Automated Diet Matrix Construction for Marine Ecosystem Models Using Generative AI

Scott Spillias[1,2*] Elizabeth A. Fulton[1,2] Fabio Boschetti[2,3,4]
Cathy Bulman[1] Joanna Strzelecki[3] Rowan Trebilco[1,2]

May 19, 2025

[1]CSIRO Environment, Hobart, Australia
[2]Centre for Marine Socio-Ecology, University of Tasmania, Hobart, Australia
[3]CSIRO Environment, IOMRC Crawley, Australia
[4]Oceans Institute, The University of Western Australia, Crawley, WA

Abstract

This study introduces a novel Large Language Model (LLM) driven framework for automated species grouping and diet matrix generation in Ecopath with Ecosim (EwE) ecosystem models, addressing a critical bottleneck in model development. The framework (i) retrieves a marine species list from an area; (ii) uses LLMs to classify them into functional groups; and (iii) synthesises trophic interactions from diverse data sources including global biodiversity databases, species interaction repositories, and unstructured user-provided text. We evaluate the framework across four large Australian marine regions to assess both consistency and ecological accuracy of the resulting functional groups and diet proportions. The framework demonstrates high reproducibility in species grouping decisions (>99.7% consistency) and diet matrix construction, with 51-59% of predator-prey interactions showing consistent diet proportions across multiple runs. Validation against expert-derived matrices for the Great Australian Bight ecosystem reveals strong ecological alignment and accuracy, with 92.6% of taxonomic assignments being at least partially correct (>75% fully correct), and correctly identifying 85% of trophic interactions, while estimating diet proportions within 0.2 of expert values for 80% of interactions. These findings demonstrate the framework's potential to generate reproducible, ecologically meaningful components for ecosystem model development while significantly reducing development time.

1 Introduction

Ecosystem modelling is a critical tool for understanding and managing complex environments, with Ecopath with Ecosim (EwE) being a well-established framework with thousands of implementations used to model marine ecosystems and predict their responses to external pressures (Christensen and Walters, 2004; Colléter et al., 2015). EwE

*Corresponding author: scott.spillias@csiro.au

models provide quantitative insights into ecosystem structure and function, enabling researchers to assess cumulative impacts of multiple stressors and support ecosystem-based fisheries management (EBFM) decisions (Coll et al., 2015; Villasante et al., 2016; Geary et al., 2020). However, constructing these models presents significant challenges, particularly in constructing and parameterizing diet matrices that capture the complex web of trophic interactions within an ecosystem.

Traditional approaches to EwE model development rely heavily on extensive literature review, data collation and expert knowledge, which are time-consuming and resource-intensive (Holden et al., 2024a). The process of assembling diet matrices is particularly challenging, requiring synthesis of diverse data sources including field studies, literature reviews, and expert opinion. This creates a significant bottleneck in model development, especially when applying models to new geographical contexts (Holden et al., 2024b). Recent advances in artificial intelligence (AI) offer new opportunities to streamline the model development process and avoid such bottlenecks (Spillias et al., 2024b). AI tools have demonstrated success in both knowledge/evidence synthesis tasks (Spillias et al., 2024c; Keck et al., 2025; Castro et al., 2024; Spillias et al., 2024a; Zheng et al., 2023; Nugraha et al., 2024), ecological and environmental tasks (Fernandes and D'Mello, 2024; Li et al., 2024; Chen and Xu, 2024; Dorm et al., 2025; Noleto Filho et al., 2024) and modelling tasks (Lapeyrolerie et al., 2022; Tuia et al., 2022; Karniadakis et al., 2021), but their application to process-based ecosystem modelling remains nascent. The key challenge lies in ensuring that AI-driven approaches can effectively synthesise available information while maintaining ecological validity.

We present a novel and flexible framework for assembling and synthesizing user-defined and online resources to parameterise EwE diet matrices using Large Language Models (LLMs). Our approach integrates multiple data sources, including global biodiversity databases, species interaction repositories, and locally-held unstructured or structured text, to automate key steps in model development. The framework employs user-selected LLMs to group species into functional units and estimate trophic interaction strengths. We evaluate the system in four distinct Australian marine ecosystems - the Northern Australia, South East shelf, and South East offshore regions, where we assess the reproducibility of the approach, and in the Great Australian Bight where we assess the accuracy of the approach. Specifically, we test the precision (repeatability) and scientific accuracy of automated species grouping decisions, and the precision and accuracy of the resulting diet matrix proportions, with accuracy defined in terms of similarity to expert estimates. These regions offer contrasting environmental conditions, species assemblages, and ecological dynamics, providing a robust test of the framework's adaptability and reliability.

2 Methods

2.1 AI-Assisted Framework Overview

The development of ecosystem models requires substantial time organizing species into functional groups and determining their interactions. This framework automates these tasks through a five-stage process that integrates artificial intelligence with ecological databases (Figure 1).

The first step in the process is to define a model domain and the resultant shapefile is used to derive a comprehensive species list from Ocean Biodiversity Information

System (OBIS) (Grassle and Stocks, 1999). In Stage 2, this species list is enriched with ecological data from FishBase and SeaLifeBase (Froese et al., 2010), which provide life history traits, ecological parameters, and diet information, and with trophic interactions from the Global Biotic Interactions (GLOBI) database (Poelen et al., 2014).

Stage 3 employs a LLM for species grouping. We use Claude Sonnet-3.5 (hereafter referred to as 'Claude')(Anthropic, 2024), though other LLMs can be incorporated. The grouping process considers both the research focus specified by the user and a generic grouping template, assigning species into functional groups.

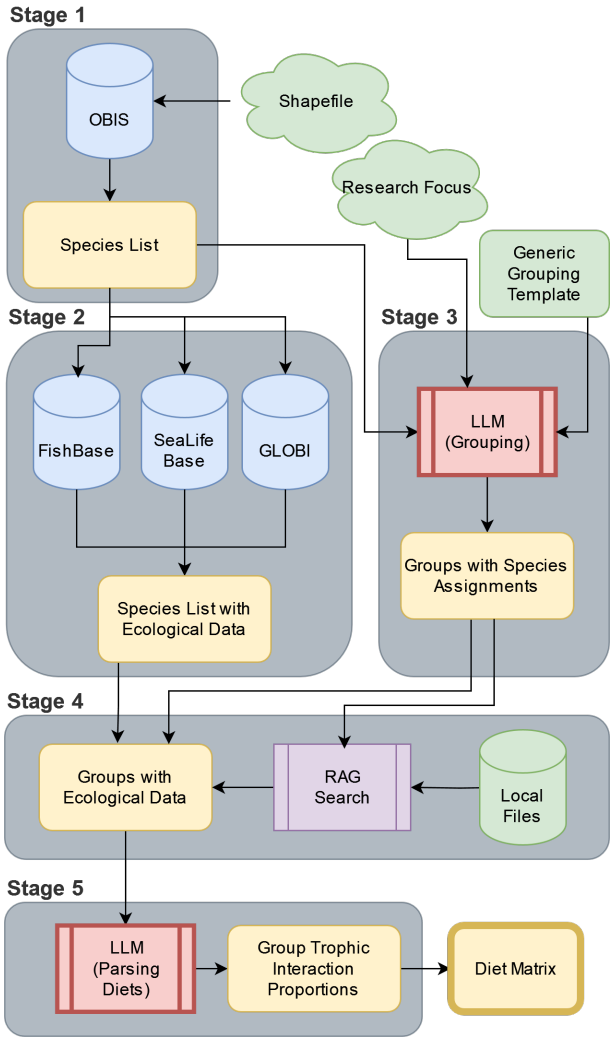

Figure 1: Overview of the AI-assisted framework for ecosystem model development. The process consists of five stages: species identification, ecological data collection, functional group organization, diet data synthesis, and diet matrix construction. Each stage integrates multiple data sources and analytical approaches, with user-provided inputs (shown in green) guiding key decisions.

Stage 4 uses a retrieval augmented generation (RAG) system (e.g. (Keck et al., 2025)) to synthesize the species-level ecological data into group-level summaries and incorporates user-provided local knowledge from a vector storage database (a database that stores and retrieves information based on semantic similarity rather than exact text string matching), Chroma (Chroma, 2024), to produce group-level diet composition estimates. Finally, Stage 5 uses the LLM to parse and structure this combined data into a

diet matrix, determining trophic interaction proportions among functional groups.

Section S4 of the supplementary material contains detailed documentation of all processing steps, including database queries, literature search criteria, and ecological classification rules. The complete codebase and configuration files reside at https://github.com/s-spillias/AI-EwE-Diets.

2.1.1 Species Identification

The framework begins by taking a user-defined shape file that defines the study region boundaries. It then accesses the Ocean Biodiversity Information System (Grassle and Stocks, 1999) through the `robis` R package (Chamberlain, 2020), which enables automated querying and data processing. We chose OBIS as our primary data source due to its extensive marine species coverage and standardized taxonomic classifications.

The framework uses the 'checklist' function from `robis` to retrieve scientific names and complete taxonomic classifications from kingdom to species level for all recorded species within these boundaries. Whilst collecting all occurrences would help with estimating distributions and biomasses for other modelling purposes, the time and computational resources required to process such large datasets are prohibitive and so we focus on presence only.

To limit the amount of processing required by the LLM, the raw OBIS data is transformed in two steps. First, the R script filters the dataset using OBIS's `is_marine` flag to eliminate terrestrial species that may occur in coastal records. Second, it removes taxonomic redundancy using a rank-based approach that retains only the most specific classification level available. For example, if our dataset contains both *Pagrus auratus* (species-level) and *Pagrus* (genus-level) entries for the same organism, our algorithm retains only the species-level entry. Our approach processes taxonomic ranks from most specific (scientific name) to most general (kingdom), keeping only the entry with the highest taxonomic resolution for each organism.

The final species list is stored in a structured CSV file containing verified marine species and their complete taxonomic hierarchies. The complete R implementation, including rank-based filtering algorithms and geographic processing functions, is available in the project repository.

2.1.2 Data Harvesting

Following species identification, the framework gathers ecological and life history information for each identified species. From SeaLifeBase and FishBase (Froese et al., 2010), it extracts a range of information, including habitat preferences (marine, brackish, or freshwater), depth range distributions, maximum body lengths, and diet data. These databases are accessed through their publicly available PARQUET files using DuckDB for efficient querying of large datasets.

Our species filtering protocol implements specific constraints to align with Ecopath with Ecosim modelling requirements. When processing taxonomic data, we prioritize species-level entries over genus-level classifications, only defaulting to genus-level when species-specific data is unavailable. For diet composition data, we extract data on food items including prey species codes ('SpecCode'), food groups, and prey stages (as is found in FishBase and SeaLifeBase; see [Github repository for specific queries to these databases]). We specifically filter for juvenile and/or adult life stages, as larval

stages are typically incorporated into planktonic functional groups rather than treated as separate components of adult diets.

We supplement the base biological data with interaction information from the Global Biotic Interactions (GLOBI) database (Poelen et al., 2014). For each species, we query the GLOBI API using URL-encoded species names to retrieve interaction records in CSV format. The GLOBI data processing preserves the raw interaction data and treats directional relationships ('eats'/'preysOn' and 'eatenBy'/'preyedUponBy') as complementary evidence of trophic interactions. For each predator-prey group pair, we tally the total number of observed interactions, which provides information about the relative frequency of feeding relationships between groups. We further enrich this data through retrieval-augmented generation (RAG) searches of regional literature (detailed in Section S4.2 of the supplementary material), focusing on specific feeding relationships and dietary preferences.

Technical implementation details are provided in Section S1 of the supplementary material.

2.1.3 Species Grouping

We implemented a template-based approach where the LLM is provided with a suggested list of functional groups (template) and is instructed to assign taxonomic groups to those groups - with permission to expand the list if it cannot match a taxonomic group to one of the provided functional groups, and the possibility of returning fewer groups than are provided. The framework uses a user-defined grouping template (provided in Section S4 of the supplementary material) that leverages a user's ecosystem modelling experience while allowing for regional customization by either the user or LLM. Due to the complexity of defining ecological groups for EwE models, we have implemented additional template-generation options but do not use them for validation in this study (See the code repository for more details).

Because OBIS can return thousands of species for a given region, instead of using an LLM to classify each species individually, which is time- and cost-prohibitive, we group species hierarchically to reduce the number of classifications required. The framework does this iteratively traversing the resulting OBIS database, from kingdom to species, classifying taxonomic groups into functional groups at finer and finer resolutions. Starting at the Kingdom level, the LLM is asked to classify taxa into functional groups. Taxa that the LLM does not think fall neatly into a specific functional group undergo evaluation at finer taxonomic levels until reaching a definitive group assignment or finally reaching the taxonomic level of species, at which point an assignment must be made.

