

1 **Automated Diet Matrix Construction for Marine**
2 **Ecosystem Models Using Generative AI**

3

4 **Abstract**

5 This study introduces a novel Large Language Model (LLM) driven frame-
6 work for automated species grouping and diet matrix generation in Ecopath with
7 Ecosim (EwE) ecosystem models, addressing a critical bottleneck in model devel-
8 opment. The framework (i) retrieves a marine species list from an area; (ii) uses
9 LLMs to classify them into functional groups; and (iii) synthesises trophic interac-
10 tions from diverse data sources including global biodiversity databases, species
11 interaction repositories, and unstructured user-provided text. We evaluate the
12 framework across four large Australian marine regions to assess both consistency
13 and ecological accuracy of the resulting functional groups and diet proportions.
14 The framework demonstrates high reproducibility in species grouping decisions
15 (>99.7% consistency) and diet matrix construction, with 51-59% of predator-prey
16 interactions showing consistent diet proportions across multiple runs. Validation
17 against expert-derived matrices for the Great Australian Bight ecosystem reveals
18 strong ecological alignment and accuracy, with 92.6% of taxonomic assignments
19 being at least partially correct (>75% fully correct), and correctly identifying 85% of
20 trophic interactions, while estimating diet proportions within 0.2 of expert values
21 for 80% of interactions. These findings demonstrate the framework's potential to
22 generate reproducible, ecologically meaningful components for ecosystem model
23 development while significantly reducing development time.

24 **1 Introduction**

25 Ecosystem modelling is a critical tool for understanding and managing complex envi-
26 ronments, with Ecopath with Ecosim (EwE) being a well-established framework with
27 thousands of implementations used to model marine ecosystems and predict their re-
28 sponses to external pressures (Christensen and Walters, 2004; Colléter et al., 2015). EwE
29 models provide quantitative insights into ecosystem structure and function, enabling
30 researchers to assess cumulative impacts of multiple stressors and support ecosystem-
31 based fisheries management (EBFM) decisions (Coll et al., 2015; Villasante et al., 2016;
32 Geary et al., 2020). However, constructing these models presents significant challenges,
33 particularly in constructing and parameterizing diet matrices that capture the complex
34 web of trophic interactions within an ecosystem.

35 Traditional approaches to EwE model development rely heavily on extensive lit-
36 erature review, data collation and expert knowledge, which are time-consuming and
37 resource-intensive (Holden et al., 2024a). The process of assembling diet matrices is
38 particularly challenging, requiring synthesis of diverse data sources including field

39 studies, literature reviews, and expert opinion. This creates a significant bottleneck in
40 model development, especially when applying models to new geographical contexts
41 (Holden et al., 2024b). Recent advances in artificial intelligence (AI) offer new op-
42 portunities to streamline the model development process and avoid such bottlenecks
43 (Spillias et al., 2024b). AI tools have demonstrated success in both knowledge/evidence
44 synthesis tasks (Spillias et al., 2024c; Keck et al., 2025; Castro et al., 2024; Spillias et al.,
45 2024a; Zheng et al., 2023; Nugraha et al., 2024), ecological and environmental tasks
46 (Fernandes and D'Mello, 2024; Li et al., 2024; Chen and Xu, 2024; Dorm et al., 2025;
47 Noleto Filho et al., 2024) and modelling tasks (Lapeyrolerie et al., 2022; Tuia et al.,
48 2022; Karniadakis et al., 2021), but their application to process-based ecosystem mod-
49 elling remains nascent. The key challenge lies in ensuring that AI-driven approaches
50 can effectively synthesize available information while maintaining ecological validity.

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

66 2 Methods

67 2.1 AI-Assisted Framework Overview

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

72 The first step in the process is to define a model domain and the resultant shape-
73 file is used to derive a comprehensive species list from Ocean Biodiversity Information
74 System (OBIS) (Grassle and Stocks, 1999). In Stage 2, this species list is enriched with
75 ecological data from FishBase and SeaLifeBase (Froese et al., 2010), which provide life
76 history traits, ecological parameters, and diet information, and with trophic interac-
77 tions from the Global Biotic Interactions (GLOBI) database (Poelen et al., 2014).

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

82 Stage 4 uses a retrieval augmented generation (RAG) system (e.g. (Keck et al.,
83 2025)) to synthesize the species-level ecological data into group-level summaries and

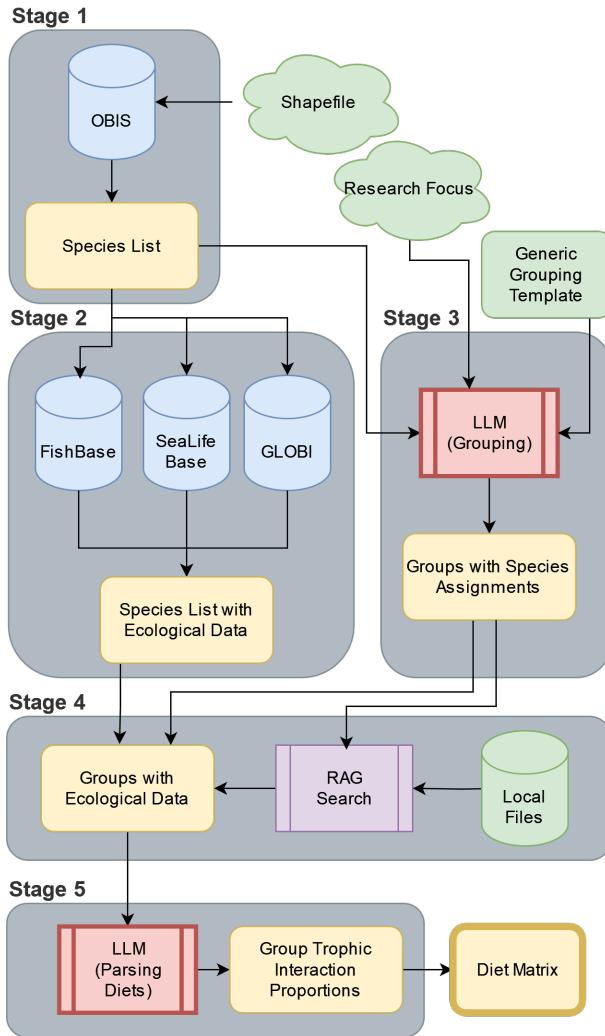


Figure 1: Overview of the AI-assisted framework for ecosystem model development. The process consists of five stages: species identification, ecological data collection, functional group organization, diet data synthesis, and diet matrix construction. Each stage integrates multiple data sources and analytical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

84 incorporates user-provided local knowledge from a vector storage database (a database
85 that stores and retrieves information based on semantic similarity rather than exact text
86 string matching), Chroma (Chroma, 2024), to produce group-level diet composition es-
87 timates. Finally, Stage 5 uses the LLM to parse and structure this combined data into a
88 diet matrix, determining trophic interaction proportions among functional groups.

