whether this {item_name} should be used when ranking or prioritizing
natural gas power plants for phaseout under the {justice_type}.

DEFAULT RULE:
- If you cannot clearly articulate how this metric or indicator would change the prioritization or
ranking of plants, assign Low (not Medium).

Assign exactly one weight for the item under {justice_type}:
- weight ∈ {json.dumps(ALLOWED_WEIGHTS)}
- priority ∈ {json.dumps(ALLOWED_PRIORITIES)}
- reasoning (≤ {max_reason_words} words)

EVIDENCE & PRIORITY RULES (ONLY TWO MODES):

Priority 1 — EVIDENCE-LED (Chunk-based):
- You may use priority = 1 ONLY if the retrieved evidence contains explicit textual support for
the item_definition.
- Explicit textual support means the chunk text literally mentions a concrete element named in
the item_definition.
- For priority = 1:
  - Your reasoning MUST be based ONLY on what is stated in the cited chunks.
  - cited_chunks MUST contain BETWEEN 3 AND 5 chunks that meet the above criterion.
  - If you cannot identify at least 3 such chunks, priority = 1 is NOT allowed.

Priority 2 — THEORY-LED (General reasoning):
- Use priority = 2 ONLY if NO retrieved chunk provides explicit textual support for the
item_definition.
- For priority = 2:
  - cited_chunks MUST be an empty list [].
  - Your reasoning MUST rely on the justice definition and general normative reasoning using
the retreived evidence if helpful.
```

```
CITATION RULES (MANDATORY):
- If priority = 1:
  - cited_chunks MUST contain BETWEEN 3 AND 5 objects.
  - Each object MUST have EXACT keys: "chunk_id", "chunk_text".
  - chunk_id MUST match one of the retrieved chunk_id values shown above.
  - chunk_text MUST be an empty string "" (leave it blank)..
- If priority = 2:
  - cited_chunks MUST be [].

OUTPUT RULES (CRITICAL):
- Output ONLY a single valid JSON object.
- Do NOT include headings, explanations, labels, or any text outside the JSON.
- Do NOT add any keys beyond those defined in the schema.

OUTPUT JSON SCHEMA:
{{
  "item_type": "{item_type}",
  "item_name": "{item_name}",
  "justice_type": "{justice_type}",
  "weight": "High|Medium|Low|None",
  "priority": 1|2,
  "reasoning": "string",
  "cited_chunks": []
}}

HARD STOP:
- Begin your response with the first character '{{'.
- End your response with the matching '}}'.
""".strip()

    parts.append(textwrap.dedent(task))
    return "\n".join(parts)

print("Prompt builder ready.")
```

Prompt for Equity Classification:

```
header = f"""
You are an evidence-bound climate change and energy expert tasked with CLASSIFYING each
METRIC or INDICATOR into exactly ONE equity dimension from the allowed set
{json.dumps(ALLOWED_EQUITY_TYPES)}.

IMPORTANT CONTEXT:
- Classification means identifying whether this metric or indicator primarily relates to
DISTRIBUTIONAL, CONTEXTUAL, or PROCEDURAL equity.
```

- You are NOT evaluating whether a metric or indicator is generally valuable, interesting, or morally relevant.
- You ARE determining which equity dimension best captures what this metric or indicator measures.

DEFINITION LOCK:
- Use ONLY the exact definitions provided below.
- If a chunk contains different terminology, ignore it and use the provided definitions only.
- Allowed equity types: {json.dumps(ALLOWED_EQUITY_TYPES)}
- Priority levels allowed: {json.dumps(ALLOWED_PRIORITIES)}

Authoritative equity dimension definitions:

Distributional Equity:
{equity_type_defs["Distributional"]}

Procedural Equity:
{equity_type_defs["Procedural"]}

Contextual Equity:
{equity_type_defs["Contextual"]}

Item to classify:
- item_type: {item_type}
- item_name: {item_name}
- item_definition (authoritative): {item_definition}

Retrieved evidence (K={len(retrieved_chunks)}):
""".strip()

    parts = [textwrap.dedent(header)] # VIMP: It removes common leading whitespace from every line in a multi-line string.

    for ch in retrieved_chunks:
        parts.append("--- CHUNK BEGIN ---")
        parts.append(f"chunk_id: {ch['chunk_id']}")
        parts.append("text:")
        parts.append(ch["text"])
        parts.append("--- CHUNK END ---")

    task = f"""

TASK:

EQUITY CLASSIFICATION RULE:
You are classifying the following item:

- item_type: {item_type}
- item_name: {item_name}
- item_definition (authoritative): {item_definition}

Assign this indicator to exactly ONE equity dimension: ["Distributional", "Contextual", "Procedural"]

CLASSIFICATION GUIDANCE:
- Distributional: Choose this if the indicator measures outcomes, resource allocation, or distribution of benefits/burdens (e.g., emissions, health impacts, economic effects on specific populations).
- Procedural: Choose this if the indicator relates to decision-making processes, stakeholder engagement, transparency, or fairness of methods (e.g., community participation, consultation processes).
- Contextual: Choose this if the indicator reflects local conditions, historical factors, cultural considerations, or place-specific vulnerabilities (e.g., demographics, historical pollution burden, cultural resources).

DEFAULT RULE:
- If an indicator could fit multiple categories, choose the PRIMARY dimension that best captures what it measures.
- Most quantitative outcome measures → Distributional
- Process/engagement measures → Procedural
- Place-based/demographic measures → Contextual

Assign exactly one equity type:
- equity_type ∈ {json.dumps(ALLOWED_EQUITY_TYPES)}
- priority ∈ {json.dumps(ALLOWED_PRIORITIES)}
- reasoning (≤ {max_reason_words} words)

EVIDENCE & PRIORITY RULES (ONLY TWO MODES):

Priority 1 — EVIDENCE-LED (Chunk-based):
- You may use priority = 1 ONLY if the retrieved evidence contains explicit textual support for the item_definition.
- Explicit textual support means the chunk text literally mentions a concrete element named in the item_definition.
- For priority = 1:
  - Your reasoning MUST be based ONLY on what is stated in the cited chunks.
  - cited_chunks MUST contain BETWEEN 3 AND 5 chunks that meet the above criterion.
  - If you cannot identify at least 3 such chunks, priority = 1 is NOT allowed.

Priority 2 — THEORY-LED (General reasoning):
- Use priority = 2 ONLY if NO retrieved chunk provides explicit textual support for the item_definition.

```
- For priority = 2:
  - cited_chunks MUST be an empty list [].
  - Your reasoning MUST rely on the equity definitions and general normative reasoning using the retrieved evidence if helpful.

CITATION RULES (MANDATORY):
- If priority = 1:
  - cited_chunks MUST contain BETWEEN 3 AND 5 objects.
  - Each object MUST have EXACT keys: "chunk_id", "chunk_text".
  - chunk_id MUST match one of the retrieved chunk_id values shown above.
  - chunk_text MUST be an empty string "" (leave it blank).
- If priority = 2:
  - cited_chunks MUST be [].

OUTPUT RULES (CRITICAL):
- Output ONLY a single valid JSON object.
- Do NOT include headings, explanations, labels, or any text outside the JSON.
- Do NOT add any keys beyond those defined in the schema.