For example, when classifying something like the Western Australian Dhufish (*Glaucosoma hebraicum*), after passing through the Kingdom Animalia, the phylum Chordata is evaluated. Since Chordata includes diverse feeding strategies from filter-feeding tunicates to predatory fish, the LLM, possessing this knowledge innately from its training, marks it for resolution at a finer level. At the class level, the LLM evaluates Actinopterygii, which is again marked for resolution due to its diverse feeding strategies. Continuing through the taxonomic hierarchy, the family Glaucosomatidae is eventually reached, where all members share similar ecological roles as demersal predators, allowing classification into the demersal carnivore functional group. This hierarchical approach significantly reduces the number of required classifications, although

is vulnerable to misclassifications at higher taxonomic levels if the LLM does not have sufficient ecological capability. The success of this approach is highly dependent on the ability of the groups in the template to properly capture the overall ecological relations that are needed to model the research question. Success is also dependent on the LLM's ability to understand the ecological roles of taxa and is a key target for validation in this study. We provide an initial evaluation of the quality of this LLM-generated grouping in Section S3.

At each taxonomic level, the LLM evaluates taxa against the selected grouping template using the following prompt (where square brackets indicate dynamically updated variables):

> You are classifying marine organisms into functional groups for an Ecopath with Ecosim (EwE) model. Functional groups can be individual species or groups of species that perform a similar function in the ecosystem, i.e. have approximately the same growth rates, consumption rates, diets, habitats, and predators. They should be based on species that occupy similar niches, rather than of similar taxonomic groups.
>
> Examine these taxa at the [rank] level and assign each to an ecological functional group.
>
> Rules for assignment:
>
> - If a taxon contains members with different feeding strategies or trophic levels, assign it to 'RESOLVE'
> - Examples requiring 'RESOLVE':
> - A phylum containing both filter feeders and predators
> - An order with both herbivores and carnivores
> - A class with species across multiple trophic levels
> - If all members of a taxon share similar ecological roles, assign to an appropriate group
> - Only consider the adult phase of the organisms, larvae and juveniles will be organized separately
> - Only assign a definite group if you are confident ALL members of that taxon belong to that group
>
> Taxa to classify: [List of taxa]
>
> Available ecological groups (name: description): [List of available groups and their descriptions]
>
> Return only a JSON object with taxa as keys and assigned groups as values.

When the research focus indicates groups requiring higher resolution (e.g., commercial fisheries species, or a specific species of conservation concern), the following additional guidance is added to the prompt:

Special consideration for research focus: The model's research focus is: [research focus]

When classifying taxa that are related to this research focus:

- Consider creating more detailed, finer resolution groupings
- Keep species of particular interest as individual functional groups
- For taxa that interact significantly with the focal species/groups, maintain higher resolution groupings
- For other taxa, broader functional groups may be appropriate

The framework maintains complete provenance information, including the source of group definitions and any AI-suggested modifications. The system automatically includes a Detritus functional group to represent non-living organic matter in the ecosystem. For fisheries-related work, users could also include a discards group that is split off of this general 'Detritus' category. Finally, a detailed grouping report is produced which documents all of the classification decisions for later human review. This allows for a human user to quickly assess the LLM's decision-making and flag any potential mistakes.

2.1.4 Diet Matrix Construction

After the framework has assigned species to groups, the species-level diet and ecological information collected in Stage 2 is re-assigned to the new functional groups. The diet matrix construction involves two LLM-driven steps. First, the framework assembles text data from various sources (RAG search results, diet data, and GLOBI interaction data) into a structured profile for each group. This profile is passed to the LLM to generate an initial diet composition summary. The following prompt guides this first LLM analysis:

Based on the following information about the diet composition of [group], provide a summary of their diet. Include the prey items and their estimated proportions in the diet.

Available functional groups and their details: [List of groups with descriptions and top species]

Here is the diet data for [group]: [Combined data including RAG search results, compressed food categories, and GLOBI interactions]

Format your response as a list, with each item on a new line in the following format:

Prey Item: Percentage

For example:

```
Small fish: 40%
Zooplankton: 30%
Algae: 20%
```

```
Detritus: 10%
```

If exact percentages are not available, estimate percentages based on the information you have been provided. Ensure that all percentages add up to approximately 100%. Consider the RAG search results, compressed food categories, and GLOBI data when creating your summary. Pay special attention to the GLOBI interaction counts, which indicate frequency of observed feeding relationships. Note that some species may feed on juvenile or larval forms of other species, which are often classified in different functional groups than the adults.

Sometimes these responses contain functional groups that are not included in the list of accepted groups or do not add up to 100%. Therefore, the initial diet summaries are passed to a second LLM step that standardizes the proportions and maps any yet undefined group to the already-defined functional groups. This second step converts the approximate summaries into a structured diet matrix, with prey items as rows and predators as columns. Each cell contains the proportion of the predator's diet comprised of that prey item. The diet matrix is then output as a CSV file for use in Ecopath with Ecosim models.

When prey items do not exactly match functional group names, we employ a hierarchical matching system. The system first attempts exact matches, then falls back to case-insensitive partial matching using species names. For example, if the AI returns a prey item "snapper" that doesn't exactly match any functional group, the system would match it to a functional group containing "snapper" in its name such as "Pink Snapper".

This process is the second target of validation in this study and is evaluated in Section S2. The complete codebase and configuration files are available at [GitHub repository URL].

2.2 Validation

Our validation framework assesses both the precision (consistency across multiple runs) and accuracy (comparison with expert-created matrices) of the LLM-driven model construction process. This dual approach provides comprehensive insights into the framework's reliability and ecological validity. We used four distinct Australian marine regions for these assessments: three regions (Northern Australia, South East shelf, and South East Offshore) to evaluate precision, and one region (Great Australian Bight) to evaluate accuracy.

We executed model generation across three distinct phases. In phase one, we established baseline configurations for each study region by processing species occurrence data and downloading relevant species data. In phase two, where the LLM is first called, we executed five independent iterations per region, maintaining fixed input parameters while allowing the LLM's stochastic decision processes to generate variation in outputs. In phase three, we conducted detailed statistical analyses of both precision across iterations and accuracy compared to expert-created matrices.

For the precision assessment, we calculated the Jaccard similarity coefficients between all possible pairs of iterations and analysed the coefficient of variation in diet proportions. For the accuracy assessment, we averaged the diet proportions of the five

LLM-generated matrices and compared the resulting matrix with an expert-created matrix for the Great Australian Bight ecosystem (C. Bulman pers. comm.) that was used to inform (Fulton et al., 2018).

2.2.1 Study Regions

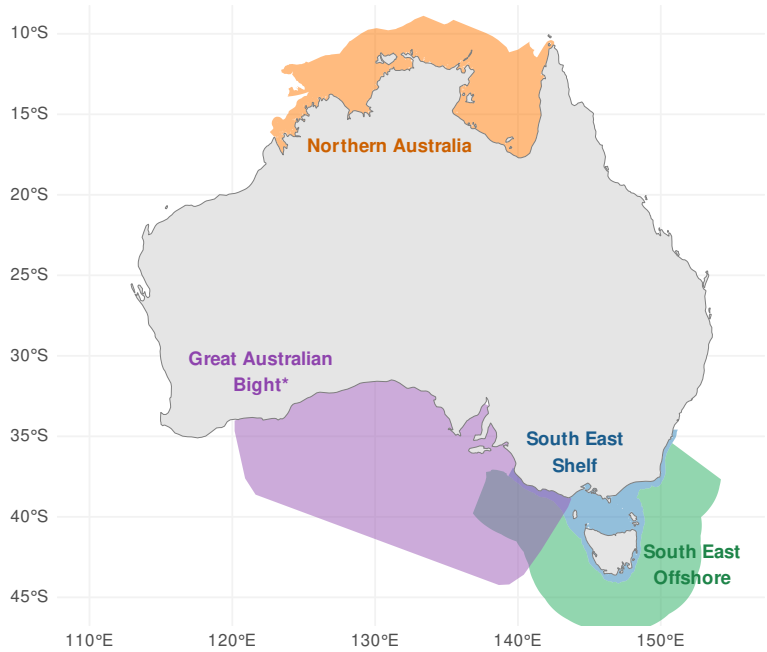

Figure 2: Map of the four study regions used to validate our approach: three of the regions were used to test the precision (consistency) of the approach: Northern Australia (orange), South East shelf (blue), South East Offshore (green). We used the fourth region, the Great Australian Bight (purple) to test the accuracy of our approach by comparing it to the initial diet matrix used in (Fulton et al., 2018)

We selected four Australian marine regions that present distinct ecological characteristics and modelling challenges for our framework (Figure 2). For precision assessment, we used three regions: the Northern Australia region, which represents a tropical ecosystem characterised by a broad shelf and complex mix of reef systems, seagrass meadows, mangrove forests and bare sediment communities with seasonal monsoon influences; the South East shelf region, a temperate coastal system with a network of rocky reefs and kelp forests, rapidly changing environmental conditions due to climate change, comprehensive diet information in established databases, well-documented EwE models spanning multiple years, and active research programs; and the South East Offshore region, a deep-water ecosystem that challenges the framework with data-limited conditions and unique ecological patterns relating to oceanic through flow, shifting current patterns, low productivity patches interspersed with production concentrating canyons and seamounts.

For accuracy assessment, we used the fourth region: the Great Australian Bight (GAB), which represents a region of high conservation significance spanning diverse habitats. The GAB has been extensively studied, with research characterising the shelf

and slope ecosystems from phytoplankton through to marine mammals and birds (Goldsworthy et al., 2013; Fulton et al., 2018). We used this region to compare the resulting diet matrix from our process to the expert-created matrix used in (Fulton et al., 2018), allowing us to evaluate the accuracy of our approach.

The contrasting characteristics of these four regions provide a robust test of the framework's adaptability across different ecological contexts.

2.2.2 Grouping Precision Analysis

To assess the precision of AI-generated species groupings, we developed quantitative measures of grouping precision. For each of the three regions used for precision assessment (Northern Australia, South East shelf, and South East Offshore), we conducted five independent iterations, resulting in five grouping outcomes per region. We analyzed precision within each region separately.

For each region, we tracked each species' group assignments across the five iterations and calculated a precision score:

$$\text{Precision Score} = \frac{\text{Number of occurrences in most common group}}{\text{Total number of iterations (5)}}$$

This metric quantifies the framework's decision-making reliability for individual species within a specific ecological context. We classified species with precision scores below 1.0 as unstable, indicating variable group assignments across iterations. A species with a precision score of 1.0 was assigned to the same functional group in every iteration, demonstrating high precision in the framework's decision-making.

While the precision score measures stability at the individual species level, we also needed to evaluate stability at the group level. To do this, we assessed group stability using the Jaccard similarity coefficient between all possible pairs of iterations within each region:

$$J(i, j) = \frac{|M_i \cap M_j|}{|M_i \cup M_j|}$$

where M_i and M_j represent the sets of species members in iterations i and j. Unlike the precision score, which focuses on whether individual species are consistently assigned to the same group, the Jaccard similarity measures whether a group consistently contains the same set of species across iterations. For example, a group might maintain a stable core of species while experiencing minor variations in peripheral members, which would be captured by the Jaccard similarity but not by individual species precision scores.

For each region, we calculated the overall stability score by averaging Jaccard similarities across all possible pairs of iterations. This approach reveals how consistently the framework identifies and maintains ecologically meaningful groupings across different runs. One-way ANOVA tests on these stability measures across regions, supplemented with Cohen's f effect size calculations, demonstrate the framework's precision across different marine ecosystems while maintaining consistent decision-making patterns.

2.2.3 Grouping Accuracy Assessment

To assess the ecological validity of AI-assigned functional groups, we conducted a manual validation of taxonomic grouping decisions. We did not use the original expert groupings from (Fulton et al., 2018) for two reasons: first, the AI system handled many more taxonomic groupings and species than the human-created groupings, and second, comprehensive data for each functional group was no longer avaialable to do a direct comparison. Therefore, we undertook a direct evaluation where the lead author, SS, reviewed each taxonomic entity (ranging from species to higher taxonomic levels) for a single iteration of the GAB (n=675 AI-decisions) and evaluated whether the AI system had correctly assigned it to an appropriate functional group based on known ecological characteristics.

For each taxonomic group assigned by the AI system to a functional group, we researched the known ecological characteristics of that taxon, including feeding behavior, habitat preferences, and trophic position. We compared these ecological characteristics to the description of the functional group provided in the grouping template. Based on this comparison, we classified each assignment as either "Correct" (the taxon fits well within the functional group), "Partial" (the taxon partially fits the functional group but has some characteristics that don't align), "Incorrect" (the taxon was inappropriately assigned), or "Not Sure" (insufficient information was available to make a determination).

2.2.4 Diet Matrix Precision Assessment

To evaluate the precision of AI-generated trophic interactions and assess the framework's ability to capture distinct ecological patterns, we developed a multi-metric analysis approach. For each region separately (with five iterations per region), we calculated the following metrics for each predator-prey interaction:

1. Presence ratio across iterations:

$$P_{ij} = \frac{\text{Number of iterations with interaction}}{n}$$

   where n is the total number of iterations (5 per region), and an interaction is present when the diet proportion $x_{ijk} > 0$ for predator i consuming prey j in iteration k.