89 Section S4 of the supplementary material contains detailed documentation of all
90 processing steps, including database queries, literature search criteria, and ecological
91 classification rules. The complete codebase and configuration files reside at [GitHub
92 repository URL, removed for peer-review].

93 **2.1.1 Species Identification**

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

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

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

114 The final species list is stored in a structured CSV file containing verified marine
115 species and their complete taxonomic hierarchies. The complete R implementation, in-
116 cluding rank-based filtering algorithms and geographic processing functions, is avail-
117 able in the project repository.

118 **2.1.2 Data Harvesting**

119 Following species identification, the framework gathers ecological and life history in-
120 formation for each identified species. From SeaLifeBase and FishBase (Froese et al.,
121 2010), it extracts a range of information, including habitat preferences (marine, brack-
122 ish, or freshwater), depth range distributions, maximum body lengths, and diet data.
123 These databases are accessed through their publicly available PARQUET files using
124 DuckDB for efficient querying of large datasets.

125 Our species filtering protocol implements specific constraints to align with Ecopath
126 with Ecosim modelling requirements. When processing taxonomic data, we priori-
127 tize species-level entries over genus-level classifications, only defaulting to genus-level
128 when species-specific data is unavailable. For diet composition data, we extract data

129 on food items including prey species codes ('SpecCode'), food groups, and prey stages
130 (as is found in FishBase and SeaLifeBase; see [Github repository for specific queries to
131 these databases]). We specifically filter for juvenile and/or adult life stages, as larval
132 stages are typically incorporated into planktonic functional groups rather than treated
133 as separate components of adult diets.

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

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

147 2.1.3 Species Grouping

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

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

168 For example, when classifying something like the Western Australian Dhufish (*Glau-*
169 *cosoma hebraicum*), after passing through the Kingdom Animalia, the phylum Chordata
170 is evaluated. Since Chordata includes diverse feeding strategies from filter-feeding tu-
171 niciples to predatory fish, the LLM, possessing this knowledge innately from its train-
172 ing, marks it for resolution at a finer level. At the class level, the LLM evaluates
173 Actinopterygii, which is again marked for resolution due to its diverse feeding strate-
174 gies. Continuing through the taxonomic hierarchy, the family Glaucomatidae is

175 eventually reached, where all members share similar ecological roles as demersal predators,
176 allowing classification into the demersal carnivore functional group. This hierarchical approach significantly reduces the number of required classifications, although
177 it is vulnerable to misclassifications at higher taxonomic levels if the LLM does not have
178 sufficient ecological capability. The success of this approach is highly dependent on
179 the ability of the groups in the template to properly capture the overall ecological relations
180 that are needed to model the research question. Success is also dependent on the
181 LLM's ability to understand the ecological roles of taxa and is a key target for validation
182 in this study. We provide an initial evaluation of the quality of this LLM-generated
183 grouping in Section S3.

184 At each taxonomic level, the LLM evaluates taxa against the selected grouping
185 template using the following prompt (where square brackets indicate dynamically up-
186 dated 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.

188 When the research focus indicates groups requiring higher resolution (e.g., com-

¹⁸⁹ mercial fisheries species, or a specific species of conservation concern), the following
¹⁹⁰ 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

¹⁹¹ The framework maintains complete provenance information, including the source
¹⁹² of group definitions and any AI-suggested modifications. The system automatically in-
¹⁹³ cludes a Detritus functional group to represent non-living organic matter in the ecosys-
¹⁹⁴ tem. For fisheries-related work, users could also include a discards group that is split
¹⁹⁵ off of this general 'Detritus' category. Finally, a detailed grouping report is produced
¹⁹⁶ which documents all of the classification decisions for later human review. This allows
¹⁹⁷ for a human user to quickly assess the LLM's decision-making and flag any potential
¹⁹⁸ mistakes.

¹⁹⁹ 2.1.4 Diet Matrix Construction

²⁰⁰ After the framework has assigned species to groups, the species-level diet and eco-
²⁰¹ logical information collected in Stage 2 is re-assigned to the new functional groups.
²⁰² The diet matrix construction involves two LLM-driven steps. First, the framework
²⁰³ assembles text data from various sources (RAG search results, diet data, and GLOBI
²⁰⁴ interaction data) into a structured profile for each group. This profile is passed to the
²⁰⁵ LLM to generate an initial diet composition summary. The following prompt guides
²⁰⁶ this first LLM analysis:

²⁰⁷ 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.

208

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

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

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

226 2.2 Validation

227 Our validation framework assesses both the precision (consistency across multiple
228 runs) and accuracy (comparison with expert-created matrices) of the LLM-driven model
229 construction process. This dual approach provides comprehensive insights into the
230 framework's reliability and ecological validity. We used four distinct Australian ma-

231 five regions for these assessments: three regions (Northern Australia, South East shelf,
232 and South East Offshore) to evaluate precision, and one region (Great Australian Bight)
233 to evaluate accuracy.

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

241 For the precision assessment, we calculated the Jaccard similarity coefficients be-
242 tween all possible pairs of iterations and analysed the coefficient of variation in diet
243 proportions. For the accuracy assessment, we averaged the diet proportions of the five
244 LLM-generated matrices and compared the resulting matrix with an expert-created
245 matrix for the Great Australian Bight ecosystem (C. Bulman pers. comm.) that was
246 used to inform (Fulton et al., 2018).