OUTPUT JSON SCHEMA:
{{
  "item_type": "{item_type}",
  "item_name": "{item_name}",
  "equity_type": "Distributional|Contextual|Procedural",
  "priority": 1|2,
  "reasoning": "string",
  "cited_chunks": []
}}

HARD STOP:
- Begin your response with the first character '{{'.
- End your response with the matching '}}'.
""".strip()

    parts.append(textwrap.dedent(task))
    return "\n".join(parts)

print("Prompt builder for Mistral is ready.")
```

References

1. Governor's Office of Planning and Research, "Equitable & Resilient Communities," in *General Plan Guidelines* (2017), 195–97, PDF, https://lci.ca.gov/docs/OPR_C5_final.pdf
2. Office of the Governor of California. 2022. *Executive Order N-16-22*. September 13, 2022. PDF. https://www.gov.ca.gov/wp-content/uploads/2022/09/9.13.22-EO-N-16-22-Equity.pdf
3. California Legislature. 2022. "SB 17 (2021–2022 Reg. Sess.): Racial Equity Commission." https://leginfo.legislature.ca.gov/faces/billCompareClient.xhtml?bill_id=202120220SB17&showamends=false
4. The White House. 2024. "Advancing Equity and Racial Justice through the Federal Government." February 14, 2024. https://www.whitehouse.gov/equity/. Accessed December 19, 2024. (Page no longer available)
5. Chu, Eric K., and C. E. Cannon. 2021. "Equity, Inclusion, and Justice as Criteria for Decision-Making on Climate Adaptation in Cities." *Current Opinion in Environmental Sustainability* 51: 85–94. https://doi.org/10.1016/j.cosust.2021.02.009
6. Minow, Martha. 2021. "Equality vs. Equity." *American Journal of Law and Equality* 1: 167–193. https://doi.org/10.1162/ajle_a_00019
7. Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *arXiv preprint* arXiv:2005.11401. https://arxiv.org/abs/2005.11401
8. Tarekegne, Bethel W., Barbara Grace R. Pennell, Danielle C. Preziuso, and Rebecca S. O'Neil. 2021. *Review of Energy Equity Metrics*. Pacific Northwest National Laboratory. https://doi.org/10.2172/1830804
9. Crosman, Katherine M., Edward H. Allison, Yoshitaka Ota, Andrés M. Cisneros-Montemayor, Gerald G. Singh, Wilf Swartz, Megan Bailey, et al. 2022. "Social Equity Is Key to Sustainable Ocean Governance." *npj Ocean Sustainability* 1 (1): 4. https://doi.org/10.1038/s44183-022-00001-7

10. Lewis, Elizabeth O., Don MacKenzie, and Jacqueline Kaminsky. 2021. “Exploring Equity: How Equity Norms Have Been Applied Implicitly and Explicitly in Transportation Research and Practice.” *Transportation Research Interdisciplinary Perspectives* 9: 100332. https://doi.org/10.1016/j.trip.2021.100332
11. Grasso, M. A normative ethical framework in climate change. *Climatic Change* **81**, 223–246 (2007). https://doi.org/10.1007/s10584-006-9158-7
12. Chu, Eric K., and Clare E. B. Cannon. 2021. “Equity, Inclusion, and Justice as Criteria for Decision-Making on Climate Adaptation in Cities.” *Current Opinion in Environmental Sustainability* 51: 85–94. https://doi.org/10.1016/j.cosust.2021.02.009
13. Maia, Rodrigo Gomes Távora, and Katia Cristina Garcia. 2023. “What They Say, What They Do and How They Do It: An Evaluation of the Energy Transition and GHG Emissions of Electricity Companies.” *Energy Policy* 174: 113462. https://doi.org/10.1016/j.enpol.2023.113462
14. Yang, Yu, Siyou Xia, Ping Huang, and Junxi Qian. 2024. “Energy Transition: Connotations, Mechanisms and Effects.” *Energy Strategy Reviews* 52: 101320. https://doi.org/10.1016/j.esr.2024.101320
15. Carley, Sanya, and David M. Konisky. 2020. “The Justice and Equity Implications of the Clean Energy Transition.” *Nature Energy* 5 (8): 569–577. https://doi.org/10.1038/s41560-020-0641-6
16. Baker, Erin, Anna P. Goldstein, and Inês M. L. Azevedo. 2021. “A Perspective on Equity Implications of Net Zero Energy Systems.” *Energy and Climate Change* 2: 100047. https://doi.org/10.1016/j.egycc.2021.100047
17. Raimi, Daniel, Monika Ehrman, and Andrew Curley. 2023. “Native Nations in the Energy Transition.” *Resources for the Future*, November 28, 2023. https://www.resources.org/common-resources/native-nations-in-the-energy-transition/
18. Chen, Huiting, Sung Gheel Jang, Yan Zhang, and Yaolin Liu. 2024. “Clean Energy for All? Mapping Inequity Potential in the Clean Energy Transition in the United States.” *Energy Research & Social Science* 108: 103400. https://doi.org/10.1016/j.erss.2023.103400

19. Kalra, Nidhi, Swaptik Chowdhury, Kelly Klima, and Liam Regan. 2022. *Equity Metrics for Climate Adaptation in the Electricity Sector*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1721-1.html
20. Barlow, Jay, Rebecca Tapio, and Bethel Tarekegne. 2022. "Advancing the State of Energy Equity Metrics." *The Electricity Journal* 35 (10): 107208. https://doi.org/10.1016/j.tej.2022.107208
21. Crespo Montañés, Cristina, Eric O'Shaughnessy, Dominic Bednar, Destenie Nock, Sanya Carley, Daniel M. Kammen, and David M. Konisky. 2023. "Enabling and Centering Equity and Justice in Clean Energy Transition Research." *Joule* 7 (3): 437–441. https://doi.org/10.1016/j.joule.2023.02.005
22. Heffron, Raphael J., Darren McCauley, and Benjamin K. Sovacool. 2015. "Resolving Society's Energy Trilemma through the Energy Justice Metric." *Energy Policy* 87: 168–176. https://doi.org/10.1016/j.enpol.2015.08.033
23. Hall, Sarah Marie, Sarah Hards, and Harriet Bulkeley. 2013. "New Approaches to Energy: Equity, Justice and Vulnerability. Introduction to the Special Issue." *Local Environment* 18 (4): 413–421. https://doi.org/10.1080/13549839.2012.759337
24. California Energy Commission. n.d. "Energy Equity Indicators." Accessed April 12, 2026. https://www.energy.ca.gov/rules-and-regulations/energy-suppliers-reporting/clean-energy-and-pollution-reduction-act-sb-350-3
25. California Energy Commission. 2023. *Justice Access Equity Diversity Inclusion (JAEDI) Framework*. PDF. https://www.energy.ca.gov/sites/default/files/2023-11/CEC-JAEDI-Framework_ada.pdf
26. Office of Environmental Health Hazard Assessment. n.d. "SB 535 Disadvantaged Communities." Accessed February 21, 2025. https://oehha.ca.gov/calenviroscreen/sb535
27. Sovacool, Benjamin K., and Michael H. Dworkin. 2015. "Energy Justice: Conceptual Insights and Practical Applications." *Applied Energy* 142: 435–444. https://doi.org/10.1016/j.apenergy.2015.01.002
28. Heffron, Raphael J., and Darren McCauley. 2017. "The Concept of Energy Justice across the Disciplines." *Energy Policy* 105: 658–667. https://doi.org/10.1016/j.enpol.2017.03.018

29. McCauley, Darren A., Raphael J. Heffron, Hannes Stephan, and Kirsten Jenkins. 2013. "Advancing Energy Justice: The Triumvirate of Tenets." *International Energy Law Review* 32 (3): 107–10.