2. Mean diet proportion:

$$\mu_{ij} = \frac{1}{n} \sum_{k=1}^{n} x_{ijk}$$

   where x_{ijk} represents diet proportion for predator i consuming prey j in iteration k.

3. Stability score:

   We first calculate a normalized deviation score for each predator-prey interaction:

$$D_{ij} = \frac{1}{n} \sum_{k=1}^{n} \frac{|x_{ijk} - \mu_{ij}|}{\max_k(|x_{ijk}|)}$$

where μ_{ij} is the mean diet proportion across iterations, and $\max_k(|x_{ijk}|)$ is the maximum absolute value across iterations. Because diet proportions (x_{ijk}) are bounded between 0 and 1, this deviation score ranges from 0 to 0.5.

Then, to create a more intuitive stability score where higher values represent greater stability, we invert this measure:

$$S_{ij} = 1 - D_{ij}$$

This transformation yields a stability score bounded between 0.5 (maximum instability) and 1 (perfect stability), with higher values indicating more consistent diet proportions across iterations.

We chose this stability metric over traditional variance measures for several reasons. First, by normalizing deviations by the maximum value, the metric achieves scale independence, allowing meaningful comparisons between interactions of different magnitudes. For example, the sequences [0.2, 0.2, 0.2, 0.2, 0.1] and [0.02, 0.02, 0.02, 0.02, 0.01] would yield the same stability score despite having different absolute variances. Second, the bounded range between 0 and 1 provides an intuitive scale for assessing stability, unlike the unbounded nature of variance. Third, when diet proportions are of similar magnitude across iterations, this approach prevents minor fluctuations in small values from disproportionately influencing the stability assessment. However, in cases where iterations contain both very small and substantially larger values, the scale independence property means the stability assessment will be more sensitive to relative deviations in the larger values.

We classified interactions as unstable when their stability score fell below 0.7, corresponding to a normalized deviation of 0.3 in the original metric. This approach balances sensitivity to meaningful ecological variation while avoiding flagging minor fluctuations that are expected in complex ecological systems. Variations in predator-prey interaction strengths beyond this threshold suggest fundamental uncertainty in the trophic relationship that would propagate through ecosystem simulations and affect model predictions. To illustrate this metric:

- A stable interaction (S = 0.92) might show values [0.02, 0.02, 0.02, 0.02, 0.01], where proportions remain very similar across iterations
- An unstable interaction (S = 0.61) might show values [0.027, 0.25, 0.25, 0.067, 0.25], where proportions vary substantially between iterations, indicating inconsistent characterization of the predator-prey relationship by roughly an order of magnitude

This metric provides a continuous measure of stability that handles both presence/absence patterns and magnitude variations in a unified way. To assess the framework's ability to capture distinct ecological patterns across regions, we employed pairwise Spearman correlations between iterations to evaluate the precision of predator-prey relationships. This non-parametric approach accounts for the potentially non-normal distribution of diet proportions. We supplemented this with Kruskal-Wallis tests to identify significant differences in trophic structure across regions, providing evidence of the framework's ability to distinguish unique ecological characteristics in different marine ecosystems.

2.2.5 Diet Matrix Accuracy Assessment

To evaluate the accuracy of AI-generated diet matrices against expert knowledge, we conducted a detailed comparison using the Great Australian Bight (GAB) ecosystem model developed by Fulton et al. (2018). We obtained the original, unbalanced diet matrix constructed by expert marine ecologists (C. Bulman, personal communication) and compared it with five independently generated AI matrices for the same region. To specifically assess the diet proportion accuracy we provide the AI system with a grouping template consisting of the same list of groupings from the extant GAB model, thus testing the system's ability to sort species into the correct groups and then assign diet proportions according to those group.

The analysis examined two fundamental aspects of the diet matrices: the structural patterns of predator-prey relationships and the quantitative diet proportions. We first assessed structural agreement by identifying matching and mismatching interactions between the expert and AI matrices. This binary presence-absence analysis yielded counts of concordant interactions, where both matrices agreed on the presence or absence of a feeding relationship, and discordant interactions where one matrix indicated a link while the other did not. We quantified the overall agreement using Cohen's Kappa coefficient, supplemented by true positive and negative rates to characterise the framework's ability to replicate expert-identified trophic relationships.

For predator-prey pairs where both matrices indicated an interaction, we conducted quantitative comparisons of the diet proportions. We calculated the Pearson correlation coefficients to measure the relationship between expert and AI-assigned proportions. We chose this measure because, unlike the stability metrics used in the precision assessment which evaluate consistency across multiple iterations, correlation analysis is specifically designed to quantify the alignment between two distinct matrices—the AI-generated and expert-created matrices. This approach directly addresses the accuracy objective by measuring how well the AI-generated diet proportions correspond to expert knowledge, rather than measuring consistency across multiple AI-generated iterations. We performed these analyses both at the whole-matrix level and for individual predator groups, enabling identification of systematic patterns in the framework's performance across different taxonomic groups.

3 Results

Our validation framework assessed three key aspects of the AI-assisted ecosystem modelling approach: reproducibility of species groupings, consistency of diet matrix construction, and accuracy against expert-derived matrices.

3.1 Species Grouping Reproducibility

3.1.1 Classification Consistency Analysis

The framework successfully reduced ecological complexity while preserving meaningful biological relationships. Starting with 63 potential functional groups provided in the default template (See S4), it identified 34-37 region-specific groups. Chi-square tests confirmed the non-random nature of these groupings, showing consistent species assignments across all regions ($p < 0.001$). This statistical significance provides strong

evidence that the framework makes systematic grouping decisions rather than arbitrary assignments.

The framework achieved high classification stability for groups across all regions. Mean consistency scores, where 1.0 represents identical species assignments to groups across all groups and within-region iterations, were exceptionally high: 0.997 for both Northern Australia and South East shelf, and 0.998 for South East Offshore. This translated to very low proportions of species that were variably classified across the five iterations: only 0.99% (103 species) in Northern Australia, 1.06% (125 species) in South East shelf, and 0.73% (87 species) in South East Offshore. These results demonstrate that the framework's classifications remained stable despite the stochastic nature of the AI decision-making process.

Among the small percentage of variably classified species, we identified consistent patterns of classification instability (Table 1). These species typically oscillated between ecologically similar functional groups, such as macrozoobenthos and benthic infaunal carnivores in the Northern Australia, or piscivores and deep demersal fish in the South East Inshore region. This suggests that classification uncertainty occurs primarily at ecological boundaries where functional roles overlap.

The Jaccard similarity indices reveal high overall stability in group membership across all three regions (Figure 3), with most functional groups showing indices above 0.95. The groups labelled in the figure represent those groups with lower stability indices.

Further detailed analysis of group stability patterns across regions is provided in S3.

Table 1: Dominant patterns of species classification instability across three study regions. The table presents the most frequent oscillation patterns between functional groups for species that were inconsistently classified across the five framework iterations. For each region, the total number of variably classified species is shown (representing less than 1.1% of all species), along with the percentage distribution of different oscillation patterns.

Region	Most Common Pattern	Count	% of Total
Northern Australia (103 species)	Macrozoobenthos ↔ Benthic infaunal carnivores	28	27.2%
	Benthic filter feeders ↔ Deposit feeders	25	24.3%
	Prawns ↔ Macrozoobenthos	21	20.4%
	Other patterns	29	28.1%
South East Shelf (125 species)	Piscivores ↔ Deep demersal fish	42	33.6%
	Benthic grazers ↔ Benthic carnivores	31	24.8%
	Planktivores ↔ Mesopelagic fish	28	22.4%
	Other patterns	24	19.2%
South East Offshore (87 species)	Benthic filter feeders ↔ Benthic carnivores	25	28.7%
	Macrozoobenthos ↔ Deep demersal fish	22	25.3%
	Mesozooplankton ↔ Macrozoobenthos	18	20.7%
	Other patterns	22	25.3%

Note: Arrows indicate group assignment oscillation between iterations. Complete species-level data available in Section S3 of the supplementary material.

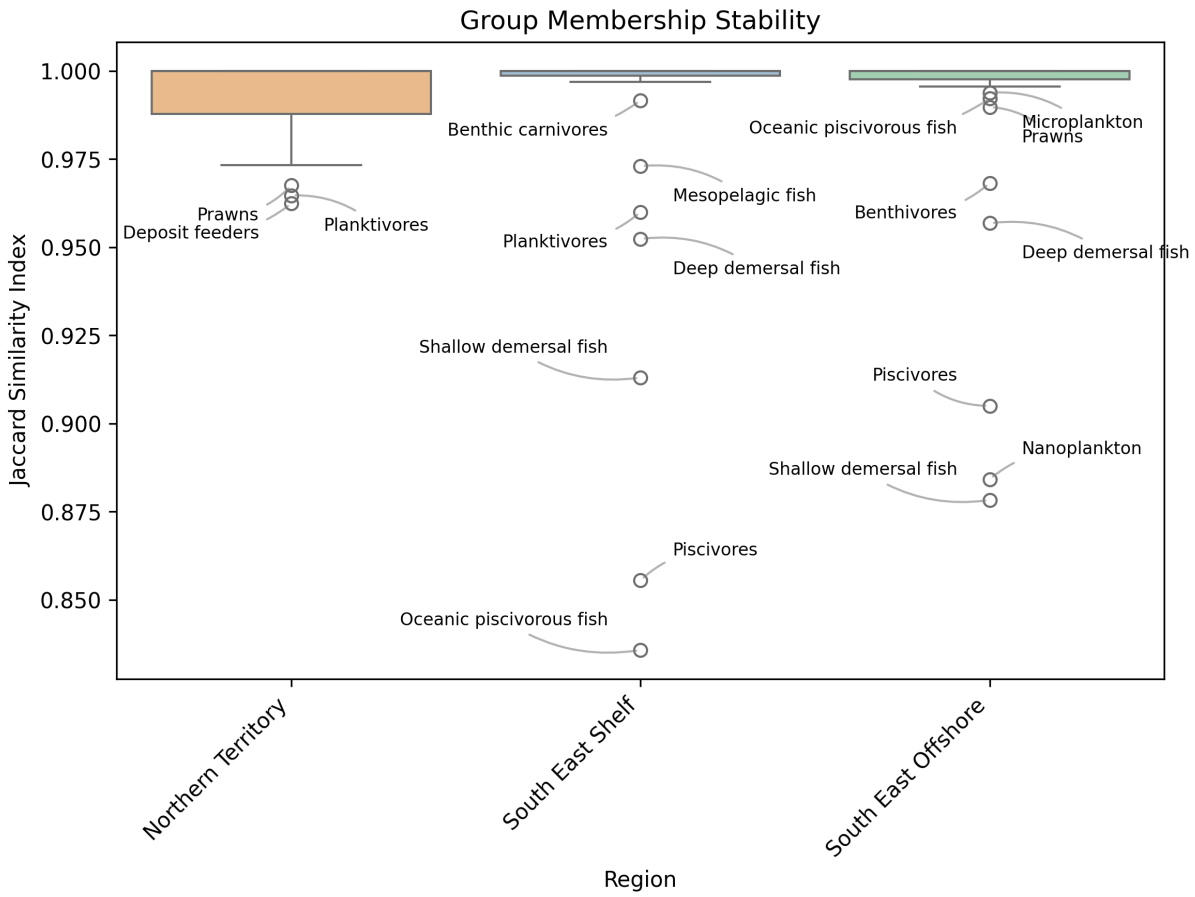

Figure 3: Group membership stability across three regions measured by Jaccard similarity index (0.85-1.0). Most groups show high stability (>0.95), with labelled points out-of-distribution outliers that exhibit lower stability.

3.2 Diet Matrix Reproducibility

3.2.1 Trophic Interaction Consistency

The framework identified consistent trophic relationships across all regions, with the Northern Australia showing 358 interactions (58.7% having stability scores > 0.7), South East shelf 380 interactions (51.3% of which were stable), and South East Offshore 477 interactions (56.0% stable). As shown in Figure 4, the distribution of stability scores across regions demonstrates that most interactions cluster above the 0.7 threshold, with a substantial proportion achieving near-perfect stability (scores approaching 1.0). Spearman correlations between iterations demonstrated that the relative proportions of different prey in predator diets remained fairly consistent across all regions (Northern Australia: $\rho = 0.72 - 0.89$; South East shelf: $\rho = 0.68 - 0.85$; South East Offshore: $\rho = 0.70 - 0.87$), even when absolute proportions varied. Detailed diet matrices for each region are provided in S2.