247 2.2.1 Study Regions

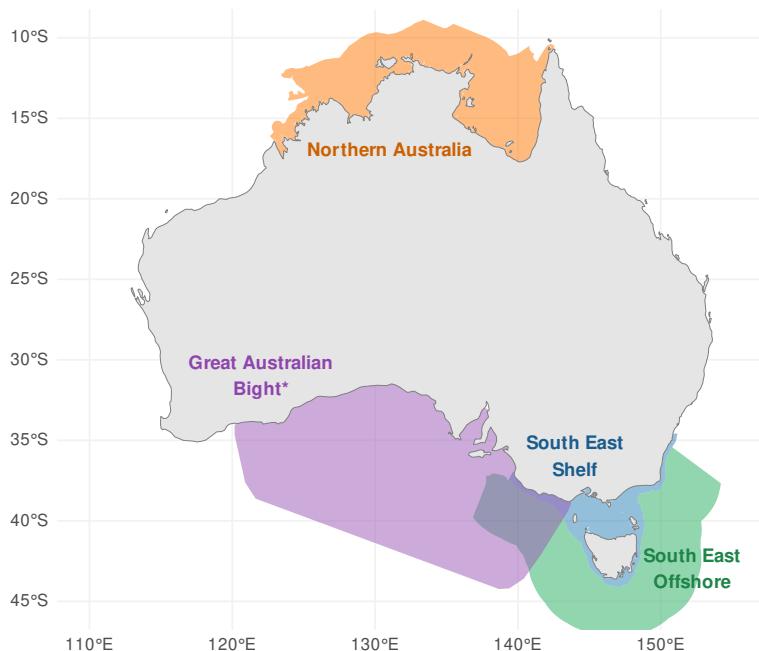


Figure 2: Map of the four study regions used to validate our approach: three of the regions were used to test the precision (consistency) of the approach: Northern Australia (orange), South East shelf (blue), South East Offshore (green). We used the fourth region, the Great Australian Bight (purple)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

248 We selected four Australian marine regions that present distinct ecological char-
249 acteristics and modelling challenges for our framework (Figure 2). For precision as-
250 sessment, 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

291 maintain a stable core of species while experiencing minor variations in peripheral
292 members, which would be captured by the Jaccard similarity but not by individual
293 species precision scores.

294 For each region, we calculated the overall stability score by averaging Jaccard sim-
295 ilarities across all possible pairs of iterations. This approach reveals how consistently
296 the framework identifies and maintains ecologically meaningful groupings across dif-
297 ferent runs. One-way ANOVA tests on these stability measures across regions, supple-
298 mented with Cohen's f effect size calculations, demonstrate the framework's precision
299 across different marine ecosystems while maintaining consistent decision-making pat-
300 terns.

301 2.2.3 Grouping Accuracy Assessment

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

312 For each taxonomic group assigned by the AI system to a functional group, we re-
313 searched the known ecological characteristics of that taxon, including feeding behav-
314 ior, habitat preferences, and trophic position. We compared these ecological charac-
315 teristics to the description of the functional group provided in the grouping template.
316 Based on this comparison, we classified each assignment as either "Correct" (the taxon
317 fits well within the functional group), "Partial" (the taxon partially fits the functional
318 group but has some characteristics that don't align), "Incorrect" (the taxon was inap-
319 propriately assigned), or "Not Sure" (insufficient information was available to make a
320 determination).

321 2.2.4 Diet Matrix Precision Assessment

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

- 326 1. Presence ratio across iterations:

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

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

330 2. Mean diet proportion:

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

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

333 3. Stability score:

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

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

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

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

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

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

343 We chose this stability metric over traditional variance measures for several rea-
344 sons. First, by normalizing deviations by the maximum value, the metric achieves
345 scale independence, allowing meaningful comparisons between interactions of differ-
346 ent magnitudes. For example, the sequences [0.2, 0.2, 0.2, 0.2, 0.1] and [0.02, 0.02, 0.02,
347 0.02, 0.01] would yield the same stability score despite having different absolute vari-
348 ances. Second, the bounded range between 0 and 1 provides an intuitive scale for
349 assessing stability, unlike the unbounded nature of variance. Third, when diet pro-
350 portions are of similar magnitude across iterations, this approach prevents minor fluc-
351 tuations in small values from disproportionately influencing the stability assessment.
352 However, in cases where iterations contain both very small and substantially larger
353 values, the scale independence property means the stability assessment will be more
354 sensitive to relative deviations in the larger values.

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

- 362 • A stable interaction ($S = 0.92$) might show values [0.02, 0.02, 0.02, 0.02, 0.01],
363 where proportions remain very similar across iterations

- 364 • An unstable interaction ($S = 0.61$) might show values [0.027, 0.25, 0.25, 0.067,
365 0.25], where proportions vary substantially between iterations, indicating inconsis-
366 tient characterization of the predator-prey relationship by roughly an order of
367 magnitude

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

377 2.2.5 Diet Matrix Accuracy Assessment

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

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

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

408 **3 Results**

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

412 **3.1 Species Grouping Reproducibility**

413 **3.1.1 Classification Consistency Analysis**

414 The framework successfully reduced ecological complexity while preserving mean-
415 ingful biological relationships. Starting with 63 potential functional groups provided
416 in the default template (See S4), it identified 34–37 region-specific groups. Chi-square
417 tests confirmed the non-random nature of these groupings, showing consistent species
418 assignments across all regions ($p < 0.001$). This statistical significance provides strong
419 evidence that the framework makes systematic grouping decisions rather than arbit-
420 rary assignments.

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

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

436 The Jaccard similarity indices reveal high overall stability in group membership
437 across all three regions (Figure 3), with most functional groups showing indices above
438 0.95. The groups labelled in the figure represent those groups with lower stability
439 indices.

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

442 **3.2 Diet Matrix Reproducibility**

443 **3.2.1 Trophic Interaction Consistency**

444 The framework identified consistent trophic relationships across all regions, with the
445 Northern Australia showing 358 interactions (58.7% having stability scores > 0.7), South
446 East shelf 380 interactions (51.3% of which were stable), and South East Offshore 477
447 interactions (56.0% stable). As shown in Figure 4, the distribution of stability scores
448 across regions demonstrates that most interactions cluster above the 0.7 threshold,
449 with a substantial proportion achieving near-perfect stability (scores approaching 1.0).