30. Sovacool, Benjamin K., Matthew Burke, Lucy Baker, Chaitanya Kumar Kotikalapudi, and Harriet Wlokas. 2017. "New Frontiers and Conceptual Frameworks for Energy Justice." *Energy Policy* 105: 677–91. https://doi.org/10.1016/j.enpol.2017.03.005.

31. Sovacool, Benjamin K., Raphael J. Heffron, Darren McCauley, and Andreas Goldthau. 2016. "Energy Decisions Reframed as Justice and Ethical Concerns." *Nature Energy* 1 (5): Article 16024. https://doi.org/10.1038/nenergy.2016.24.

32. Jenkins, Kirsten E. H., Benjamin K. Sovacool, Niek Mouter, Nick Hacking, Mary-Kate Burns, and Darren McCauley. 2021. "The Methodologies, Geographies, and Technologies of Energy Justice: A Systematic and Comprehensive Review." *Environmental Research Letters* 16 (4): 043009. https://doi.org/10.1088/1748-9326/abd78c.

33. McDermott, Melanie, Sango Mahanty, and Kate Schreckenberg. 2013. "Examining Equity: A Multidimensional Framework for Assessing Equity in Payments for Ecosystem Services." *Environmental Science & Policy* 33: 416–27. https://doi.org/10.1016/j.envsci.2012.10.006

34. Energy Equity Project. n.d. *Energy Equity Project*. Accessed April 13, 2026. https://energyequityproject.org/.

35. American Council for an Energy-Efficient Economy. n.d. "Energy Equity." *ACEEE*. Accessed April 13, 2026. https://www.aceee.org/topic/energy-equity.

36. California Energy Commission. 2023. *Justice Access Equity Diversity Inclusion (JAEDI) Framework*. February 2023. CEC-100-2022-001-CMF-APA. https://www.energy.ca.gov/sites/default/files/2023-11/CEC-JAEDI-Framework_ada.pdf.

37. Abdul-Khabir, Fatima. 2023. "Emerging Energy Technologies Part 2: Applying an Equity Framework." *The Greenlining Institute*, September 19, 2023. https://greenlining.org/2023/emerging-energy-technologies-part-2-applying-an-equity-framework/.

38. Forrester, Sydney P., and Andrew J. Satchwell. 2023. *Developing an Equity Framework for State Regulatory Decision-Making*. Lawrence Berkeley National Laboratory, August 2023. https://live-lbl-eta-publications.pantheonsite.io/sites/default/files/equity_in_reg_decision_making.pdf

39. Susser, Daniel, Adam Ceglarz, Hubert Gasching, Vangelis Stavrakas, Alexandros Flamos, Giorgos Giannakidis, and Johan Lilliestam. 2021. "Model-Based Policymaking or Policy-Based Modelling? How Energy Models and Energy Policy Interact." *Energy Research & Social Science* 75: 101984. https://doi.org/10.1016/j.erss.2021.101984

40. Mara, Thierry Alex, Ivano Azzini, Giulia Listorti, and Rossana Rosati. 2020. *Uncertainty and Sensitivity Analysis for Policy Decision Making: An Introductory Guide.* Luxembourg: Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/ff3cf22c-35e1-11eb-b27b-01aa75ed71a1/language-en

41. Prasinos, Marios, Ioannis Basdekis, Marco Anisetti, George Spanoudakis, Dimitrios Koutsouris, and Ernesto Damiani. 2022. "A Modelling Framework for Evidence-Based Public Health Policy Making." *IEEE Journal of Biomedical and Health Informatics* 26 (5): 2388–2399. https://doi.org/10.1109/JBHI.2022.3142503

42. James, Lyndon P., Joshua A. Salomon, Caroline O. Buckee, and Nicolas A. Menzies. 2021. "The Use and Misuse of Mathematical Modeling for Infectious Disease Policymaking: Lessons for the COVID-19 Pandemic." *Medical Decision Making* 41 (4): 379–385. https://doi.org/10.1177/0272989X21990391

43. Jafino, Bramka Arga, Jan H. Kwakkel, Frans Klijn, Nguyen Viet Dung, Hedwig van Delden, Marjolijn Haasnoot, and Edwin H. Sutanudjaja. 2021. "Accounting for Multisectoral Dynamics in Supporting Equitable Adaptation Planning: A Case Study on the Rice Agriculture in the Vietnam Mekong Delta." *Earth's Future* 9 (5): e2020EF001939. https://doi.org/10.1029/2020EF001939

44. Yang, Guang, Matteo Giuliani, and Andrea Castelletti. 2023. "Operationalizing Equity in Multipurpose Water Systems." *Hydrology and Earth System Sciences* 27 (1): 69–81. https://doi.org/10.5194/hess-27-69-2023.

45. Jafino, Bramka Arga, Jan H. Kwakkel, and Frans Klijn. 2022. “Evaluating the Distributional Fairness of Alternative Adaptation Policies: A Case Study in Vietnam’s Upper Mekong Delta.” *Climatic Change* 173: 17. https://doi.org/10.1007/s10584-022-03395-y.

46. Ciullo, Alessio, Jan H. Kwakkel, Karin M. de Bruijn, Neelke Doorn, and Frans Klijn. 2020. “Efficient or Fair? Operationalizing Ethical Principles in Flood Risk Management: A Case Study on the Dutch-German Rhine.” *Risk Analysis* 40 (9): 1844–1862. https://doi.org/10.1111/risa.13527

47. Ma, Gang, Xu Li, and Jianping Zheng. 2020. “Efficiency and Equity in Regional Coal De-Capacity Allocation in China: A Multiple Objective Programming Model Based on Gini Coefficient and Data Envelopment Analysis.” *Resources Policy* 66: 101621. https://doi.org/10.1016/j.resourpol.2020.101621.

48. Heleno, Miguel, Benjamin Sigrin, Natalie Popovich, Jenny Heeter, Anjuli Jain Figueroa, Michael Reiner, and Tony Reames. 2022. “Optimizing Equity in Energy Policy Interventions: A Quantitative Decision-Support Framework for Energy Justice.” *Applied Energy* 325: 119771. https://doi.org/10.1016/j.apenergy.2022.119771

49. Cardiff, P. 2020. “Aggregation Bias, Local Estimates and the Devil.” *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3739892

50. Schumacher, Ingmar. 2018. “The Aggregation Dilemma in Climate Change Policy Evaluation.” *Climate Change Economics* 9 (3): 1–20. https://doi.org/10.1142/S2010007818500082

51. Hyatt, Josh, and Jeffrey Gruenglas. 2023. “Ethical Considerations in Organizational Conflict.” In *Conflict Management: Organizational Happiness, Mindfulness, and Coping Strategies*. London: IntechOpen. Published September 20, 2023. https://doi.org/10.5772/intechopen.1002645.

52. Williams, Simone A., Susanna Eden, Sharon B. Megdal, and Valerisa Joe-Gaddy. 2023. “Diversity, Equity, Inclusion, and Justice in Water Dialogues: A Review and Conceptualization.” *Journal of Contemporary Water Research & Education* 177 (1): 113–139. https://doi.org/10.1111/j.1936-704X.2022.3386.x.

53. Müller, Daniel, and Sander Renes. 2021. “Fairness Views and Political Preferences: Evidence from a Large and Heterogeneous Sample.” *Social Choice and Welfare* 56: 679–711. https://doi.org/10.1007/s00355-020-01289-5

54. Zimm, Caroline, Kian Mintz-Woo, Elina Brutschin, Susanne Hanger-Kopp, Roman Hoffmann, Jarmo S. Kikstra, Michael Kuhn, Jihoon Min, Raya Muttarak, Shonali Pachauri, Omkar Patange, Keywan Riahi, and Thomas Schinko. 2024. “Justice Considerations in Climate Research.” *Nature Climate Change* 14 (1): 22–30. https://doi.org/10.1038/s41558-023-01869-0