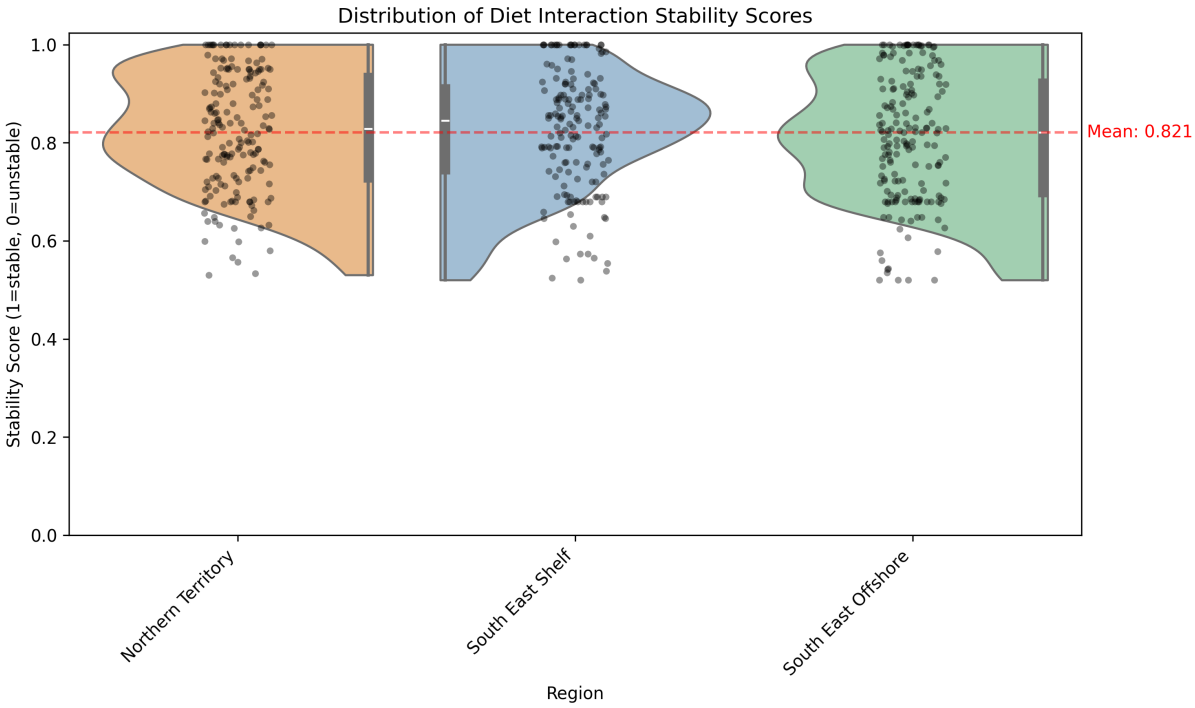

Figure 4: Distribution of diet interaction stability scores across regions for substantial interactions (those comprising more than 5% of a predator's diet). Half-violin plots show the density of stability scores (1=stable, 0=unstable), with embedded box plots indicating quartiles and median. Individual points represent specific predator-prey interactions, and the red dashed line shows the mean stability score across all regions. The distributions are bounded at one, reflecting perfect stability, with most interactions showing scores above 0.7. Stability scores quantify the consistency of predator-prey interactions across iterations, where a score of 1.0 indicates the interaction was identified with identical diet proportions in all iterations, while lower scores reflect either variable diet proportions or inconsistent identification of the interaction.

3.3 Grouping and Diet Proportion Accuracy Assessment: Great Australian Bight Case Study

3.3.1 Taxonomic Grouping Accuracy

To evaluate the ecological validity of AI-assigned functional groups, we conducted a detailed manual validation of 675 taxonomic grouping decisions. The results revealed that 75.3% (508) of the AI's taxonomic assignments were fully correct, aligning with known ecological characteristics of the taxa. Additionally, 17.3% (117) of assignments were partially correct, where the taxon fit some but not all aspects of the functional group description. For example, these included cases where the AI system designated a taxonomic group as 'deep' or 'slope' when they might inhabit both, or might designate a taxonomic group as 'large' or 'small' when members could be one or the other. Only 3.4% (23) of assignments were clearly incorrect (demonstrable incorrect feeding strategy habitats), and 4.0% (27) could not be definitively assessed due to limited ecological information about the taxa (mostly poorly researched deep-water taxonomies).

The accuracy of assignments varied considerably across functional groups (Figure 5). Many functional groups showed perfect or near-perfect assignment accuracy, including all assignments for albatross, pelagic sharks, small phytoplankton, mesozooplankton, small petrels, and several other well-defined groups. These groups typically have clear ecological niches and distinctive characteristics that facilitate accurate classification.

Functional groups with lower accuracy rates appeared to be constrained by limitations in the AI system's grouping template, particularly for specialized ecological niches. For example, waterfowl were frequently misclassified into the "Shags and cormorants" group, achieving only 33.3% correct assignments with 50% incorrect assignments. These errors revealed confusion between taxonomically related but ecologically distinct bird groups. Similarly, "Deep filter feeders" showed only 28.6% correct assignments with 71.4% partial assignments, highlighting challenges in classifying deep-sea organisms with complex or variable feeding strategies.

The analysis of partial assignments revealed several recurring patterns. Taxa associated with deep-sea environments (17 taxa) were frequently misclassified, likely due to limited ecological information and the complex nature of deep-sea ecosystems. Parasitic organisms (7 taxa) were also challenging to classify correctly, as they often have complex life cycles that span multiple functional roles. Filter feeders, detritivores, and grazers showed similar patterns of partial classification, typically due to their variable feeding strategies that may change based on environmental conditions or life stage.

3.3.2 Diet Matrix Accuracy

To evaluate the framework's accuracy against expert knowledge, we compared its output to an expert-derived Ecopath model of the Great Australian Bight ecosystem (Fulton et al., 2018). The framework demonstrated varying performance across functional groups, successfully matching 59 of 76 expert-defined groups (77.6%). The framework omitted 17 groups present in the expert matrix, including several commercially important species (Southern Bluefin Tuna, Snapper, King George whiting, and Abalone) as well as Nanozooplankton. Conversely, it generated only two groups not present in the expert matrix (Offshore pelagic invertivores large and Slope large demersal omnivores).

As shown in Figure 6a, the framework demonstrated varying levels of agreement across different functional groups in identifying trophic interactions. The analysis revealed that 8.5% of interactions were present in both matrices (dark purple), while 73.1% were correctly identified as absent in both (light grey). The framework uniquely identified 14.5% of interactions (teal) that were not present in the expert matrix, while missing 3.9% of expert-identified interactions (yellow). Overall, the framework achieved an agreement rate of 81.6% with the expert matrix, with a true positive rate (sensitivity) of 0.687 and a true negative rate (specificity) of 0.834.

The framework showed moderate success in capturing the quantitative aspects of diet proportions, with a Kappa coefficient of 0.38 with expert-assigned diet proportions. The distribution of absolute differences in diet proportions (Figure 6b) revealed that the majority of differences were relatively small, with approximately 80% of the differences being less than 0.2. Detailed analysis showed a mean absolute difference of 0.110 (median: 0.058) in diet proportions for interactions present in both matrices. However, a long tail in the distribution (maximum difference: 0.865) indicates some cases where AI-generated proportions diverged substantially from expert values. This suggests that when the framework correctly identified a trophic interaction, it often estimated diet proportions within reasonable bounds of expert values, though with notable variations across different predator-prey combinations.

Further examination of the omitted groups revealed a pattern where the framework tended to miss specialized ecological groups, particularly those comprised of only a single species (e.g., Southern Bluefin Tuna, Snapper, King George whiting). This suggests a bias toward generalized classifications that may overlook management-relevant distinctions. This limitation was most evident in commercially important species that typically receive individual attention in expert-created models but were subsumed into broader functional groups by the framework. A comprehensive visualization of these differences across all functional groups is provided in S2.

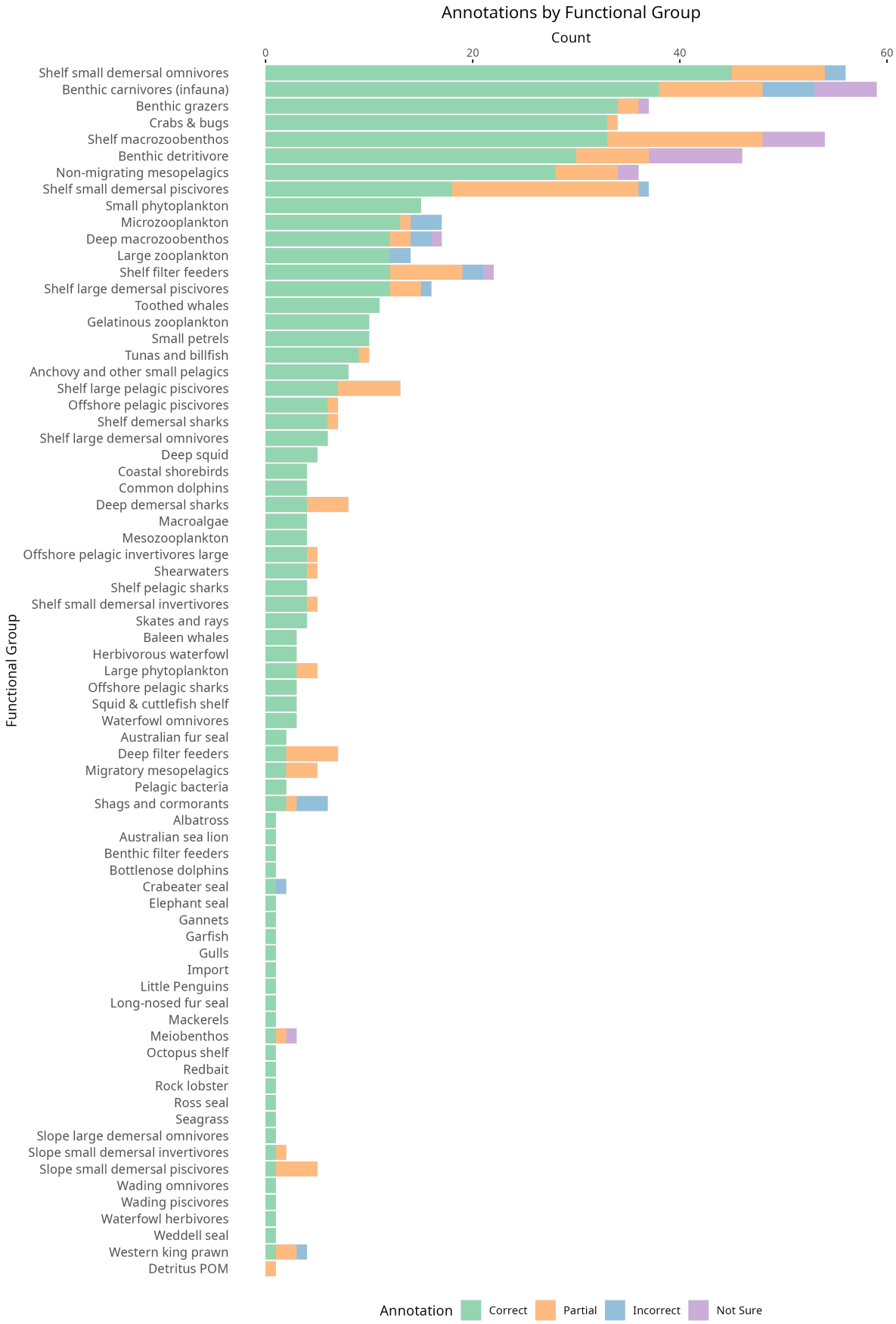

Figure 5: Accuracy of AI-assigned taxonomic groupings by functional group. The chart shows the percentage of correct (green), partial (orange), incorrect (blue), and uncertain (purple) assignments for each functional group.