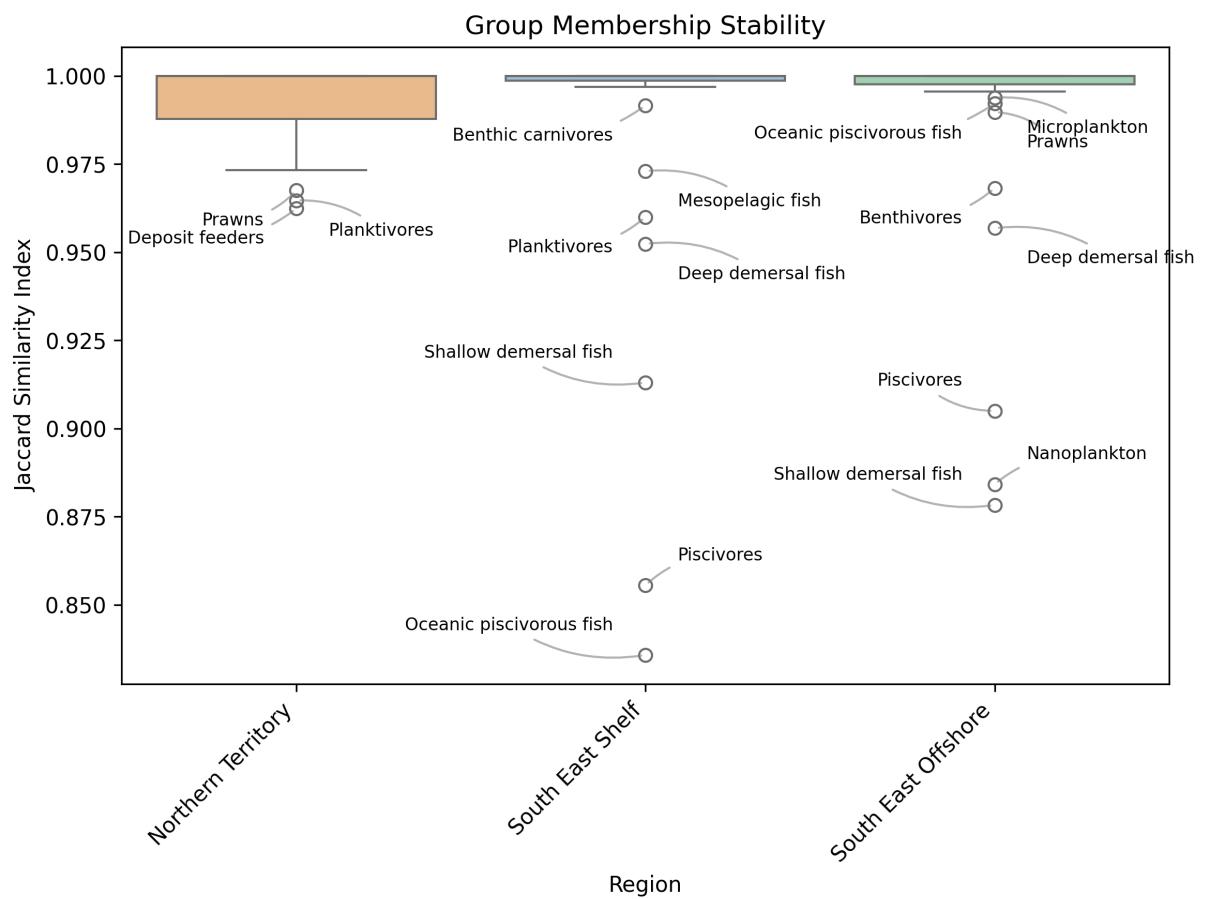


Figure 3: Group membership stability across three regions measured by Jaccard similarity index (0.85-1.0). Most groups show high stability (>0.95), with labelled points out-of-distribution outliers that exhibit lower stability.

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.

450 Spearman correlations between iterations demonstrated that the relative proportions
 451 of different prey in predator diets remained fairly consistent across all regions (Northern
 452 Australia: $\rho = 0.72 - 0.89$; South East shelf: $\rho = 0.68 - 0.85$; South East Offshore:
 453 $\rho = 0.70 - 0.87$), even when absolute proportions varied. Detailed diet matrices for
 454 each region are provided in S2.

455 3.3 Grouping and Diet Proportion Accuracy Assessment: Great Aus- 456 tralian Bight Case Study

457 3.3.1 Taxonomic Grouping Accuracy

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

469 The accuracy of assignments varied considerably across functional groups (Figure
 470 5). Many functional groups showed perfect or near-perfect assignment accuracy, in-
 471 cluding all assignments for albatross, pelagic sharks, small phytoplankton, mesozoo-
 472 plankton, small petrels, and several other well-defined groups. These groups typically

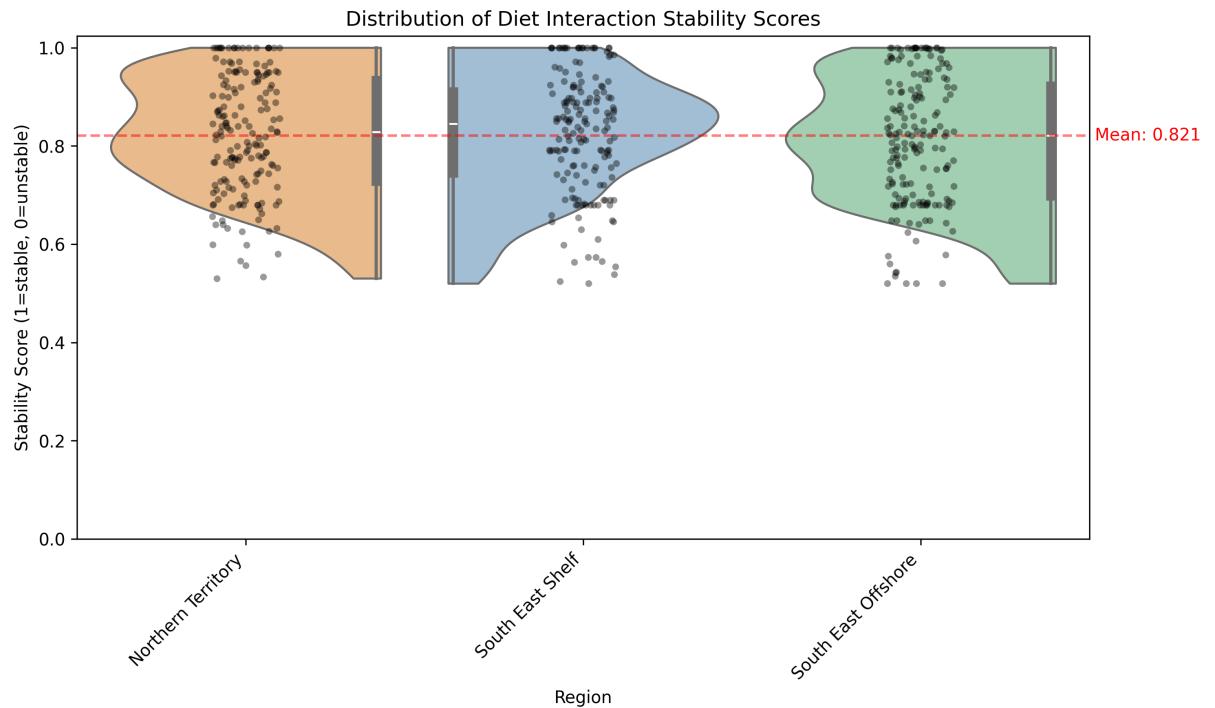


Figure 4: Distribution of diet interaction stability scores across regions for substantial interactions (those comprising more than 5% of a predator's diet). Half-violin plots show the density of stability scores (1=stable, 0=unstable), with embedded box plots indicating quartiles and median. Individual 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.