55. Giang, Amanda, Morgan R. Edwards, Sarah M. Fletcher, Rivkah Gardner-Frolick, Rowenna Gryba, Jean-Denis Mathias, Camille Venier-Cambron, John M. Anderies, Emily Berglund, Sanya Carley, Jacob Shimkus Erickson, Emily Grubert, Antonia Hadjimichael, Jason Hill, Erin Mayfield, Destenie Nock, Kimberly Kivvaq Pikok, Rebecca K. Saari, Maria S. Lezcano, and Christopher W. Tessum. 2024. “Equity and Modeling in Sustainability Science: Examples and Opportunities throughout the Process.” *Proceedings of the National Academy of Sciences of the United States of America* 121 (13): e2215688121. https://doi.org/10.1073/pnas.2215688121.

56. Weyant, John. 2017. “Some Contributions of Integrated Assessment Models of Global Climate Change.” *Review of Environmental Economics and Policy* 11 (1): 115–137. https://doi.org/10.1093/reep/rew018

57. Porter, James J., and Suraje Dessai. 2017. “Mini-Me: Why Do Climate Scientists’ Misunderstand Users and Their Needs?” *Environmental Science & Policy* 77: 9–14. https://doi.org/10.1016/j.envsci.2017.07.004

58. Mendoza, Guillermo A., and Ravi Prabhu. 2005. “Combining Participatory Modeling and Multi-Criteria Analysis for Community-Based Forest Management.” *Forest Ecology and Management* 207 (1–2): 145–156. https://doi.org/10.1016/j.foreco.2004.10.024

59. Kiker, Gregory A., Todd S. Bridges, Arun Varghese, Thomas P. Seager, and Igor Linkov. 2005. “Application of Multicriteria Decision Analysis in Environmental Decision Making.” *Integrated Environmental Assessment and Management* 1 (2): 95–108. https://doi.org/10.1897/IEAM_2004a-015.1

60. Lazo, Alejandro. 2024. “California Isn’t on Track to Meet Its Climate Change Mandates—and a New Analysis Says It’s Not Even Close.” *CalMatters*, March 14, 2024. https://calmatters.org/environment/climate-change/2024/03/california-climate-change-mandate-analysis
61. California Energy Commission. n.d. “SB 100 Joint Agency Report.” Accessed April 13, 2026. https://www.energy.ca.gov/sb100
62. California Energy Commission. n.d. “Supply and Demand of Natural Gas in California.” Accessed April 13, 2026. https://www.energy.ca.gov/data-reports/energy-almanac/californias-natural-gas-market/supply-and-demand-natural-gas-california
63. U.S. Energy Information Administration. n.d. “California State Energy Profile.” Accessed April 13, 2026. https://www.eia.gov/states/CA/overview
64. California is now the 4th largest economy in the world | Governor of California. (2025, April 23). https://www.gov.ca.gov/. https://www.gov.ca.gov/2025/04/23/california-is-now-the-4th-largest-economy-in-the-world/
65. Declet-Barreto, Juan, and Andrew A. Rosenberg. 2022. “Environmental Justice and Power Plant Emissions in the Regional Greenhouse Gas Initiative States.” *PLOS ONE* 17 (7): e0271026. https://doi.org/10.1371/journal.pone.0271026
66. Petek, Gabriel. 2022. *Climate Change Impacts Across California: Crosscutting Issues*. Sacramento: Legislative Analyst’s Office. https://lao.ca.gov/reports/2022/4575/Climate-Change-Impacts-Crosscutting-Issues-040522.pdf
67. Crimmins, Allison R., Christopher W. Avery, David R. Easterling, Kenneth E. Kunkel, Brooke C. Stewart, and Thomas K. Maycock, eds. 2023. *Fifth National Climate Assessment*. Washington, DC: United States Global Change Research Program. https://doi.org/10.7930/NCA5.2023
68. Cart, Julie, Alejandro Lazo, Alastair Bland, and Alejandra Reyes-Velarde. 2024. “Here We Go Again: California Prepares to Battle Trump over Environmental Policies.” *CalMatters*, November 11, 2024. https://calmatters.org/environment/2024/11/california-trump-environmental-policies/
69. Hui, Iris, Angela Zhao, Bruce E. Cain, and Anne Maureen Driscoll. 2020. “Actual Experience, Partisan Distortion and Recall of Wildfire Events.” Stanford University, Bill

Lane Center for the American West and Center on Food Security and the Environment, September 2020. http://west.stanford.edu/sites/west/files/apsa2020.pdf

70. Bolstad, Parker, and David G. Victor. 2024. “Growing Deviations between Elite and Non-Elite Media Coverage of Climate Change in the United States.” *Climatic Change* 177: 102. https://doi.org/10.1007/s10584-024-03750-1
71. Global Energy Monitor. 2026. “Global Oil and Gas Plant Tracker.” Accessed April 13, 2026. https://globalenergymonitor.org/projects/global-oil-gas-plant-tracker/
72. U.S. Energy Information Administration. n.d. “Emissions by Plant and by Region.” Accessed April 13, 2026. https://www.eia.gov/electricity/data/emissions/
73. Jabareen, Yosef. “Building a Conceptual Framework: Philosophy, Definitions, and Procedure.” *International Journal of Qualitative Methods* 8, no. 4 (2009): 49–62. https://doi.org/10.1177/160940690900800406
74. Sen, Amartya. 1979. “Utilitarianism and Welfarism.” *The Journal of Philosophy* 76 (9): 463–489. https://doi.org/10.2307/2025934
75. Ferranna, Maddalena, and Marc Fleurbaey. 2020. “Prioritarianism and Climate Change.” Prepared for inclusion in *Prioritarianism in Practice*, edited by Matthew Adler and Ole Norheim. Cambridge: Cambridge University Press. https://ssrn.com/abstract=3813825
76. Piacquadio, P.G. (2017), A Fairness Justification of Utilitarianism. Econometrica, 85: 1261-1276. https://doi.org/10.3982/ECTA14151
77. Rojas-Zerpa, Juan C., and José M. Yusta. 2015. “Application of Multicriteria Decision Methods for Electric Supply Planning in Rural and Remote Areas.” *Renewable and Sustainable Energy Reviews* 52: 557–571. https://doi.org/10.1016/j.rser.2015.07.139
78. Taherdoost, Hamed, and Mitra Madanchian. 2023. “Multi-Criteria Decision Making (MCDM) Methods and Concepts.” *Encyclopedia* 3 (1): 77–87. https://doi.org/10.3390/encyclopedia3010006
79. Freudenberg, M. (2003), “Composite Indicators of Country Performance: A Critical Assessment”, *OECD Science, Technology and Industry Working Papers*, No. 2003/16, OECD Publishing, Paris, https://doi.org/10.1787/405566708255.
80. Pakkar, Mohammad Sadegh. 2016. "A Hierarchical Aggregation Approach for Indicators Based on Data Envelopment Analysis and Analytic Hierarchy Process" *Systems* 4, no. 1: 6. https://doi.org/10.3390/systems4010006

81. Cui, Ryna Yiyun, Nathan Hultman, Diyang Cui, Haewon McJeon, Sha Yu, Morgan R. Edwards, Arijit Sen, Kaihui Song, Christina Bowman, Leon Clarke, Junjie Kang, Jiehong Lou, Fuqiang Yang, Jiahai Yuan, Weirong Zhang, and Mengye Zhu. 2021. "A Plant-by-Plant Strategy for High-Ambition Coal Power Phaseout in China." *Nature Communications* 12 (1): 1468. https://doi.org/10.1038/s41467-021-21786-0