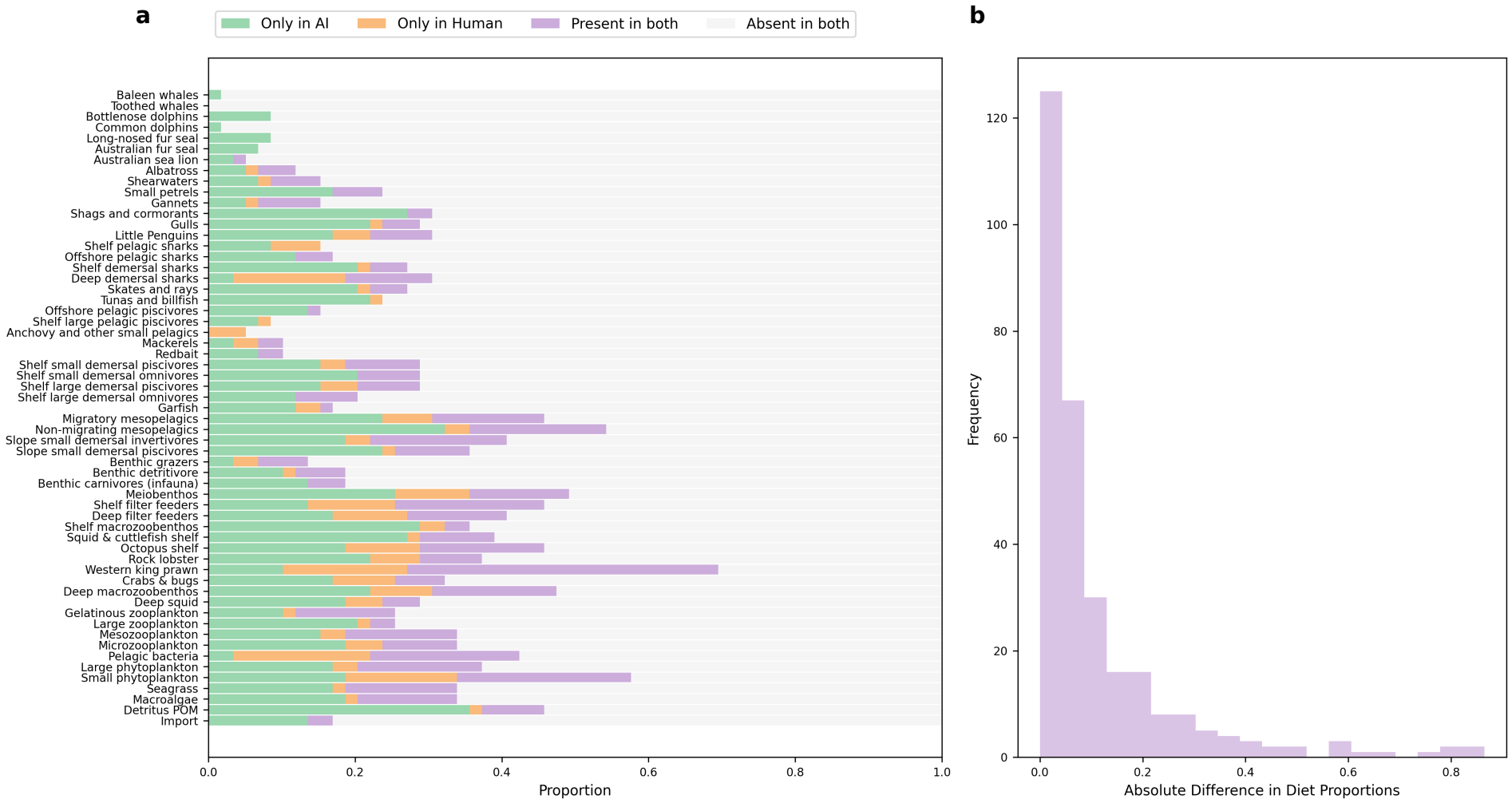

Figure 6: Comparison of expert-created and AI-generated diet matrices for the Great Australian Bight ecosystem. (a) Presence-absence patterns showing the proportion of different interaction types across functional groups. Dark purple indicates interactions present in both matrices, orange shows expert-only interactions, teal shows AI-only interactions, and light grey indicates absence in both. (b) Distribution of absolute differences in diet proportions where both matrices indicate an interaction, showing the frequency of different magnitudes of disagreement between AI and expert estimates.

3.4 Framework Implementation and Performance

3.4.1 Scale and Processing Efficiency

We evaluated our framework through five independent runs across three distinct Australian regions, processing a total of 39,722 species. The framework handled 10,621 species in the Northern Australia's tropical reef ecosystem, 17,068 in the South East shelf's coastal and pelagic environments, and 12,033 in the South East Offshore's deep-water systems.

3.4.2 Computational Efficiency

The computational requirements of the AI framework varied across regions. Total processing time ranged from 2.8 to 4.8 hours across regions. The most time-intensive stage was the downloading of biological data from online databases, accounting for approximately 70% of the total processing time. Species identification typically required 0.01 hours, while the AI-driven species grouping process averaged 0.26 hours. Diet data collection and matrix construction required 0.7 and 0.04 hours respectively, with final parameter estimation taking 0.20 hours. On average, the framework required 0.7 seconds per species for data downloading and 0.2 seconds per species for diet data collection, though these rates varied considerably between regions due to differences in data availability and species complexity.

Table 2: Computational requirements by region and processing stage

Region	Species Count	Processing Time (hours)				
		Identification	Data Download	Grouping	Diet Collection	Matrix Construction
Northern Australia	10,621	0.01	2.2	0.2	0.2	0.04
South East Shelf	17,068	0.01	2.8	0.2	1.6	0.04
South East Offshore	12,033	0.01	3.3	0.4	0.3	0.04
Great Australian Bight	6,957	0.01	1.3	0.3	0.5	0.07

4 Discussion

We have shown that an AI-driven framework is able to construct an important component, the diet matrix, of common ecosystem models, with a fair degree of reliability. This capability addresses a significant challenge in ecosystem modelling, as the increasing use of these models for environmental management and policy decisions requires efficient and accurate development approaches (Weiskopf et al., 2022; Schuwirth et al., 2019). Constructing reliable ecosystem models traditionally involves complex technical challenges, including species identification, data harvesting, and the creation of accurate diet matrices—processes that are particularly demanding in fisheries management contexts where food web models inform ecological and socioeconomic decision-making (Chakravorty et al., 2024). Our framework provides a systematic, AI-assisted solution that enhances reproducibility and dramatically reduces the time investment required from months to hours. This efficiency gain aligns with the growing recognition that effective ecological models must balance mechanistic understanding, appropriate spatial and temporal resolution, and uncertainty quantification

to support decision-making (Schuwirth et al., 2019). By streamlining technical aspects through integration of multiple data sources (Christensen and Walters, 2004; Colléter et al., 2015) and AI-driven synthesis (Spillias et al., 2024c; Noleto Filho et al., 2024), our approach allows modelers to dedicate more resources to stakeholder engagement and result communication, potentially increasing the impact of ecosystem modelling on environmental management.

4.1 Validation Assessment

Here we have demonstrated the framework's potential to generate reproducible, ecologically meaningful components for ecosystem model development while significantly reducing development time. It can complement the traditional approach to model building and expert judgement. The framework's ability to construct reliable ecosystem model components reveal both strengths and limitations. The framework demonstrates strong internal consistency, with high stability scores (98.8-99.6%) in species classifications across regions and robust correlations in predator-prey rankings ($\rho = 0.72 - 0.89$). This consistency suggests the framework makes systematic rather than arbitrary decisions in constructing ecological relationships. However, these metrics must be interpreted cautiously, as they reflect the framework's reproducibility rather than ecological accuracy.

Our performance metrics are comparable to other recent AI applications in food web modelling, such as FoodWebAI (Noleto Filho et al., 2024), which achieved 90-98% accuracy in correctly determining species' positions in the food chain (trophic levels) and 74% accuracy in identifying predator-prey relationships (trophic links) across three ecosystems. The grouping accuracy we found was less than others have reported for bespoke chatbots but higher than has been reported for frontier LLMs of the previous generation (Sahu et al., 2024).

The Great Australian Bight comparison against expert knowledge provides important insights into the framework's reliability. The high success rate in identifying absent trophic interactions (73.1%) indicates the framework effectively avoids spurious ecological connections. However, the moderate Kappa coefficient (0.38) with expert-assigned diet proportions and the tendency to miss specialised ecological groups reveals important limitations. The framework shows a bias toward generalised classifications that may overlook management-relevant distinctions, particularly when it comes to accommodating functional groups that are comprised of only a single species.

Our detailed taxonomic validation analysis further illuminates the framework's ecological accuracy, with 75.3% of taxonomic assignments being fully correct and an additional 17.3% being partially correct. This high overall accuracy rate (92.6% at least partially correct) is consistent with other analyses (Noleto Filho et al., 2024; Dorm et al., 2025) and suggests the framework generally makes ecologically sound grouping decisions. However, the analysis also revealed that the success of the grouping may depend on the comprehensiveness of the grouping template that the AI system is provided. Further, deep-sea organisms, parasitic taxa, and species with complex or variable feeding strategies were more frequently misclassified or only partially correctly assigned. These findings highlight areas where the framework's ecological knowledge may be limited or where the predefined functional group templates may not adequately capture the full range of ecological roles present in marine ecosystems. It also highlights where human triage will be needed when using these methods, show-

ing the kinds of groups where the modeller needs to focus checking or redefine model structures to reduce erroneous assignments. However, the uncertainty exhibited by the framework in classifying certain taxa, particularly deep-sea organisms and species with complex feeding strategies, should not be viewed solely as a limitation but also as a feature that appropriately reflects genuine scientific uncertainty. When knowledge about certain ecological niches is limited, a high level of certainty in classifications would be misleading. The framework's variable classifications and partial assignments for these challenging taxa actually represent a more honest representation of our current ecological understanding, aligning with principles of scientific transparency in uncertainty communication.

The framework's handling of ecological complexity shows mixed results. While it successfully captures broad trophic patterns and adapts to regional differences, its treatment of species that span multiple functional roles needs improvement. For instance, the variable classification of anemones and flatfishes between functional groups, while partially reflecting natural ecological flexibility, suggests the need for more nuanced classification approaches. Whilst here we have relied on the LLM's 'embedded' ecological knowledge to classify species, providing additional ecological information from online databases may improve the quality of grouping assignments. Or perhaps, with the rapid rate of LLM capacity improvement, future LLM's will perform better at ecological tasks such as these. The identification of additional trophic interactions not present in expert matrices (14.5%) requires careful evaluation - these could represent either over-connection or potentially valid relationships that merit further investigation.

These validation outcomes suggest the framework can serve as a useful starting point for ecosystem model development, particularly in its ability to avoid implausible ecological connections and maintain consistent broad-scale trophic structures. However, its outputs require expert review, especially for specialised ecological groups and complex trophic relationships or where there is only a qualitative understanding of ecoystem function. The balance between the framework's systematic approach and the need for ecological expertise emerges as a key consideration for its practical application.

4.2 Applications for EBFM

Given the validation outcomes, the framework shows promise as a rapid prototyping tool for model developement for ecosystem-based fisheries management. Its demonstrated ability to avoid spurious ecological connections while maintaining consistent broad-scale trophic structures makes it valuable for accelerating initial model development, reducing construction time from months to hours, with total processing times ranging from 2.6 to 4.9 hours per region. However, the framework's limitations with specialised groups and single-species functional units means it should be used as a starting and accelerator for expert refinement rather than a standalone solution. Further studies that explore the impact of prompt engineering and template choice on the quality of outcomes for a given context will help further demonstrate the frameworks capabilities and limitations.

The framework's systematic approach to uncertainty quantification helps identify where additional data collection or expert input is most needed. For instance, its higher performance in identifying absent interactions versus capturing expert-identified rela-

tionships suggests where manual review should focus. This aligns with approaches to uncertainty-aware ecosystem management (Hill et al., 2007; Link et al., 2012), while the framework's consistent methodology across regions supports standardised approaches to model development across jurisdictions.

4.3 Limitations and Uncertainties

Our validation results highlight three key limitations of the framework. First, the framework's bias toward generalised classifications, evidenced by its difficulty with single-species functional groups in the GAB comparison, reflects fundamental limitations in how the AI system processes ecological relationships. Second, the framework's reliance on Claude 3.5 Sonnet, a closed-source large language model, introduces scientific reproducibility challenges. While our validation demonstrates consistent performance, we cannot fully examine the model's decision-making process or potential biases. Third, practical implementation faces computational and data-related constraints. Data harvesting operations proved time-intensive, and the framework's performance varied with data availability across regions and ecological roles. Future iterations might benefit from exploring open-source alternatives (Kommineni et al., 2024) and developing more transparent decision-making processes. Additionally, fishery and ecosystem managers need to trust that AI or hybrid approaches can reliably construct and parameterize models, requiring careful attention to risk, uncertainty, and transparency in model development. Decision makers often perceive ecological models as "black boxes" with questionable data inputs (Boschetti et al., 2018), which may be amplified with AI-based approaches.

4.4 Future Development and Assessment

To address the identified limitations, several key areas require further development. First, the framework's handling of specialised ecological groups needs improvement, particularly for commercially important single-species units. This could involve developing more sophisticated protocols for identifying and preserving management-relevant distinctions during the grouping process, as different key groupings are more important for fishery management vs. spatial planning, for example. Second, to enhance scientific reproducibility, future versions should explore the capability of other LLMs, including open-source LLMs which can be more transparently assessed than proprietary models like Claude.

Third, systematic validation across diverse ecosystem types is needed to establish operational boundaries. This validation should encompass a range of ecosystems with varying structures, biodiversity levels, and data availability—including polar regions, coral reefs, deep ocean habitats, pelagic systems, and upwelling zones—and across multiple spatial scales from ocean basins to coastal bays. Testing should pay particular attention to how the framework handles specialised ecological roles in different contexts. Throughout this development process, the framework should maintain its role as a collaborative tool that complements rather than replaces expert judgment, focusing on rapid prototyping while preserving the critical role of ecological expertise in model refinement.