473 have clear ecological niches and distinctive characteristics that facilitate accurate clas-
474 sification.

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

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

490 3.3.2 Diet Matrix Accuracy

491 To evaluate the framework's accuracy against expert knowledge, we compared its out-
492 put to an expert-derived Ecopath model of the Great Australian Bight ecosystem (Ful-
493 ton et al., 2018). The framework demonstrated varying performance across functional
494 groups, successfully matching 59 of 76 expert-defined groups (77.6%). The framework
495 omitted 17 groups present in the expert matrix, including several commercially im-
496 portant species (Southern Bluefin Tuna, Snapper, King George whiting, and Abalone)
497 as well as Nanozooplankton. Conversely, it generated only two groups not present in
498 the expert matrix (Offshore pelagic invertivores large and Slope large demersal omni-
499 vores).

500 As shown in Figure 6a, the framework demonstrated varying levels of agreement
501 across different functional groups in identifying trophic interactions. The analysis re-
502 vealed that 8.5% of interactions were present in both matrices (dark purple), while
503 73.1% were correctly identified as absent in both (light grey). The framework uniquely
504 identified 14.5% of interactions (teal) that were not present in the expert matrix, while
505 missing 3.9% of expert-identified interactions (yellow). Overall, the framework achieved
506 an agreement rate of 81.6% with the expert matrix, with a true positive rate (sensitivity)
507 of 0.687 and a true negative rate (specificity) of 0.834.

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

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

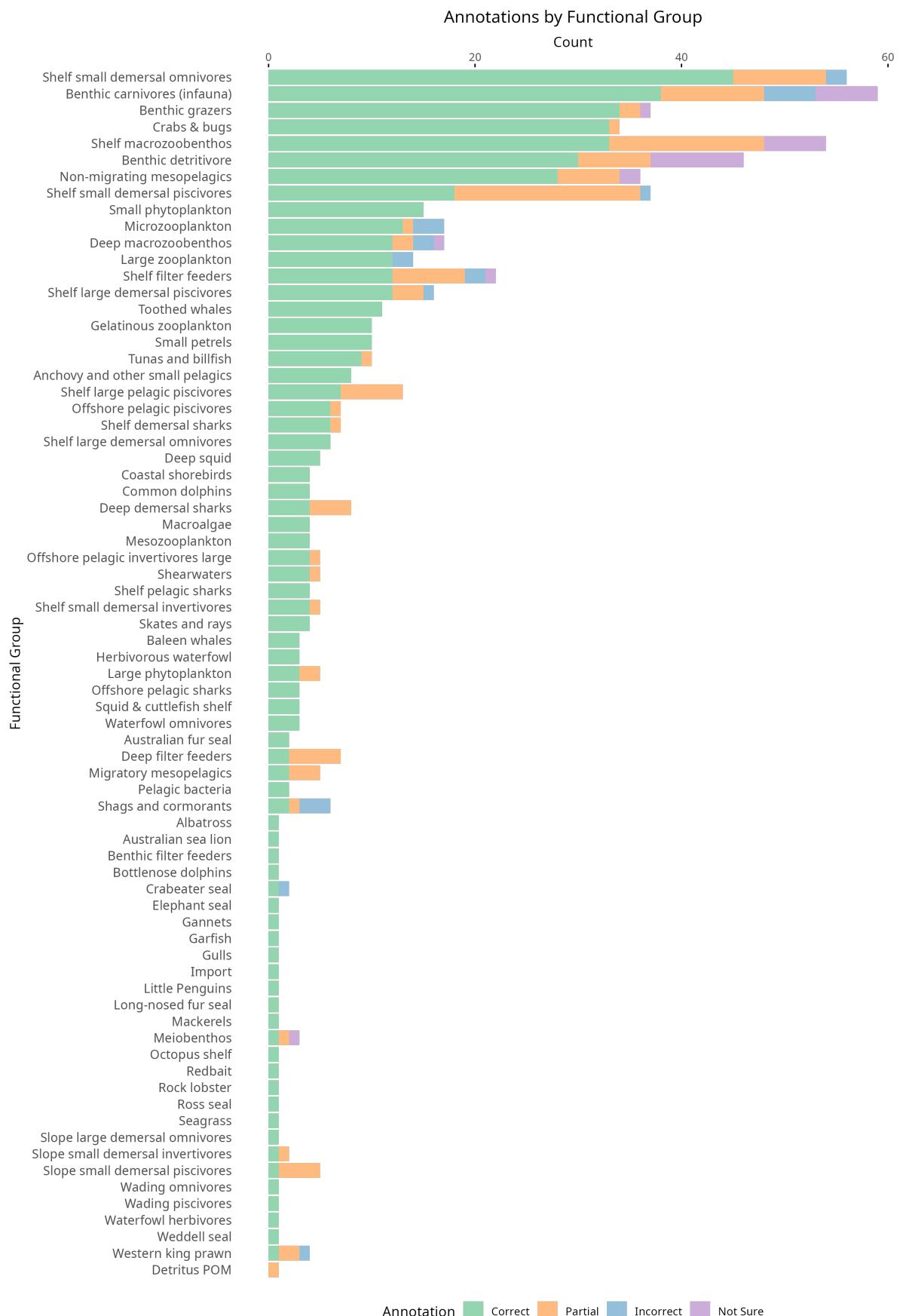


Figure 5: Accuracy of AI-assigned taxonomic groupings by functional group. The chart shows the percentage of correct (green), partial (orange), incorrect (blue), and uncertain (purple) assignments for each functional group.