82. Hanson, Gordon H. 2023. *Local Labor Market Impacts of the Energy Transition: Prospects and Policies*. NBER Working Paper 30871. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w30871
83. Jolley, G. Jason, C. Khalaf, Greg Michaud, and A. M. Sandler. 2019. "The Economic, Fiscal, and Workforce Impacts of Coal-Fired Power Plant Closures in Appalachian Ohio." *Regional Science Policy & Practice* 11: 403–422. https://doi.org/10.1111/rsp3.12191
84. Funderburg, Richard. 2024. "The Disparate Local Impacts of Closing Fossil-Fuel Power Plants." *Illinois Municipal Policy Journal* 9 (1). https://www.uis.edu/sites/default/files/inline-images/04-Disparate%20Local%20Impacts%20of%20Closing%20Fossil%20Fuel%20Power%20Plants.pdf
85. U.S. Bureau of Labor Statistics. 2024. "May 2023 Metropolitan and Nonmetropolitan Area Occupational Employment and Wage Estimates: Transportation and Material Moving Occupations (53-0000)." Last modified April 3, 2024. https://www.bls.gov/oes/2023/may/oes_nat.htm
86. Wagman, David. 2017. "Automation Is Engineering the Jobs Out of Power Plants." *IEEE Spectrum*, August 3, 2017. https://spectrum.ieee.org/automation-is-engineering-the-jobs-out-of-power-plants
87. U.S. Bureau of Economic Analysis. n.d. "What Are the Guidelines for Citing BEA Information?" https://www.bea.gov/help/faq/478
88. Zheng Tian & Paul D. Gottlieb & Stephan J. Goetz, 2020. "**Measuring industry co-location across county borders**," Spatial Economic Analysis, Taylor & Francis Journals, vol. 15(1), pages 92-113, January. https://ideas.repec.org/a/taf/specan/v15y2020i1p92-113.html
89. U.S. Bureau of Labor Statistics. 2024. "May 2023 Metropolitan and Nonmetropolitan Area Occupational Employment and Wage Estimates: Santa Maria–Santa Barbara, CA

(42200), Transportation and Material Moving Occupations (53-0000).” Last modified April 3, 2024. https://www.bls.gov/oes/2023/may/oes_42200.htm#53-0000

90. U.S. Census Bureau. “Per Capita Income.” *American Community Survey 5-Year Estimates Detailed Tables*, Table B19301, 2022. https://data.census.gov/table/ACSDT5Y2022.B19301?q=per+capita+Income&g=040XX00US06&y=2022 (accessed April 14, 2026).
91. Sergi, Briana J., Peter J. Adams, Nicholas Z. Muller, Allen L. Robinson, Steven J. Davis, Julian D. Marshall, and Inês L. Azevedo. 2020. “Optimizing Emissions Reductions from the U.S. Power Sector for Climate and Health Benefits.” *Environmental Science & Technology* 54 (12): 7513–7523. https://doi.org/10.1021/acs.est.9b06936
92. Campos Morales, Carolina, Erin L. Pakhtigian, Jacob R. Landry, Hannah Wiseman, Anh T. Pham, and Wei Peng. 2024. “Designing Retirement Strategies for Coal-Fired Power Plants To Mitigate Air Pollution and Health Impacts.” *Environmental Science & Technology* 58 (35): 15371–15380. https://doi.org/10.1021/acs.est.4c00704
93. Burney, Jennifer A. 2020. “The Downstream Air Pollution Impacts of the Transition from Coal to Natural Gas in the United States.” *Nature Sustainability* 3 (2): 152–160. https://doi.org/10.1038/s41893-019-0453-5
94. U.S. Environmental Protection Agency. 2023. *2021 and 2022 Annual Power Plant Emissions*. EPA Clean Air Markets Division. Data current as of February 7, 2023. Dataset. https://www.epa.gov/newsreleases/epa-releases-2022-power-plant-emissions-data
95. Decarb My State. n.d. “AltaGas Ripon Energy Inc.” https://decarbmystate.com/california/power-plant/altagas-ripon-energy-inc/
96. Pollin, Robert, and Brian Callaci. 2019. “The Economics of Just Transition: A Framework for Supporting Fossil Fuel–Dependent Workers and Communities in the United States.” *Labor Studies Journal* 44 (2): 93–138. https://doi.org/10.1177/0160449X18787051
97. Fofrich, Robert, Dan Tong, Katherine Calvin, Harmen Sytze De Boer, Johannes Emmerling, Oliver Fricko, Shinichiro Fujimori, Gunnar Luderer, Joeri Rogelj, and Steven J. Davis. 2020. “Early Retirement of Power Plants in Climate Mitigation Scenarios.” *Environmental Research Letters* 15 (9): 094064. https://doi.org/10.1088/1748-9326/ab96d3

98. Burton, Christopher G. 2015. "A Validation of Metrics for Community Resilience to Natural Hazards and Disasters Using the Recovery from Hurricane Katrina as a Case Study." *Annals of the Association of American Geographers* 105 (1): 67–86. https://doi.org/10.1080/00045608.2014.960039.

99. Carley, Sanya, Tom P. Evans, and David M. Konisky. 2018. "Adaptation, Culture, and the Energy Transition in American Coal Country." *Energy Research & Social Science* 37: 133–139. https://doi.org/10.1016/j.erss.2017.10.007

100. Gu, Donghwan, Maria Dillard, Michael Gerst, and Jarrod Loerzel. 2023. "Validating Commonly Used Indicators for Community Resilience Measurement." *Natural Hazards Review* 24 (2): 04023008. https://doi.org/10.1061/NHREFO.NHENG-1642

101. U.S. Census Bureau. 2026. "2022 NAICS Sector 62—Health Care and Social Assistance." Last modified February 27, 2026. https://www.census.gov/data/tables/2022/econ/economic-census/naics-sector-62.html

102. U.S. Census Bureau. 2022. "Business Dynamics Statistics: Establishments by Age (NAICS 62)." https://data.census.gov/table/BDSTIMESERIES.BDSEAGE?g=050XX00US06071&codeset=naics~62

103. Edgemon, Lesley, Carol Freeman, Carmella Burdi, John Hutchison, Karen Marsh, and Kyle Pfeiffer. 2020. *Community Resilience Indicator Analysis: County-Level Analysis of Commonly Used Indicators from Peer-Reviewed Research, 2020 Update*. Argonne National Laboratory and Federal Emergency Management Agency. https://publications.anl.gov/anlpubs/2023/04/181636.pdf

104. Isserman, Andrew M., and James Westervelt. 2006. "1.5 Million Missing Numbers: Overcoming Employment Suppression in County Business Patterns Data." *International Regional Science Review* 29 (3): 311–335. https://doi.org/10.1177/0160017606290359

105. U.S. Census Bureau. 2022. "Business Dynamics Statistics: Establishment Age (BDSEAGE), NAICS 81." https://data.census.gov/table/BDSTIMESERIES.BDSEAGE?g=050XX00US06071&codeset=naics~81 (accessed April 14, 2026)

106. Parker, Kendall M., Jay Barlow, Jason Eisdorfer, and Kamila Kazimierczuk. 2023. "Observations of an Evolving Grid: Resilience and Equity Performance Metrics." *Current*

Sustainable/Renewable Energy Reports 10: 234–242. https://doi.org/10.1007/s40518-023-00227-0