Fourth, we rely heavily on online databases which are subject to data quality issues and biases. Future work should explore how to incorporate local knowledge and

expert judgment into the framework to address these limitations. This could involve developing more sophisticated data integration methods that combine structured data from online sources with unstructured local knowledge, as well as exploring how to incorporate expert feedback into the AI decision-making process - potentially involving weighting the importance of certain sources of data for the LLM. Additionally, testing whether using local or national databases (e.g., Fishes of Australia) alongside global databases provides relevant information missed in global databases would be valuable to determine if region-specific data sources can improve the framework's ecological accuracy. Current movements towards FAIR data principles (Tanhua et al., 2019) will likely also improve the ability for AI systems to find the most accurate and relevant data sources.

Fifth, future development should incorporate established best practices for ecological model building to enhance quality and reliability. The Ecopath with Ecosim approach offers methodological standards (Christensen and Walters, 2004; Heymans et al., 2016) and pre-balance diagnostics (Link, 2010) that could strengthen AI-assisted frameworks. More broadly, Good Modelling Practice (GMP) principles (Jakeman et al., 2024) emphasize the importance of explicating modelling choices throughout the entire modelling lifecycle to build trust in model insights within their social and political contexts. Integrating these established practices would improve model assessment, facilitate evaluation by management bodies, and help normalize rigorous standards across the modelling community; particularly important as ecological network models increasingly inform resource management decisions.

Finally, the framework's utility for ecosystem-based fisheries management should be further explored through case studies that evaluate its effectiveness in building models that support management decisions. This could involve comparing the performance of models built using the framework to those built using traditional methods, as well as assessing the framework's ability to support management-relevant analyses such as scenario testing and policy evaluation.

Acknowledgements

SS was funded by a CSIRO R+ Postdoctoral Fellowship. This project was partially funded by the 'Futures of Seafood' project.

Data Availability

The complete codebase, including all scripts, configuration files, and analysis tools, is available at https://github.com/s-spillias/AI-EwE-Diets. The validation framework, including reference group definitions and classification rules, is documented in the project repository to ensure reproducibility.

Author Contributions

SS: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing,

Visualization, Project administration. BF: Validation, Writing - Review & Editing, Supervision, Funding acquisition. FB: Methodology, Software, Validation, Writing - Review & Editing, Supervision. CB: Investigation, Data Curation, Validation, Writing - Review & Editing. JS: Conceptualization, Validation, Investigation, Writing - Review & Editing. RT: Methodology, Software, Validation, Investigation, Writing - Review & Editing, Supervision, Funding acquisition.

Statement on the Use of Generative AI

Generative AI tools, specifically Claude Sonnet 3.5, were utilized in the preparation of this manuscript to assist with tasks such as language refinement, text structuring, and summarization. All scientific content, data interpretation, and conclusions were independently developed and verified by the authors to ensure accuracy and integrity.

References

Anthropic (2024). Claude 3 model card. Technical documentation, Anthropic.

Boschetti, F., Hughes, M., Jones, C., and Lozano-Montes, H. (2018). On decision makers' perceptions of what an ecological computer model is, what it does, and its impact on limiting model acceptance. *Sustainability*, 10(8):2767.

Castro, A., Pinto, J., Reino, L., Pipek, P., and Capinha, C. (2024). Large language models overcome the challenges of unstructured text data in ecology. *Ecological Informatics*, 82:102742.

Chakravorty, D., Armelloni, E. N., and de la Puente, S. (2024). A systematic review on the use of food web models for addressing the social and economic consequences of fisheries policies and environmental change. *Frontiers in Marine Science*, 11:1489984.

Chamberlain, S. (2020). *robis: An R Client for the Ocean Biodiversity Information System*. R package version 0.2.0.

Chen, M. and Xu, Z. (2024). A deep learning classification framework for research methods of marine protected area management. *Journal of Environmental Management*.

Christensen, V. and Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling*, 172(2-4):109–139.

Chroma (2024). Chroma - the ai-native open-source embedding database. Accessed: 2024.

Coll, M., Akoglu, E., Arreguín-Sánchez, F., Fulton, E. A., Gascuel, D., Heymans, J. J., Libralato, S., Mackinson, S., Palomera, I., Piroddi, C., et al. (2015). Modelling dynamic ecosystems: venturing beyond boundaries with the Ecopath approach. *Reviews in Fish Biology and Fisheries*, 25(2):413–424.

Colléter, M., Valls, A., Guitton, J., Gascuel, D., Pauly, D., and Christensen, V. (2015). Global overview of the applications of the Ecopath with Ecosim modeling approach using the EcoBase models repository. *Ecological Modelling*, 302:42–53.

Dorm, F., Millard, J., Purves, D., Harfoot, M., and Mac Aodha, O. (2025). Large language models possess some ecological knowledge, but how much? *bioRxiv*, pages 2025–02.

Fernandes, S. and D'Mello, A. (2024). Artificial intelligence in the aquaculture industry: Current state, challenges and future directions. *Aquaculture*, page 742048.

Froese, R., Pauly, D., et al. (2010). FishBase.

Fulton, E., Bulman, C., Pethybridge, H., and Goldsworthy, S. (2018). Modelling the Great Australian Bight ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, 157-158:211–235.

Geary, W. L., Bode, M., Doherty, T. S., Fulton, E. A., Nimmo, D. G., Tulloch, A. I., Tulloch, V. J., and Ritchie, E. G. (2020). A guide to ecosystem models and their environmental applications. *Nature Ecology & Evolution*, 4:1459–1471.

Goldsworthy, S. D., Page, B., Rogers, P. J., Bulman, C., Wiebkin, A., McLeay, L. J., Einoder, L., Baylis, A. M., Braley, M., Caines, R., et al. (2013). Trophodynamics of the eastern great australian bight ecosystem: Ecological change associated with the growth of australia's largest fishery. *Ecological Modelling*, 255:38–57.

Grassle, J. F. and Stocks, K. (1999). A global ocean biogeographic information system (OBIS) for the census of marine life. *Oceanography*, 12(3):12–14.

Heymans, J. J., Coll, M., Link, J. S., Mackinson, S., Steenbeek, J., Walters, C., and Christensen, V. (2016). Best practice in ecopath with ecosim food-web models for ecosystem-based management. *Ecological modelling*, 331:173–184.

Hill, S. L., Watters, G. M., Punt, A. E., McAllister, M. K., Quéré, C. L., and Turner, J. (2007). Model uncertainty in the ecosystem approach to fisheries. *Fish and Fisheries*, 8(4):315–336.

Holden, M. H., Akinlotan, M. D., Binley, A. D., Cho, F. H., Helmstedt, K. J., and Chadès, I. (2024a). Why shouldn't I collect more data? Reconciling disagreements between intuition and value of information analyses. *Methods in Ecology and Evolution*, 15:1580–1592.

Holden, M. H., Plagányi, É. E., Fulton, E. A., Campbell, A. B., Janes, R., Lovett, R. A., Wickens, M., Adams, M. P., Botelho, L. L., Dichmont, C. M., et al. (2024b). Cost–benefit analysis of ecosystem modeling to support fisheries management. *Journal of Fish Biology*, 104:1667–1674.

Jakeman, A. J., Elsawah, S., Wang, H.-H., Hamilton, S. H., Melsen, L., and Grimm, V. (2024). Towards normalizing good practice across the whole modeling cycle: its instrumentation and future research topics. *Socio-Environmental Systems Modelling*, 6:18755–18755.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440.

Keck, F., Broadbent, H., and Altermatt, F. (2025). Extracting massive ecological data on state and interactions of species using large language models. *bioRxiv*, pages 2025–01.

Kommineni, V. K., König-Ries, B., and Samuel, S. (2024). Harnessing multiple LLMs for information retrieval: A case study on deep learning methodologies in biodiversity publications. *arXiv preprint arXiv:2411.09269*.

Lapeyrolerie, M., Chapman, M. S., Norman, K. E., and Boettiger, C. (2022). Deep reinforcement learning for conservation decisions. *Methods in Ecology and Evolution*, 13:2649–2662.

Li, H., Liu, J., Wang, Z., Luo, S., Jia, X., and Yao, H. (2024). Lite: Modeling environmental ecosystems with multimodal large language models. *Conference on Language Models (COLM)*.

Link, J. S. (2010). Adding rigor to ecological network models by evaluating a set of pre-balance diagnostics: a plea for prebal. *Ecological modelling*, 221(12):1580–1591.

Link, J. S., Ihde, T., Harvey, C., Gaichas, S. K., Field, J., Brodziak, J., Townsend, H., and Peterman, R. (2012). Dealing with uncertainty in ecosystem models: the paradox of use for living marine resource management. *Progress in Oceanography*, 102:102–114.

Noleto Filho, E. M., Keppeler, F. W., Reis-Filho, J. A., Giarrizzo, T., Keenlyside, N., Carvalho, A. R., Coll, M., Steenbeek, J. G., and Angelini, R. (2024). Foodwebai: Creating and augmenting food webs with natural language models. *Preprint*.

Nugraha, F. S., Saputra, I., Triana, H., Wahyudi, M., Radiyah, U., Prasetyo, A., Faisal, M., Siregar, J., et al. (2024). From traditional to innovation: Large language models in fisheries data extraction. In *2024 International Conference on Information Technology Research and Innovation (ICITRI)*, pages 305–310. IEEE.

Poelen, J. H., Simons, J. D., and Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, 24:148–159.

Sahu, N., Baburajan, R., and Saravanane, N. (2024). Taxobot: A chatbot for the taxonomy of marine organisms. *Available at SSRN 4946942*.

Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuemmerlen, M., Langhans, S. D., Martínez-López, J., and Vermeiren, P. (2019). How to make ecological models useful for environmental management. *Ecological Modelling*, 411:108784.

Spillias, S., Ollerhead, K., Andreotta, M., Annand-Jones, R., Boschetti, F., Duggan, J., Karcher, D., Paris, C., Shellock, R., and Trebilco, R. (2024a). Evaluating generative ai to extract qualitative data from peer-reviewed documents. *Environmental Evidence*.

Spillias, S., Trebilco, R., Adams, M. P., Boschetti, F., Constable, A., Dunstan, P., Ferrier, S., Porobic, J., Grimberg, E., Grigg, N., et al. (2024b). The future of artificial intelligence in ecosystem modelling. *SSRN preprint*.

Spillias, S., Tuohy, P., Andreotta, M., Annand-Jones, R., Boschetti, F., Cvitanovic, C., Duggan, J., Fulton, E. A., Karcher, D. B., Paris, C., et al. (2024c). Human-ai collaboration to identify literature for evidence synthesis. *Cell Reports Sustainability*, 1(7).

Tanhua, T., Pouliquen, S., Hausman, J., O'brien, K., Bricher, P., De Bruin, T., Buck, J. J., Burger, E. F., Carval, T., Casey, K. S., et al. (2019). Ocean fair data services. *Frontiers in Marine Science*, 6:440.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13:1–15.

Villasante, S., Arreguín-Sánchez, F., Heymans, J. J., Libralato, S., Piroddi, C., Christensen, V., and Coll, M. (2016). Modelling complex systems of multiple species for ecosystem based management. *Ecological Modelling*, 326:68–76.

Weiskopf, S. R., Harmáčková, Z. V., Johnson, C. G., Londoño-Murcia, M. C., Miller, B. W., Myers, B. J., Pereira, L., Arce-Plata, M. I., Blanchard, J. L., Ferrier, S., et al. (2022). Increasing the uptake of ecological model results in policy decisions to improve biodiversity outcomes. *Environmental Modelling & Software*, 149:105318.

Zheng, Z., Zhang, J., Vu, T. A., Diao, S., Tim, Y. H. W., and Yeung, S. K. (2023). MarineGPT: Unlocking secrets of ocean to the public. *arXiv preprint arXiv:2310.13596*.

Supplementary Material

S1. Data Harvesting Implementation

Our data harvesting system employs DuckDB for efficient querying of PARQUET files, enabling complex joins and aggregations without full memory loading. For species matching across databases, we use structured SQL queries that join on concatenated genus and species names:

```
SELECT
    SpecCode, PreySpecCode, AlphaCode,
    Foodgroup, Foodname, PreyStage, PredatorStage, FoodI, FoodII, FoodIII
    Commoness, CommonessII, PreyTroph, PreySeTroph
FROM sealifebase_df
WHERE SpecCode IN ({',’.join(map(str, valid_codes))})
AND (PreyStage LIKE ’%adult%’ OR PreyStage LIKE ’%juv%’)
AND (PredatorStage LIKE ’%adult%’ OR PredatorStage LIKE ’%juv%’)
```

When combining interaction data from GLOBI with diet information, we implement a comprehensive interaction mapping system that creates bidirectional records:

```
interaction_data[source_group]['preys_on'][target_group] = count
interaction_data[target_group]['is_preyed_on_by'][source_group] = count
```

Our data cleaning protocol standardises types by converting numerical values to consistent formats and timestamps to ISO format. We handle null values by removing empty values, ‘NA’ strings, and null entries while preserving data structure. Source tracking maintains database origin information for all data points.