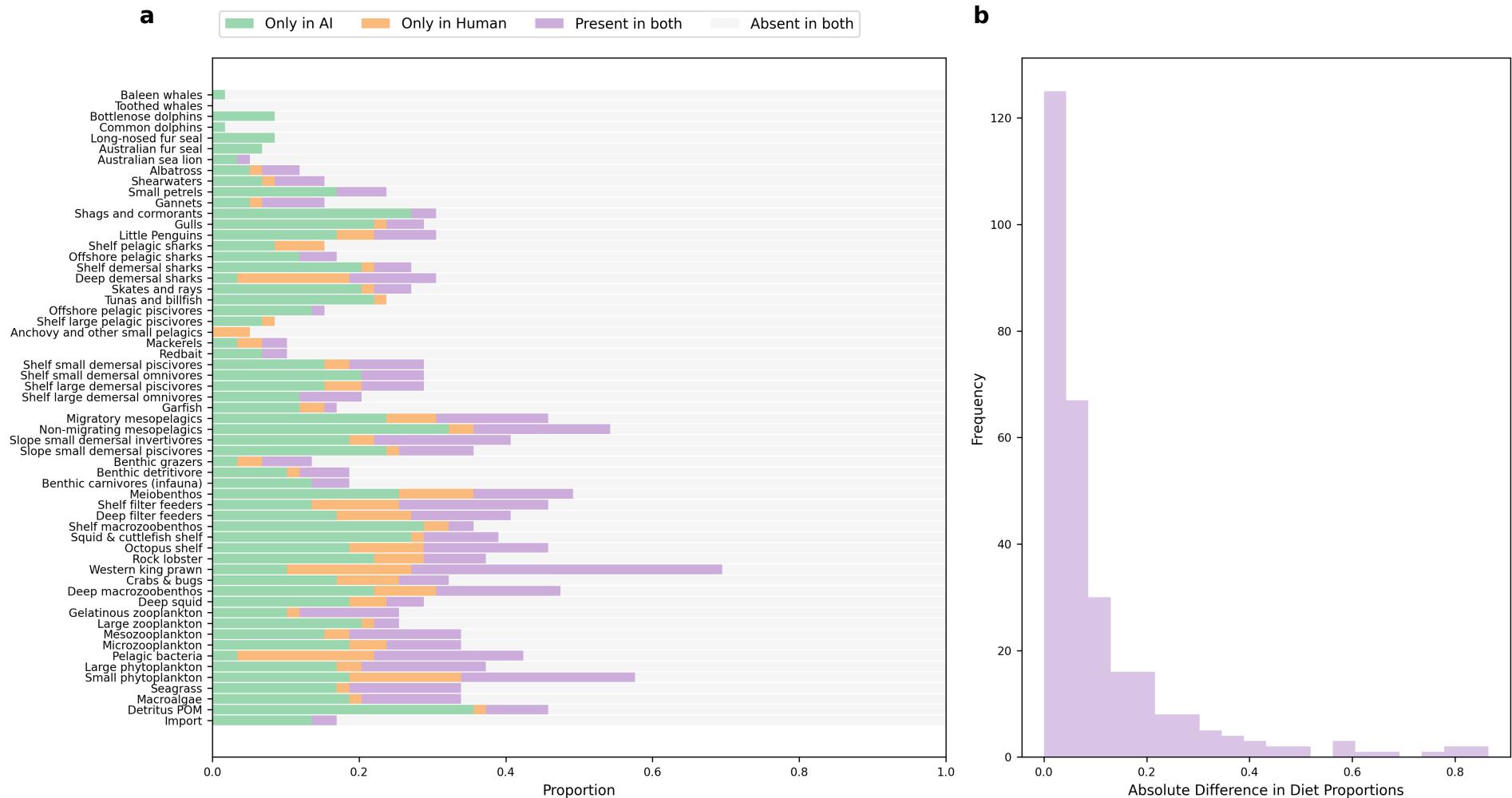


Figure 6: Comparison of expert-created and AI-generated diet matrices for the Great Australian Bight ecosystem. (a) Presence-absence patterns showing the proportion of different interaction types across functional groups. Dark purple indicates interactions present in both matrices, orange shows expert-only 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.

527 **3.4 Framework Implementation and Performance**

528 **3.4.1 Scale and Processing Efficiency**

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

534 **3.4.2 Computational Efficiency**

535 The computational requirements of the AI framework varied across regions. Total pro-
536 cessing time ranged from 2.8 to 4.8 hours across regions. The most time-intensive stage
537 was the downloading of biological data from online databases, accounting for approx-
538 imately 70% of the total processing time. Species identification typically required 0.01
539 hours, while the AI-driven species grouping process averaged 0.26 hours. Diet data
540 collection and matrix construction required 0.7 and 0.04 hours respectively, with fi-
541 nal parameter estimation taking 0.20 hours. On average, the framework required 0.7
542 seconds per species for data downloading and 0.2 seconds per species for diet data
543 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

544

545 **4 Discussion**

546 We have shown that an AI-driven framework is able to construct an important com-
547 ponent, the diet matrix, of common ecosystem models, with a fair degree of reli-
548 ability. This capability addresses a significant challenge in ecosystem modelling, as
549 the increasing use of these models for environmental management and policy deci-
550 sions requires efficient and accurate development approaches (Weiskopf et al., 2022;
551 Schuwirth et al., 2019). Constructing reliable ecosystem models traditionally involves
552 complex technical challenges, including species identification, data harvesting, and the
553 creation of accurate diet matrices—processes that are particularly demanding in fish-
554 eries management contexts where food web models inform ecological and socioeco-
555 nomic decision-making (Chakravorty et al., 2024). Our framework provides a system-
556 atic, AI-assisted solution that enhances reproducibility and dramatically reduces the
557 time investment required from months to hours. This efficiency gain aligns with the
558 growing recognition that effective ecological models must balance mechanistic under-
559 standing, appropriate spatial and temporal resolution, and uncertainty quantification

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

566 4.1 Validation Assessment

567 Here we have demonstrated the framework's potential to generate reproducible, eco-
568 logically meaningful components for ecosystem model development while significantly
569 reducing development time. It can complement the traditional approach to model
570 building and expert judgement. The framework's ability to construct reliable ecosys-
571 tem model components reveal both strengths and limitations. The framework demon-
572 strates strong internal consistency, with high stability scores (98.8-99.6%) in species
573 classifications across regions and robust correlations in predator-prey rankings ($\rho =$
574 $0.72 - 0.89$). This consistency suggests the framework makes systematic rather than
575 arbitrary decisions in constructing ecological relationships. However, these metrics
576 must be interpreted cautiously, as they reflect the framework's reproducibility rather
577 than ecological accuracy.