107. Drehobl, Ariel, Lauren Ross, and Roxana Ayala. 2020. *How High Are Household Energy Burdens? An Assessment of National and Metropolitan Energy Burden across the United States*. Washington, DC: American Council for an Energy-Efficient Economy. https://www.aceee.org/research-report/u2006.
108. Bohr, Jeremiah, and Anna C. McCreery. 2020. "Do Energy Burdens Contribute to Economic Poverty in the United States? A Panel Analysis." *Social Forces* 99 (1): 155–177. https://doi.org/10.1093/sf/soz131
109. U.S. Department of Energy, Office of State and Community Energy Programs. n.d. "Low-Income Energy Affordability Data (LEAD) Tool." https://www.energy.gov/scep/slsc/lead-tool
110. U.S. Department of Energy, Office of State and Community Energy Programs. 2024. *Low-Income Energy Affordability Data (LEAD) Tool Methodology*. December 2024. https://www.energy.gov/sites/default/files/2024-12/lead-methodology_122024.pdf
111. International Energy Agency. 2024. *Strategies for Affordable and Fair Clean Energy Transitions: World Energy Outlook Special Report*. Paris: International Energy Agency. https://www.iea.org/reports/strategies-for-affordable-and-fair-clean-energy-transitions
112. U.S. Department of Energy. 2024. *The Future of Resource Adequacy: Solutions for Clean, Reliable, Secure, and Affordable Electricity*. Washington, DC: U.S. Department of Energy
113. U.S. Department of Energy. 2024. *The Future of Resource Adequacy: Solutions for Clean, Reliable, Secure, and Affordable Electricity*. Washington, DC: U.S. Department of Energy.
114. Schleifer, Anna, Abigail King, Juliet Homer, and Todd Levin. 2023. *Assessing the Resource Adequacy and Community Impacts of an Evolving Grid: Considerations for State Regulators*. ANL-23/35. Lemont, IL: Argonne National Laboratory and Pacific Northwest National Laboratory.
115. Burke, Paul J., Rohan Best, and Frank Jotzo. 2019. "Closures of Coal-Fired Power Stations in Australia: Local Unemployment Effects." *Australian Journal of Agricultural and Resource Economics* 63 (1): 142–165. https://doi.org/10.1111/1467-8489.12289

116. U.S. Energy Information Administration. 2024**.** "What Is the Efficiency of Different Types of Power Plants?" *Frequently Asked Questions (FAQs)*. Last updated May 15, 2024.
117. California Energy Commission**.** n.d. "QFER Power Plant Heat Rates." *Energy Almanac*.
118. Nyberg, Michael**.** 2020. *Thermal Efficiency of Natural Gas-Fired Generation in California: 2019 Update*. California Energy Commission. Publication Number CEC-200-2020-003
119. California Office of Environmental Health Hazard Assessment. n.d. "SB 535 Disadvantaged Communities." *CalEnviroScreen*. Accessed April 14, 2026. https://oehha.ca.gov/calenviroscreen/sb535
120. Tarekegne, Bethel, Devyn Powell, Konstantinos Oikonomou, Eva Jacroux, and Rebecca O'Neil. 2022. "Analysis of Energy Justice and Equity Impacts from Replacing Peaker Plants with Energy Storage." Paper presented at the **2022 IEEE Electrical Energy Storage Applications and Technologies Conference (EESAT)**, November 8, 2022. https://doi.org/10.1109/EESAT55007.2022.9998034
121. Specht, Mark G., and Laura M. Wisland. 2018. *Turning Down the Gas in California: The Role of Natural Gas in the State's Clean Electricity Future*. Union of Concerned Scientists. https://www.jstor.org/stable/pdf/resrep24067.1.pdf
122. Goforth, Teagan and Levin, Todd and Nock, Destenie, Incorporating Energy Justice and Equity into Power System Models: A Review of Current Practices and Paths Forward (April 30, 2023). Available at SSRN: https://ssrn.com/abstract=4591242 or http://dx.doi.org/10.2139/ssrn.4591242
123. Trinidad, Stephen, Cole Brokamp, Andres Mor Huertas, Andrew F. Beck, Carley L. Riley, Erika Rasnick, Richard Falcone, and Meera Kotagal. 2022. "Use Of Area-Based Socioeconomic Deprivation Indices: A Scoping Review And Qualitative Analysis." *Health Affairs* 41 (12): 1804–1811. https://doi.org/10.1377/hlthaff.2022.00482
124. Martenies, Sheena E., Arash Akherati, Shantanu Jathar, and Sheryl Magzamen. 2019. "Health and Environmental Justice Implications of Retiring Two Coal-Fired Power Plants in the Southern Front Range Region of Colorado." *GeoHealth* 3 (9): 266–283. https://doi.org/10.1029/2019GH000206
125. Global Energy Monitor. 2024. *Global Oil and Gas Plant Tracker*. January 2024 release. https://globalenergymonitor.org/projects/global-oil-gas-plant-tracker

126. Global Energy Monitor. n.d. *Global Energy Monitor*. https://globalenergymonitor.org
127. Dagdelen, John, Alex Dunn, Sanghoon Lee, Nicholas Walker, Andrew S. Rosen, Gerbrand Ceder, Kristin A. Persson, and Anubhav Jain. 2024. "Structured Information Extraction from Scientific Text with Large Language Models." *Nature Communications* 15: 1418. https://doi.org/10.1038/s41467-024-45563-x.
128. Shamsabadi, Mahsa, Jennifer D'Souza, and Sören Auer. 2024. "Large Language Models for Scientific Information Extraction: An Empirical Study for Virology." In *Findings of the Association for Computational Linguistics: EACL 2024*, 374–92. St. Julian's, Malta: Association for Computational Linguistics.
129. Polak, Maciej P., Shrey Modi, Anna Latosinska, Jinming Zhang, Ching-Wen Wang, Shaonan Wang, Ayan Deep Hazra, and Dane Morgan. 2024. "Flexible, Model-Agnostic Method for Materials Data Extraction from Text Using General Purpose Language Models." *Digital Discovery* 3: 1221–35. https://doi.org/10.1039/D4DD00016A
130. Sentence Transformers. 2026. "all-MiniLM-L6-v2." Hugging Face model card. Accessed March 15, 2026. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2
131. Douze, Matthijs, Alexandre Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2024. "The Faiss Library." *arXiv preprint* arXiv:2401.08281. https://doi.org/10.48550/arXiv.2401.08281
132. Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*
133. Gemma Team, Morgane Rivière, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Leme, Jemmy Leme Anton Tsitsulin, Nino Vieillard, Piotr Stanczyk, Sertan Girgin, Nikola Momchev, Matt Hoffman, Shantanu Thakoor, Jean-Bastien Grill, Behnam Neyshabour, Olivier Bachem, et al. 2024. *"Gemma 2: Improving Open Language Models at a Practical Size."* arXiv preprint arXiv:2408.00118

134. Yang, An, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. 2024. *"Qwen2.5 Technical Report."* arXiv preprint arXiv:2412.15115.

135. Jiang, Albert Q., Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. *"Mistral 7B."* arXiv preprint arXiv:2310.06825.

136. Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2019. "HuggingFace's Transformers: State-of-the-art Natural Language Processing." arXiv (Cornell University), October. https://doi.org/10.48550/arxiv.1910.03771.