The system implements file locking mechanisms for concurrent access, with separate locks for species data and interaction networks. We use exponential backoff retry logic for API interactions, with configurable parameters including maximum retries (5), initial delay (1 second), and maximum delay (60 seconds).

The completion check system verifies the presence of required fields including:

- Complete taxonomic hierarchy
- Species-specific database records (when available)
- Interaction data
- Source attribution
- Data quality indicators

The final JSON output maintains a consistent structure across all species entries, facilitating automated processing in subsequent framework stages.

S2. Diet Matrix Analysis

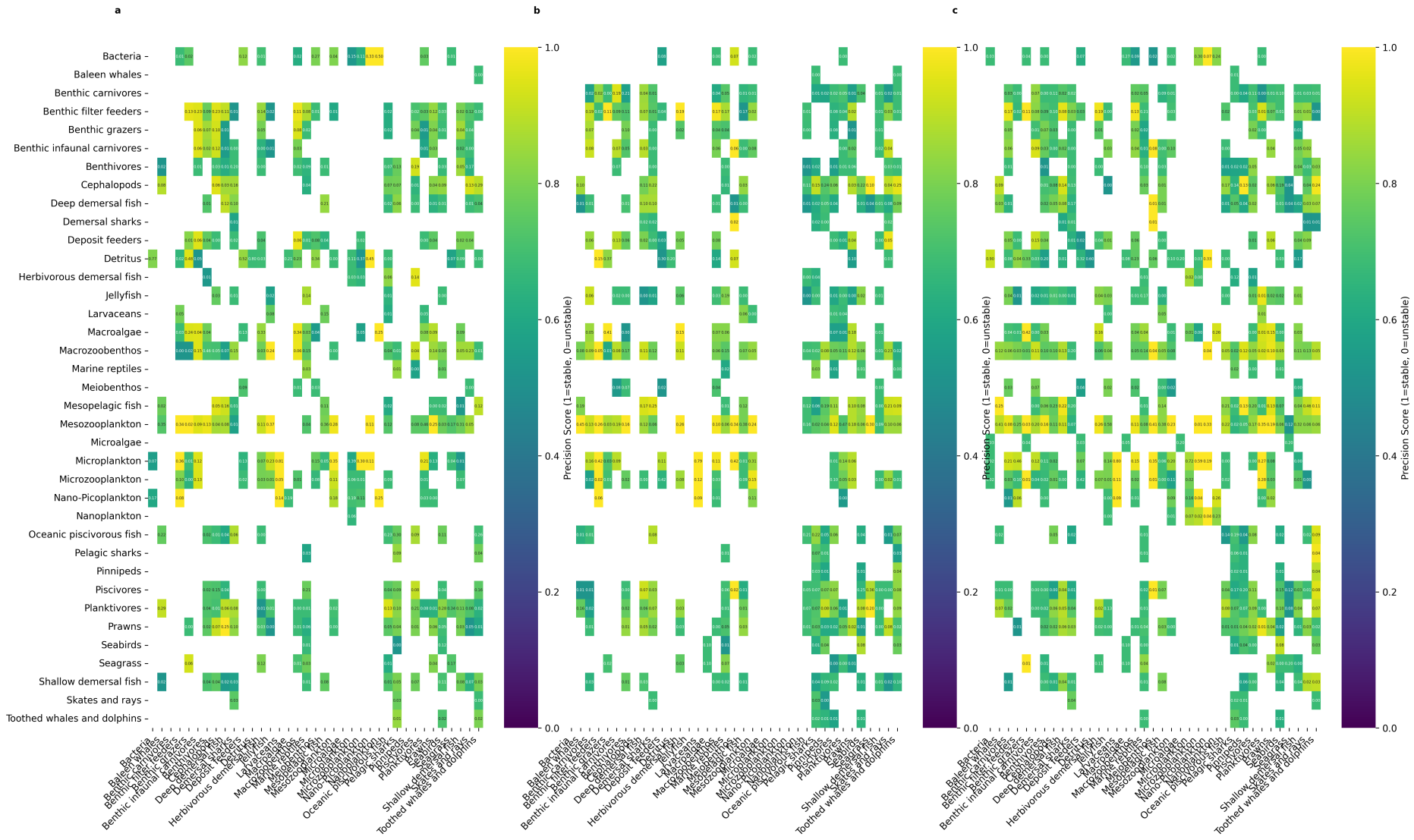

Figure S1: Detailed diet matrix consistency across five iterations for each geographic region. Column names represent predator groups and row names represent their prey groups. Numbers in each cell indicate the mean diet proportions across five iterations, while cell colors indicate the stability score (0-1, where 0 represents perfect stability and 1 represents maximum variation). White cells represent absent feeding relationships. This comprehensive visualization complements the stability score distributions and predator-specific analyses presented in the main text (Figures 4 and 5).

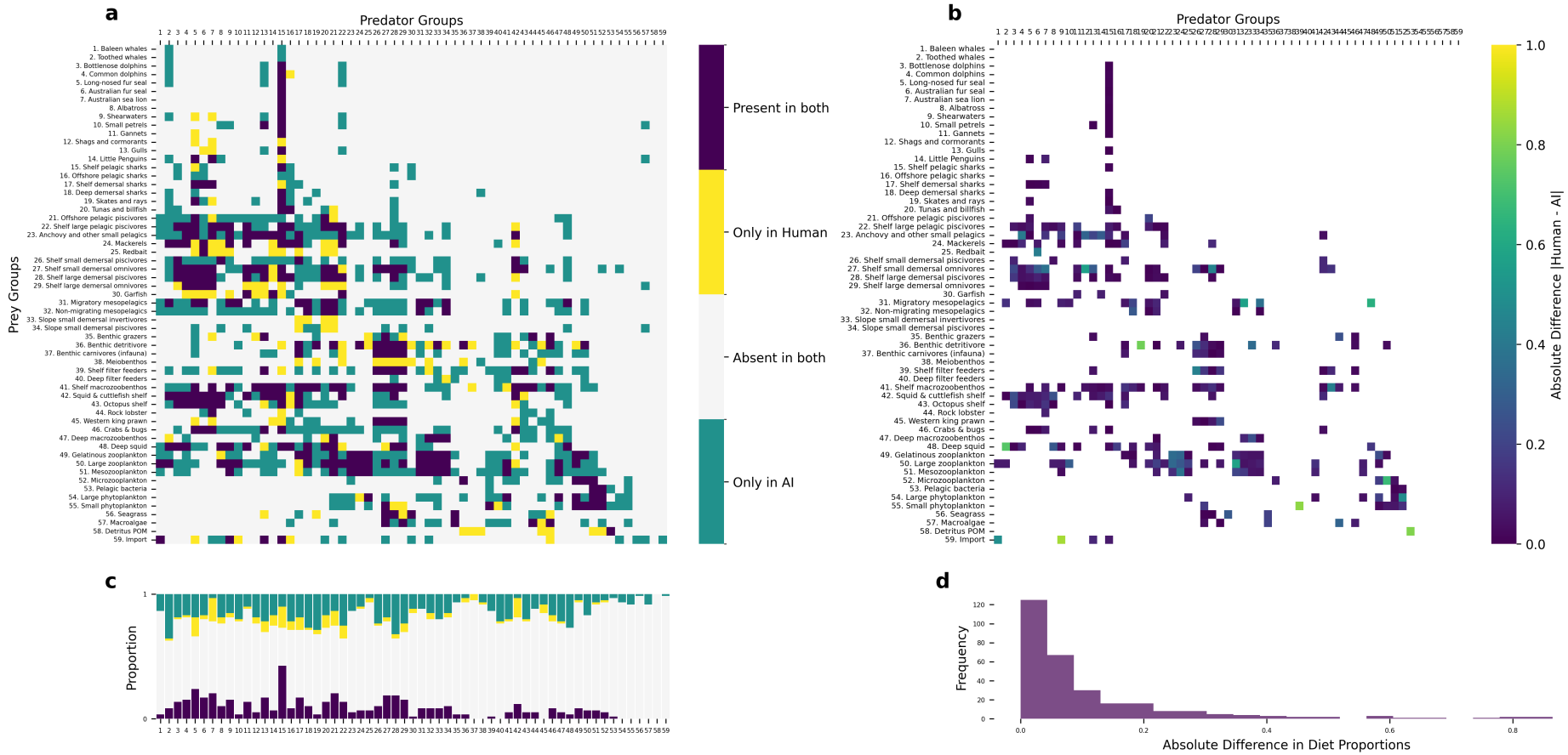

Figure S2: Detailed comparison of diet matrix elements between expert-derived and AI-generated matrices for the Great Australian Bight ecosystem. Panel (a) shows the complete diet matrix with color-coded interaction types: dark purple indicates interactions present in both matrices, yellow shows expert-only interactions, teal shows AI-only interactions, and light grey indicates absence in both. Panel (b) displays the absolute differences in diet proportions between expert and AI matrices where interactions are present in both, with colors ranging from purple (small differences) to yellow (large differences). Panel (c) shows the proportional breakdown of interaction types for each predator group, while panel (d) presents the frequency distribution of absolute differences in diet proportions. This comprehensive visualization expands on Figure 6 from the main text by providing a detailed view of each predator-prey relationship and quantifying the differences between expert and AI assessments.

S3. Group Stability Analysis

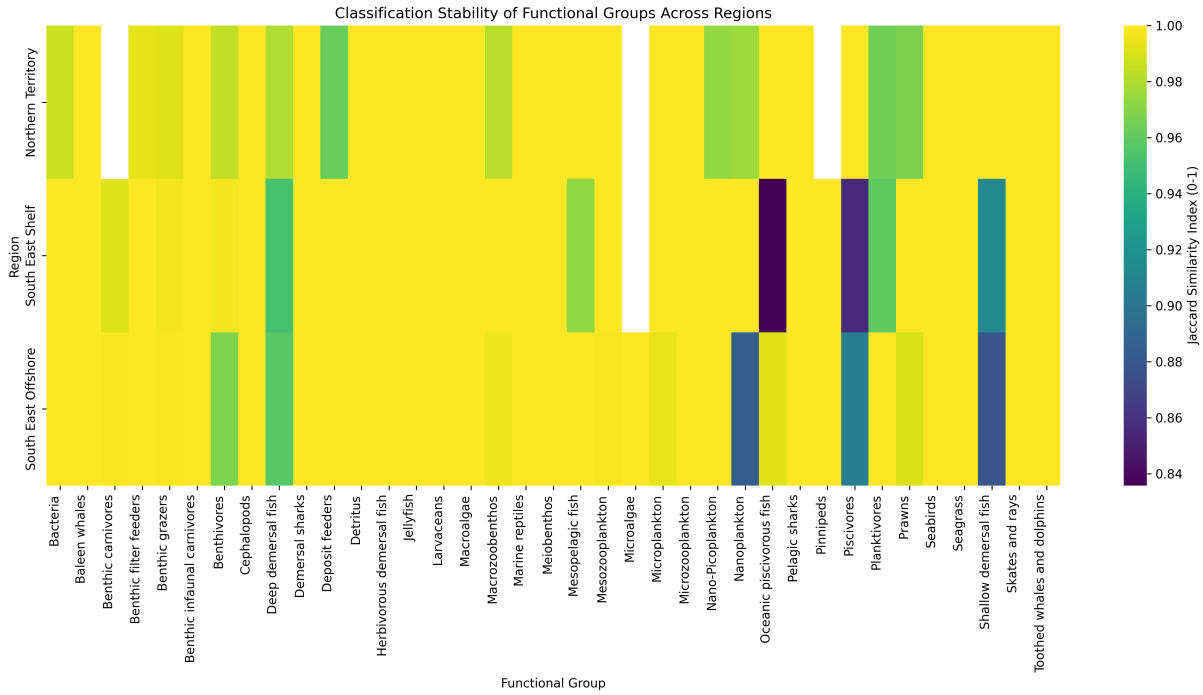

Figure S3: Heatmap showing the stability of functional group classifications across regions. Each cell displays the Jaccard similarity score (ranging from 0.975 to 1.000) between consecutive framework iterations, where 1.000 indicates perfect consistency in species assignments. Yellow colors represent higher stability (scores near 1.000), while darker purple colors indicate more variable classifications (scores closer to 0.975). Most functional groups show high stability (>0.99) across all regions, with occasional variations in groups like benthic grazers and deposit feeders, particularly in the Northern Australia region. White indicates groups that were not assigned by the AI system for that region.

S4. Technical Implementation

S4.1. Default Grouping with Descriptions

Table S1 presents the complete template of potential functional groups used by the system. This template serves as a reference for group classification, though the system can create new groups or modify existing ones based on specific ecosystem characteristics.