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

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

593 Our detailed taxonomic validation analysis further illuminates the framework's
594 ecological accuracy, with 75.3% of taxonomic assignments being fully correct and an
595 additional 17.3% being partially correct. This high overall accuracy rate (92.6% at least
596 partially correct) is consistent with other analyses (Noleto Filho et al., 2024; Dorm et al.,
597 2025) and suggests the framework generally makes ecologically sound grouping deci-
598 sions. However, the analysis also revealed that the success of the grouping may de-
599 pend on the comprehensiveness of the grouping template that the AI system is pro-
600 vided. Further, deep-sea organisms, parasitic taxa, and species with complex or vari-
601 able feeding strategies were more frequently misclassified or only partially correctly
602 assigned. These findings highlight areas where the framework's ecological knowl-
603 edge may be limited or where the predefined functional group templates may not
604 adequately capture the full range of ecological roles present in marine ecosystems. It
605 also highlights where human triage will be needed when using these methods, show-

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

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

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

637 **4.2 Applications for EBFM**

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

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

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

656 4.3 Limitations and Uncertainties

657 Our validation results highlight three key limitations of the framework. First, the
658 framework's bias toward generalised classifications, evidenced by its difficulty with
659 single-species functional groups in the GAB comparison, reflects fundamental limi-
660 tations in how the AI system processes ecological relationships. Second, the frame-
661 work's reliance on Claude 3.5 Sonnet, a closed-source large language model, intro-
662 duces scientific reproducibility challenges. While our validation demonstrates consis-
663 tent performance, we cannot fully examine the model's decision-making process or
664 potential biases. Third, practical implementation faces computational and data-related
665 constraints. Data harvesting operations proved time-intensive, and the framework's
666 performance varied with data availability across regions and ecological roles. Future
667 iterations might benefit from exploring open-source alternatives (Kommineni et al.,
668 2024) and developing more transparent decision-making processes. Additionally, fish-
669 ery and ecosystem managers need to trust that AI or hybrid approaches can reliably
670 construct and parameterize models, requiring careful attention to risk, uncertainty, and
671 transparency in model development. Decision makers often perceive ecological mod-
672 els as “black boxes” with questionable data inputs (Boschetti et al., 2018), which may
673 be amplified with AI-based approaches.

674 4.4 Future Development and Assessment

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

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

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

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

707 Fifth, future development should incorporate established best practices for eco-
708 logical model building to enhance quality and reliability. The Ecopath with Ecosim
709 approach offers methodological standards (Christensen and Walters, 2004; Heymans
710 et al., 2016) and pre-balance diagnostics (Link, 2010) that could strengthen AI-assisted
711 frameworks. More broadly, Good Modelling Practice (GMP) principles (Jakeman et al.,
712 2024) emphasize the importance of explicating modelling choices throughout the en-
713 tire modelling lifecycle to build trust in model insights within their social and political
714 contexts. Integrating these established practices would improve model assessment,
715 facilitate evaluation by management bodies, and help normalize rigorous standards
716 across the modelling community; particularly important as ecological network models
717 increasingly inform resource management decisions.

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

724 References

- 725 Anthropic (2024). Claude 3 model card. Technical documentation, Anthropic.
- 726 Boschetti, F., Hughes, M., Jones, C., and Lozano-Montes, H. (2018). On decision mak-
727 ers' perceptions of what an ecological computer model is, what it does, and its im-
728 pact on limiting model acceptance. *Sustainability*, 10(8):2767.
- 729 Castro, A., Pinto, J., Reino, L., Pipek, P., and Capinha, C. (2024). Large language models
730 overcome the challenges of unstructured text data in ecology. *Ecological Informatics*,
731 82:102742.
- 732 Chakravorty, D., Armelloni, E. N., and de la Puente, S. (2024). A systematic review on
733 the use of food web models for addressing the social and economic consequences of
734 fisheries policies and environmental change. *Frontiers in Marine Science*, 11:1489984.
- 735 Chamberlain, S. (2020). *robis: An R Client for the Ocean Biodiversity Information System*.
736 R package version 0.2.0.

- 737 Chen, M. and Xu, Z. (2024). A deep learning classification framework for research
738 methods of marine protected area management. *Journal of Environmental Management*
739 *management*.
- 740 Christensen, V. and Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities
741 and limitations. *Ecological Modelling*, 172(2-4):109–139.
- 742 Chroma (2024). Chroma - the ai-native open-source embedding database. Accessed:
743 2024.
- 744 Coll, M., Akoglu, E., Arreguín-Sánchez, F., Fulton, E. A., Gascuel, D., Heymans, J. J., Li-
745 bralato, S., Mackinson, S., Palomera, I., Piroddi, C., et al. (2015). Modelling dynamic
746 ecosystems: venturing beyond boundaries with the Ecopath approach. *Reviews in*
747 *Fish Biology and Fisheries*, 25(2):413–424.
- 748 Colléter, M., Valls, A., Guittot, J., Gascuel, D., Pauly, D., and Christensen, V. (2015).
749 Global overview of the applications of the Ecopath with Ecosim modeling approach
750 using the EcoBase models repository. *Ecological Modelling*, 302:42–53.
- 751 Dorm, F., Millard, J., Purves, D., Harfoot, M., and Mac Aodha, O. (2025). Large lan-
752 guage models possess some ecological knowledge, but how much? *bioRxiv*, pages
753 2025–02.
- 754 Fernandes, S. and D'Mello, A. (2024). Artificial intelligence in the aquaculture industry:
755 Current state, challenges and future directions. *Aquaculture*, page 742048.
- 756 Froese, R., Pauly, D., et al. (2010). FishBase.
- 757 Fulton, E., Bulman, C., Pethybridge, H., and Goldsworthy, S. (2018). Modelling the
758 Great Australian Bight ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanog-
759 raphy*, 157-158:211–235.
- 760 Geary, W. L., Bode, M., Doherty, T. S., Fulton, E. A., Nimmo, D. G., Tulloch, A. I.,
761 Tulloch, V. J., and Ritchie, E. G. (2020). A guide to ecosystem models and their envi-
762 ronmental applications. *Nature Ecology & Evolution*, 4:1459–1471.
- 763 Goldsworthy, S. D., Page, B., Rogers, P. J., Bulman, C., Wiebkin, A., McLeay, L. J.,
764 Einoder, L., Baylis, A. M., Braley, M., Caines, R., et al. (2013). Trophodynamics of
765 the eastern great australian bight ecosystem: Ecological change associated with the
766 growth of australia's largest fishery. *Ecological Modelling*, 255:38–57.
- 767 Grassle, J. F. and Stocks, K. (1999). A global ocean biogeographic information system
768 (OBIS) for the census of marine life. *Oceanography*, 12(3):12–14.
- 769 Heymans, J. J., Coll, M., Link, J. S., Mackinson, S., Steenbeek, J., Walters, C., and
770 Christensen, V. (2016). Best practice in ecopath with ecosim food-web models for
771 ecosystem-based management. *Ecological modelling*, 331:173–184.
- 772 Hill, S. L., Watters, G. M., Punt, A. E., McAllister, M. K., Quéré, C. L., and Turner, J.
773 (2007). Model uncertainty in the ecosystem approach to fisheries. *Fish and Fisheries*,
774 8(4):315–336.