137. Leo Breiman, "Random Forests," *Machine Learning* 45, no. 1 (2001): 5–32, https://doi.org/10.1023/A:1010933404324

138. Scornet, Erwan. 2020. "Trees, Forests, and Impurity-Based Variable Importance." *arXiv* preprint arXiv:2001.04295. https://doi.org/10.48550/arXiv.2001.04295

139. Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. "Conditional Variable Importance for Random Forests." *BMC Bioinformatics* 9: 307. https://doi.org/10.1186/1471-2105-9-307

140. Anwoy Chatterjee, H. S. V. N. S. Kowndinya Renduchintala, Sumit Bhatia, and Tanmoy Chakraborty, "POSIX: A Prompt Sensitivity Index for Large Language Models," in *Findings of the Association for Computational Linguistics: EMNLP 2024* (Miami, FL: Association for Computational Linguistics, 2024), 14550–14565, https://aclanthology.org/2024.findings-emnlp.852

141. Qinggang Zhang, Zhishang Xiang, Yilin Xiao, Le Wang, Junhui Li, Xinrun Wang, and Jinsong Su, “FaithfulRAG: Fact-Level Conflict Modeling for Context-Faithful Retrieval-Augmented Generation,” in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vienna, Austria: Association for Computational Linguistics, 2025), 21863–21882, https://aclanthology.org/2025.acl-long.1062/
142. Lovisa Hagström, Sara Vera Marjanović, Haeun Yu, Arnav Arora, Christina Lioma, Maria Maistro, Pepa Atanasova, and Isabelle Augenstein, “A Reality Check on Context Utilisation for Retrieval-Augmented Generation,” *arXiv*, December 22, 2024, https://doi.org/10.48550/arXiv.2412.17031
143. Zhou, Jeffrey, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. 2023. "Instruction-Following Evaluation for Large Language Models." arXiv preprint arXiv:2311.07911
144. Antoniadis, Anestis, Sophie Lambert-Lacroix, and Jean-Michel Poggi. 2021. “Random Forests for Global Sensitivity Analysis: A Selective Review.” *Reliability Engineering & System Safety* 206: 107312. https://doi.org/10.1016/j.ress.2020.107312
145. Schonlau, Matthias, and Rosie Yuyan Zou. 2020. “The Random Forest Algorithm for Statistical Learning.” *The Stata Journal* 20 (1): 3–29. https://doi.org/10.1177/1536867X20909688
146. Payabvash, Saeed, Maryam Aboian, Tarik Tihan, and Soonmee Cha. 2020. “Machine Learning Decision Tree Models for Differentiation of Posterior Fossa Tumors Using Diffusion Histogram Analysis and Structural MRI Findings.” *Frontiers in Oncology* 10: 71. https://doi.org/10.3389/fonc.2020.00071
147. Probst, Philipp, and Anne-Laure Boulesteix. 2017. “To Tune or Not to Tune the Number of Trees in Random Forest?” *arXiv preprint* arXiv:1705.05654. https://doi.org/10.48550/arXiv.1705.05654
148. Zafari, Azar, Raul Zurita-Milla, and Emma Izquierdo-Verdiguier. 2019. "Evaluating the Performance of a Random Forest Kernel for Land Cover Classification" *Remote Sensing* 11, no. 5: 575. https://doi.org/10.3390/rs11050575
149. California Public Utilities Commission. 2024a. *24-25 TPP Busbar Mapping: Gas Capacity Not Retained Criteria Application and Selection*. Energy Division Staff.

February 15, 2024. https://www.cpuc.ca.gov/-/media/cpuc-website/divisions/energy-division/documents/integrated-resource-plan-and-long-term-procurement-plan-irp-ltpp/2023-irp-cycle-events-and-materials/assumptions-for-the-2024-2025-tpp/gasnotretained_overview.pdf

150. California Public Utilities Commission. 2024b. *Gas Capacity Not Retained Assumption List for the Base Case and Sensitivity Portfolios*. Microsoft Excel workbook. Posted February 15, 2024. https://www.cpuc.ca.gov/-/media/cpuc-website/divisions/energy-division/documents/integrated-resource-plan-and-long-term-procurement-plan-irp-ltpp/2023-irp-cycle-events-and-materials/assumptions-for-the-2024-2025-tpp/gasnotretained_mappingresults.xlsx
151. Lahdelma, R., Salminen, P. (2010). Stochastic Multicriteria Acceptability Analysis (SMAA). In: Ehrgott, M., Figueira, J., Greco, S. (eds) Trends in Multiple Criteria Decision Analysis. International Series in Operations Research & Management Science, vol 142. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-5904-1_10
152. Welcome to PyMuPDF. (2026, May 4). https://pymupdf.readthedocs.io/en/latest/. Retrieved May 19, 2026, from https://pymupdf.readthedocs.io/en/latest/
153. Park, A. & Urban Sustainability Directors Network. (2014). Equity in Sustainability : An equity scan of local government sustainability programs. In Urban Sustainability Directors Network [Report]. Urban Sustainability Directors Network (USDN). Retrieved May 19, 2026,from https://www.usdn.org/uploads/cms/documents/usdn_equity_scan_sept_2014_final.pdf
154. Shetty, K., Nunez, Y., Heath, N., Efstathiou, C., Bilsback, K., Union of Concerned Scientists, Healthy Climate Wisconsin, & PSE Healthy Energy. (2025). Health, equity, and economic impacts of proposed gas power plants in Wisconsin: Oak Creek and Paris projects. https://blog.ucs.org/wp-content/uploads/2025/01/Health-Equity-and-Economic-Impacts-of-Proposed-Gas-Power-Plants-in-Wisconsin-Oak-Creek-and-Paris-Projects.pdf
155. Roy, B. (1968). Classement et choix en présence de points de vue multiples. Revue Française D Informatique Et De Recherche Opérationnelle, 2(8), 57–75. https://doi.org/10.1051/ro/196802v100571

156. Pohekar, S., & Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning—A review. Renewable and Sustainable Energy Reviews, 8(4), 365–381. https://doi.org/10.1016/j.rser.2003.12.007
157. Opricovic, S., & Tzeng, G. (2003). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research, 156(2), 445–455. https://doi.org/10.1016/s0377-2217(03)00020-1
158. L Saaty, T. (1980). The analytic hierarchy process : planning, priority setting, resource allocation [Internet Archive]. New York ; London : McGraw-Hill International Book Co. https://archive.org/details/analytichierarch0000saat/page/n9/mode/2up
159. Kandakoglu, A., Frini, A., & Amor, S. B. (2019). Multicriteria decision making for sustainable development: A systematic review. Journal of Multi-Criteria Decision Analysis, 26(5–6), 202–251. https://doi.org/10.1002/mcda.1682
160. Esangbedo, M. O., & Che, A. (2016). Grey Weighted sum model for evaluating business environment in West Africa. Mathematical Problems in Engineering, 2016, 1–14. https://doi.org/10.1155/2016/3824350
161. Hwang, C., & Yoon, K. (1981). Multiple attribute decision making. In Lecture notes in economics and mathematical systems. https://doi.org/10.1007/978-3-642-48318-9
162. Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2023. “Retrieval-Augmented Generation for Large Language Models: A Survey.” arXiv.Org. December 18, 2023. https://arxiv.org/abs/2312.10997.
163. Huang, Lei, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, et al. 2024a. “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions.” ACM Transactions on Information Systems 43 (2): 1–55. https://doi.org/10.1145/3703155.
164. Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, et al. 2020. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” arXiv.Org. May 22, 2020. https://arxiv.org/abs/2005.11401.

165. Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient Estimation of Word Representations in Vector Space.” arXiv.Org. January 16, 2013. https://arxiv.org/abs/1301.3781.
166. Reimers, Nils, and Iryna Gurevych. 2019. “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks.” arXiv.Org. August 27, 2019. https://arxiv.org/abs/1908.10084.
167. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” 2017. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.
168. Mahmud, Tarek, Bin Duan, Corina Pasareanu, and Guowei Yang. 2025. “Enhancing LLM Code Generation With Ensembles: A Similarity-Based Selection Approach.” arXiv.Org. March 20, 2025. https://arxiv.org/abs/2503.15838.
169. Salton, Gerard, and Michael J McGill. 1983. Introduction to Modern Information Retrieval. McGraw-Hill, New York, ©1983.
170. Smith, Brandon, Mohamed Reda Bouadjenek, Tahsin Alamgir Kheya, Phillip Dawson, and Sunil Aryal. 2025. “A Comprehensive Analysis of Large Language Model Outputs: Similarity, Diversity, and Bias.” arXiv.Org. May 14, 2025. https://arxiv.org/abs/2505.09056.
171. Xu, Shaochen, Zihao Wu, Huaqin Zhao, Peng Shu, Zhengliang Liu, Wenxiong Liao, Sheng Li, Andrea Sikora, Tianming Liu, and Xiang Li. 2024. “Reasoning Before Comparison: LLM-Enhanced Semantic Similarity Metrics for Domain Specialized Text Analysis.” arXiv.Org. February 17, 2024. https://arxiv.org/abs/2402.11398.
172. Lempert, Robert, Steven Popper, and Steven Bankes. 2002. “Confronting Surprise.” *Social Science Computer Review* 20 (4): 420–40. https://doi.org/10.1177/089443902237320.
173. Beutler, Philipp, Tove A. Larsen, Max Maurer, Philipp Staufer, and Judit Lienert. 2024. “A Participatory Multi-criteria Decision Analysis Framework Reveals Transition Potential Towards Non-grid Wastewater Management.” Journal of Environmental Management 367 (August): 121962. https://doi.org/10.1016/j.jenvman.2024.121962.

174. Lempert, Robert J. 2025. “Decision Making Under Deep Uncertainty and the Great Acceleration: The Role of Experts and Policy Analysis in a World in Transition.” RAND. April 24, 2025. https://www.rand.org/pubs/perspectives/PEA3789-1.html.
175. Fleiss, Joseph L. 1971. “Measuring Nominal Scale Agreement Among Many Raters.” Psychological Bulletin 76 (5): 378–82. https://doi.org/10.1037/h0031619.
176. Koo, Terry K., and Mae Y. Li. 2016. “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.” Journal of Chiropractic Medicine 15 (2): 155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
177. Shrout, Patrick E., and Joseph L. Fleiss. 1979. “Intraclass Correlations: Uses in Assessing Rater Reliability.” Psychological Bulletin 86 (2): 420–28. https://doi.org/10.1037/0033-2909.86.2.420.
178. Tricco, Andrea C., Erin Lillie, Wasifa Zarin, Kelly K. O’Brien, Heather Colquhoun, Danielle Levac, David Moher, et al. 2018. “PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation.” Annals of Internal Medicine 169 (7): 467–73. https://doi.org/10.7326/m18-0850.
179. Broekhuizen, Henk, Catharina G. M. Groothuis-Oudshoorn, Janine A. van Til, J. Marjan Hummel, and Maarten J. IJzerman. 2015. "A Review and Classification of Approaches for Dealing with Uncertainty in Multi-Criteria Decision Analysis for Healthcare Decisions." *Pharmacoeconomics* 33 (5): 445–455. https://doi.org/10.1007/s40273-014-0251-x
180. Tebaldi, Claudia, and Reto Knutti. 2007. "The Use of the Multi-Model Ensemble in Probabilistic Climate Projections." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (1857): 2053–2075. https://doi.org/10.1098/rsta.2007.2076
181. Parker, Wendy S. 2013. "Ensemble Modeling, Uncertainty and Robust Predictions." *Wiley Interdisciplinary Reviews: Climate Change* 4 (3): 213–223. https://doi.org/10.1002/wcc.220
182. Morgan, M. Granger. 2014. "Use (and Abuse) of Expert Elicitation in Support of Decision Making for Public Policy." *Proceedings of the National Academy of Sciences* 111 (20): 7176–7184. https://doi.org/10.1073/pnas.1319946111
183. Hemming, Victoria, Mark A. Burgman, Anca M. Hanea, Marissa F. McBride, and Bonnie C. Wintle. 2018. "A Practical Guide to Structured Expert Elicitation Using the IDEA

Protocol." *Methods in Ecology and Evolution* 9 (1): 169–180. https://doi.org/10.1111/2041-210X.12857

184. Briggs, Andrew H., Milton C. Weinstein, Elisabeth A.L. Fenwick, Jonathan Karnon, Mark J. Sculpher, and A. David Paltiel. 2012. "Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6." Value in Health 15 (6): 835–42. https://doi.org/10.1016/j.jval.2012.04.014
185. Knutti, Reto, Reinhard Furrer, Claudia Tebaldi, Jan Cermak, and Gerald A. Meehl. 2009. "Challenges in Combining Projections From Multiple Climate Models." Journal of Climate 23 (10): 2739–58. https://doi.org/10.1175/2009jcli3361.1
186. O'Hagan, Anthony. 2019. "Expert Knowledge Elicitation: Subjective but Scientific." The American Statistician 73 (sup1): 69–81. https://doi.org/10.1080/00031305.2018.1518265
187. Adler, Matthew D. 2024. "Social Welfare Functions and Health Policy: A New Approach." Journal of Benefit-Cost Analysis 15 (S1): 26–51. https://doi.org/10.1017/bca.2024.6
188. Dyer, James S., and Rakesh K. Sarin. 1979. "Measurable Multiattribute Value Functions." Operations Research 27 (4): 810–22. https://doi.org/10.1287/opre.27.4.810
189. Frankfurt, Harry. 1987. "Equality as a Moral Ideal." *Ethics* 98 (1): 21–43. https://www.jstor.org/stable/2381290?seq=1.
190. Crisp, Roger. 2003. "Equality, Priority, and Compassion." Ethics 113 (4): 745–63. https://doi.org/10.1086/373954.
191. Cota-Robles, Marc. 2019. "Garcetti Announces Decision to Forgo Plans to Rebuild Coastal Natural-gas Plants." ABC7 Los Angeles, February 13, 2019. https://abc7.com/post/garcetti-scraps-plans-to-rebuild-3-coastal-natural-gas-plants/5134409.
192. Becker, Rachel. 2023. "Southern California's Natural Gas Plants to Stay Open Through 2026." Https://Calmatters.Org/. August 15, 2023. Accessed June 5, 2026. https://calmatters.org/environment/2023/08/southern-california-natural-gas-plants-remain-open/