Table S1: Complete Functional Group Template

Group Name	Description
Skates and rays	Bottom-dwelling cartilaginous fish that play a role in controlling benthic prey populations

Continued on next page

Table S1 – Continued

Group Name	Description
Nearshore and smaller seabirds	Small gulls, terns etc that feed near shore (possibly include penguins here too) - avian predators that link marine and terrestrial ecosystems
Albatrosses	Large seabirds that forage exclusively at sea, feeding on marine prey (fishes, squids, gelatinous organisms)
Skuas and giant petrels	Large predatory seabirds that feed both at sea and on land, including predation on other birds
Fish-eating pinnipeds	Marine mammals (seals, sea lions) that primarily prey on fish in coastal and pelagic ecosystems
Invertebrate-eating pinnipeds	Marine mammals (particularly Antarctic seals) that primarily feed on krill and other invertebrates
Baleen whales	Large filter-feeding marine mammals that regulate zooplankton populations and contribute to nutrient cycling
Orcas	Apex predators that uniquely prey upon other top predators including marine mammals, sharks, and large fish
Sperm whales	Deep-diving cetaceans that primarily feed on deep-water squid and fish
Small toothed whales and dolphins	Smaller cetaceans that primarily feed on fish and squid in surface and mid-waters
Sea snakes	Marine reptiles that prey primarily on fish, particularly eels and fish eggs
Crocodiles	Large predatory reptiles in coastal and estuarine waters that prey on fish, birds, and mammals
Turtles	Herbivores and omnivores that breed on land
Planktivores	Small fishes that feed on plankton, crucial in transferring energy from plankton to larger predators
Flying fish	Epipelagic fish capable of gliding above the water surface, important prey for many predators
Remoras	Fish that form commensal relationships with larger marine animals, feeding on parasites and food scraps
Large oceanic piscivorous fish	Fish-eating predators in open ocean environments, mid-sized non-migratory species (e.g. barracuda)
Tuna and Billfish	Large oceanic predatory fish, highly mobile, often dive to feed deeper into the water column
Shelf small benthivores	Small bodied fish that feed on benthic organisms, playing a key role in benthic-pelagic coupling, live in shelf waters
Shelf demersal omnivorous fish	Medium sized demersal fish that feed on invertebrates as well as smaller fish, live in shelf waters
Shelf medium demersal piscivores	Medium sized demersal fish living near the bottom in shallow waters, often important in benthic food webs, feed on other fish primarily, live in shelf waters
Shelf large piscivores	Fish-eating predatory fishes found in various marine habitats, important in controlling prey fish populations
Herbivorous demersal fish	Bottom-associated fish that primarily feed on plants, important in controlling algal growth

Continued on next page

Table S1 – Continued

Group Name	Description
Slope/deep water benthivores	Small to mid sized fish that feed on benthic organisms and live on the shelf or seamounts
Slope/deep demersal omnivorous fish	Medium sized demersal fish that feed on invertebrates as well as smaller fish, live in slope or seamount waters
Slope/deep medium demersal piscivores	Medium sized demersal fish that feed on other fish primarily, live in slope or seamount waters
Slope/deep large piscivores	Fish-eating predatory fishes found in various marine habitats in deeper water, live in slope or seamount waters
Migratory mesopelagic fish	Fish living in the mesopelagic zone, undertake diel vertical migration, important in energy transfer between depths
Non-migratory mesopelagic fish	Fish living in the mesopelagic zone, non-migratory species, important in energy transfer between depths
Reef sharks	Top predators in coral reef ecosystems, controlling fish populations and maintaining reef health
Pelagic sharks	Open-ocean predators that help regulate populations of fishes and squids
Demersal sharks	Bottom-dwelling sharks, including dogfishes, that control populations of fishes and invertebrates on and near the seafloor
Cephalopods	Intelligent mollusks like squid and octopus, important predators in many marine ecosystems
Hard corals	Reef-building colonial animals that create complex habitat structure through calcium carbonate deposition
Soft corals	Colonial animals that contribute to reef habitat complexity without building calcium carbonate structures
Sea anemones	Predatory anthozoans that can form symbiotic relationships with fish and crustaceans
Hydrothermal vent communities	Specialized organisms living around deep-sea vents, including chemosynthetic bacteria and associated fauna
Cold seep communities	Organisms adapted to methane and sulfide-rich environments on the seafloor
Deep-sea glass sponges	Filter-feeding animals that create complex deep-water habitats and are important in silicon cycling
Sea cucumbers	Deposit-feeding echinoderms important in sediment processing and bioturbation
Sea urchins	Herbivorous echinoderms that can control macroalgal abundance and affect reef structure
Crown-of-thorns starfish	Coral-eating sea stars that can significantly impact reef health during population outbreaks
Benthic filter feeders	Bottom-dwelling organisms that filter water for food, important in nutrient cycling and regulating water quality in various depths - bivalves, crinoids, sponges
Macrozoobenthos	Mobile large bottom-dwelling invertebrates in both shallow and deep waters, important in benthic food webs and bioturbation (predatory or omnivorous)

Continued on next page

Table S1 – Continued

Group Name	Description
Benthic grazers	Bottom-dwelling organisms that graze on algae and detritus, influencing benthic community structure
Prawns	Small crustaceans that are important in benthic and pelagic food webs
Meiobenthos	Tiny bottom-dwelling organisms, important in sediment processes and as food for larger animals
Deposit feeders	Animals that feed on organic matter in sediments, important in nutrient cycling
Benthic infaunal carnivores	Predatory animals living within the seafloor sediments
Sedimentary Bacteria	Microscopic organisms crucial in nutrient cycling and the microbial loop in marine ecosystems
Large carnivorous zooplankton	Fish larvae, arrow worms and other large predatory zooplankton
Antarctic krill	Key species in Antarctic food webs, particularly important as prey for whales, seals, and seabirds
Ice-associated algae	Microalgae living within and on the underside of sea ice, important primary producers in polar regions
Ice-associated fauna	Specialized invertebrates living in association with sea ice, important in polar food webs
Mesozooplankton	Medium-sized zooplankton (200 µm to 2 cm) that feed on smaller plankton and serve as food for larger animals
Microzooplankton	Tiny zooplankton (20 µm to 200 µm) that graze on phytoplankton and bacteria, forming a crucial link in the microbial food web
Pelagic tunicates	Including larvaceans, salps, and pyrosomes, important in marine snow formation and carbon cycling
Jellyfish	Predatory gelatinous species
Diatoms	Larger phytoplankton (20 µm to 200 µm), silica dependent important primary producers in marine ecosystems
Dinoflagellates	Mixotrophic species (20 µm to 200 µm) that can switch between primary production and consumption as needed
Nanoplankton	Plankton ranging from 2 µm to 20 µm in size, including small algae and protozoans
Picoplankton	Plankton ranging from 0.2 µm to 2 µm in size, including both photosynthetic and heterotrophic organisms
Microalgae (microphytobenthos)	Microscopic algae that live on the seafloor or attached to other organisms
Pelagic bacteria	Watercolumn dwelling bacteria, consume marine snow amongst other things
Seagrass	Marine flowering plants that form important coastal habitats and nursery areas
Mangroves	Salt-tolerant trees forming critical coastal nursery habitats and protecting shorelines
Salt marsh plants	Coastal vegetation adapted to periodic flooding, important in nutrient cycling and shoreline protection

Continued on next page

Table S1 – Continued

Group Name	Description
Macroalgae	Seaweeds of various sizes that provide habitat and food for many species, including both canopy and understory forms
Symbiotic zooxanthellae	Photosynthetic dinoflagellates living within coral and other marine invertebrates
Cleaner fish and shrimp	Species that remove parasites from other marine animals, important in reef health
Discards	Carrion and freshly discarded material from fisheries activities
Detritus	Labile components of natural death and waste

S4.2. Retrieval-Augmented Generation Implementation

We implement a retrieval-augmented generation system using ChromaDB for vector storage and document management. Document processing begins with LlamaParse conversion of source materials to markdown format, preserving structural elements while enabling consistent text extraction across document types. We segment documents using a token-aware chunking strategy with a 2000-token maximum size, determined through empirical testing to balance context preservation with model limitations.

Document processing follows a two-phase approach. The initial phase generates embeddings for each document chunk using Azure OpenAI's text-embedding-3-small model, storing them in ChromaDB's PersistentClient. The system maintains an indexed_files.json registry to track processed documents. The second phase handles incremental updates, identifying and processing only new content when documents are added to the source directory.

For diet composition analysis, we implement a two-stage query process. The first stage employs a simple query to retrieve relevant document chunks:

> What do [group] eat?

The system embeds this query using the same Azure OpenAI model and performs vector similarity search to identify relevant document chunks. These results combine with structured data sources including species occurrence frequencies, food category classifications, and GLOBI interaction data to form a comprehensive input for the second stage.

We implement comprehensive error handling throughout the pipeline. The system employs exponential backoff retry logic for API interactions, with configurable parameters including maximum retries (10), initial delay (1 second), and maximum delay (300 seconds). For model interactions, we utilise LlamaIndex's query engine with zero-temperature sampling to ensure deterministic responses. The system supports multiple language model backends including Claude-3 Sonnet (200k token context), GPT-4, and AWS Claude, enabling flexible deployment based on availability and performance requirements.

The system maintains separate storage contexts for different document collections through ChromaDB's collection management. This separation prevents cross-contamination

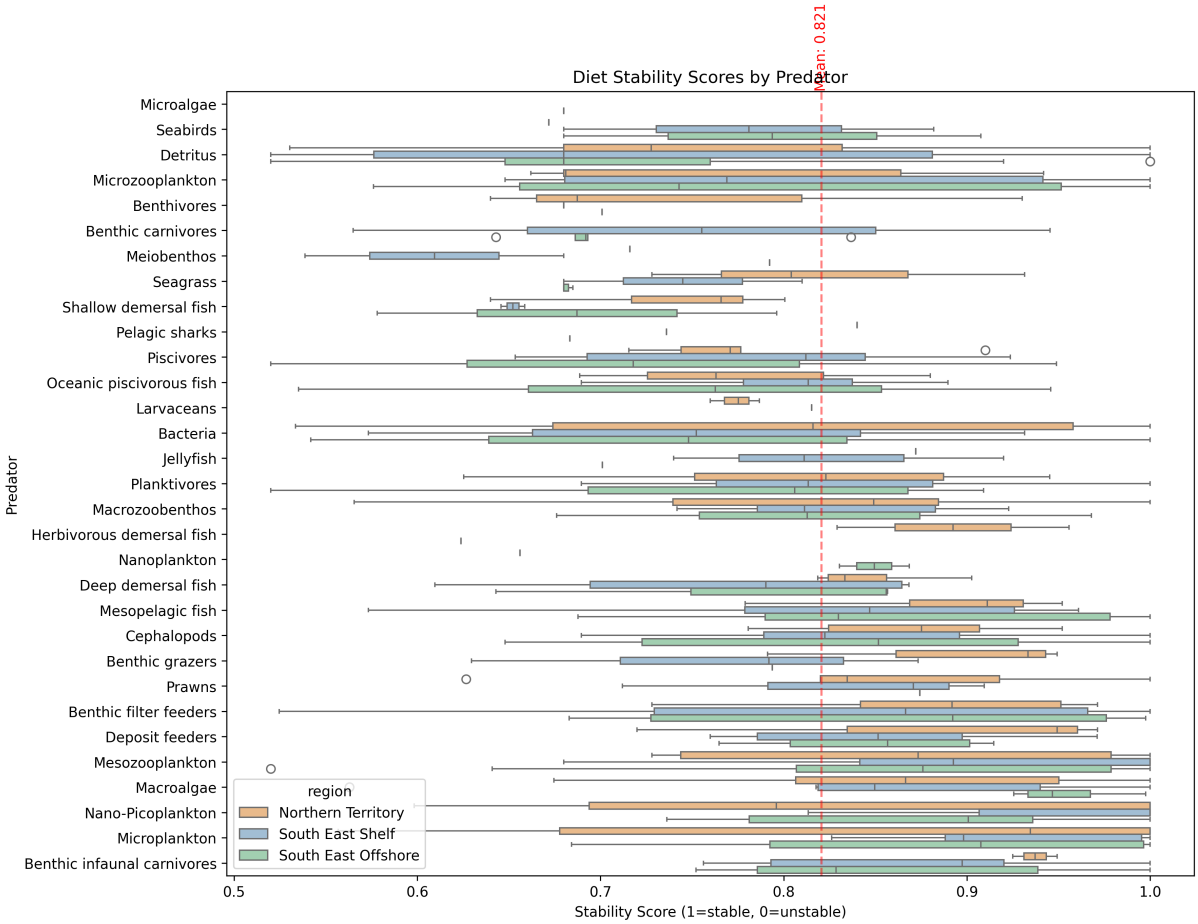

Figure S4: Diet stability scores for substantial interactions (those comprising more than 5% of a predator's diet) grouped by predator, ordered by median stability. Box plots show the distribution of stability scores for each predator's diet across regions (colored by region). The red dashed line indicates the mean stability score across all included predator-prey interactions. Higher scores indicate more consistent diet compositions across framework iterations.

between knowledge bases while enabling efficient parallel processing. We track document citations throughout the retrieval process, maintaining provenance information for all retrieved content. The complete implementation, including embedding generation, chunking algorithms, and query processing functions, is available in the project repository.