- 775 Holden, M. H., Akinlotan, M. D., Binley, A. D., Cho, F. H., Helmstedt, K. J., and
776 Chadès, I. (2024a). Why shouldn't I collect more data? Reconciling disagreements
777 between intuition and value of information analyses. *Methods in Ecology and Evolution*, 15:1580–1592.
- 779 Holden, M. H., Plagányi, É. E., Fulton, E. A., Campbell, A. B., Janes, R., Lovett, R. A.,
780 Wickens, M., Adams, M. P., Botelho, L. L., Dichmont, C. M., et al. (2024b). Cost–
781 benefit analysis of ecosystem modeling to support fisheries management. *Journal of*
782 *Fish Biology*, 104:1667–1674.
- 783 Jakeman, A. J., Elsawah, S., Wang, H.-H., Hamilton, S. H., Melsen, L., and Grimm,
784 V. (2024). Towards normalizing good practice across the whole modeling cycle: its
785 instrumentation and future research topics. *Socio-Environmental Systems Modelling*,
786 6:18755–18755.
- 787 Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021).
788 Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440.
- 789 Keck, F., Broadbent, H., and Altermatt, F. (2025). Extracting massive ecological data on
790 state and interactions of species using large language models. *bioRxiv*, pages 2025–
791 01.
- 792 Kommineni, V. K., König-Ries, B., and Samuel, S. (2024). Harnessing multiple LLMs for
793 information retrieval: A case study on deep learning methodologies in biodiversity
794 publications. *arXiv preprint arXiv:2411.09269*.
- 795 Lapeyrolerie, M., Chapman, M. S., Norman, K. E., and Boettiger, C. (2022). Deep re-
796 inforcement learning for conservation decisions. *Methods in Ecology and Evolution*,
797 13:2649–2662.
- 798 Li, H., Liu, J., Wang, Z., Luo, S., Jia, X., and Yao, H. (2024). Lite: Modeling environ-
799 mental ecosystems with multimodal large language models. *Conference on Language*
800 *Models (COLM)*.
- 801 Link, J. S. (2010). Adding rigor to ecological network models by evaluating a set of
802 pre-balance diagnostics: a plea for prebal. *Ecological modelling*, 221(12):1580–1591.
- 803 Link, J. S., Ihde, T., Harvey, C., Gaichas, S. K., Field, J., Brodziak, J., Townsend, H., and
804 Peterman, R. (2012). Dealing with uncertainty in ecosystem models: the paradox of
805 use for living marine resource management. *Progress in Oceanography*, 102:102–114.
- 806 Noleto Filho, E. M., Keppeler, F. W., Reis-Filho, J. A., Giarrizzo, T., Keenlyside, N., Car-
807 valho, A. R., Coll, M., Steenbeek, J. G., and Angelini, R. (2024). Foodwebai: Creating
808 and augmenting food webs with natural language models. *Preprint*.
- 809 Nugraha, F. S., Saputra, I., Triana, H., Wahyudi, M., Radiyah, U., Prasetyo, A., Faisal,
810 M., Siregar, J., et al. (2024). From traditional to innovation: Large language models
811 in fisheries data extraction. In *2024 International Conference on Information Technology*
812 *Research and Innovation (ICITRI)*, pages 305–310. IEEE.
- 813 Poelen, J. H., Simons, J. D., and Mungall, C. J. (2014). Global biotic interactions: An
814 open infrastructure to share and analyze species-interaction datasets. *Ecological In-*
815 *formatics*, 24:148–159.

- 816 Sahu, N., Baburajan, R., and Saravanane, N. (2024). Taxobot: A chatbot for the taxon-
817 omy of marine organisms. *Available at SSRN* 4946942.
- 818 Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D.,
819 Kuemmerlen, M., Langhans, S. D., Martínez-López, J., and Vermeiren, P. (2019).
820 How to make ecological models useful for environmental management. *Ecological
821 Modelling*, 411:108784.
- 822 Spillias, S., Ollerhead, K., Andreotta, M., Annand-Jones, R., Boschetti, F., Duggan, J.,
823 Karcher, D., Paris, C., Shellock, R., and Trebilco, R. (2024a). Evaluating generative ai
824 to extract qualitative data from peer-reviewed documents. *Environmental Evidence*.
- 825 Spillias, S., Trebilco, R., Adams, M. P., Boschetti, F., Constable, A., Dunstan, P., Fer-
826 rier, S., Porobic, J., Grimberg, E., Grigg, N., et al. (2024b). The future of artificial
827 intelligence in ecosystem modelling. *SSRN preprint*.
- 828 Spillias, S., Tuohy, P., Andreotta, M., Annand-Jones, R., Boschetti, F., Cvitanovic, C.,
829 Duggan, J., Fulton, E. A., Karcher, D. B., Paris, C., et al. (2024c). Human-ai collabora-
830 tion to identify literature for evidence synthesis. *Cell Reports Sustainability*, 1(7).
- 831 Tanhua, T., Pouliquen, S., Hausman, J., O'brien, K., Bricher, P., De Bruin, T., Buck, J. J.,
832 Burger, E. F., Carval, T., Casey, K. S., et al. (2019). Ocean fair data services. *Frontiers
833 in Marine Science*, 6:440.
- 834 Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A.,
835 Mathis, M. W., van Langevelde, F., Burghardt, T., et al. (2022). Perspectives in ma-
836 chine learning for wildlife conservation. *Nature Communications*, 13:1–15.
- 837 Villasante, S., Arreguín-Sánchez, F., Heymans, J. J., Libralato, S., Piroddi, C., Chris-
838 tensen, V., and Coll, M. (2016). Modelling complex systems of multiple species for
839 ecosystem based management. *Ecological Modelling*, 326:68–76.
- 840 Weiskopf, S. R., Harmáčková, Z. V., Johnson, C. G., Londoño-Murcia, M. C., Miller,
841 B. W., Myers, B. J., Pereira, L., Arce-Plata, M. I., Blanchard, J. L., Ferrier, S., et al.
842 (2022). Increasing the uptake of ecological model results in policy decisions to im-
843 prove biodiversity outcomes. *Environmental Modelling & Software*, 149:105318.
- 844 Zheng, Z., Zhang, J., Vu, T. A., Diao, S., Tim, Y. H. W., and Yeung, S. K. (2023).
845 MarineGPT: Unlocking secrets of ocean to the public. *arXiv preprint arXiv:2310.13596*.

846 **Supplementary Material**

847 **S1. Data Harvesting Implementation**

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

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

860 When combining interaction data from GLOBI with diet information, we imple-
861 ment a comprehensive interaction mapping system that creates bidirectional records:

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

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

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

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

- 873 • Complete taxonomic hierarchy
- 874 • Species-specific database records (when available)
- 875 • Interaction data
- 876 • Source attribution
- 877 • Data quality indicators

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

880 **S2. Diet Matrix Analysis**

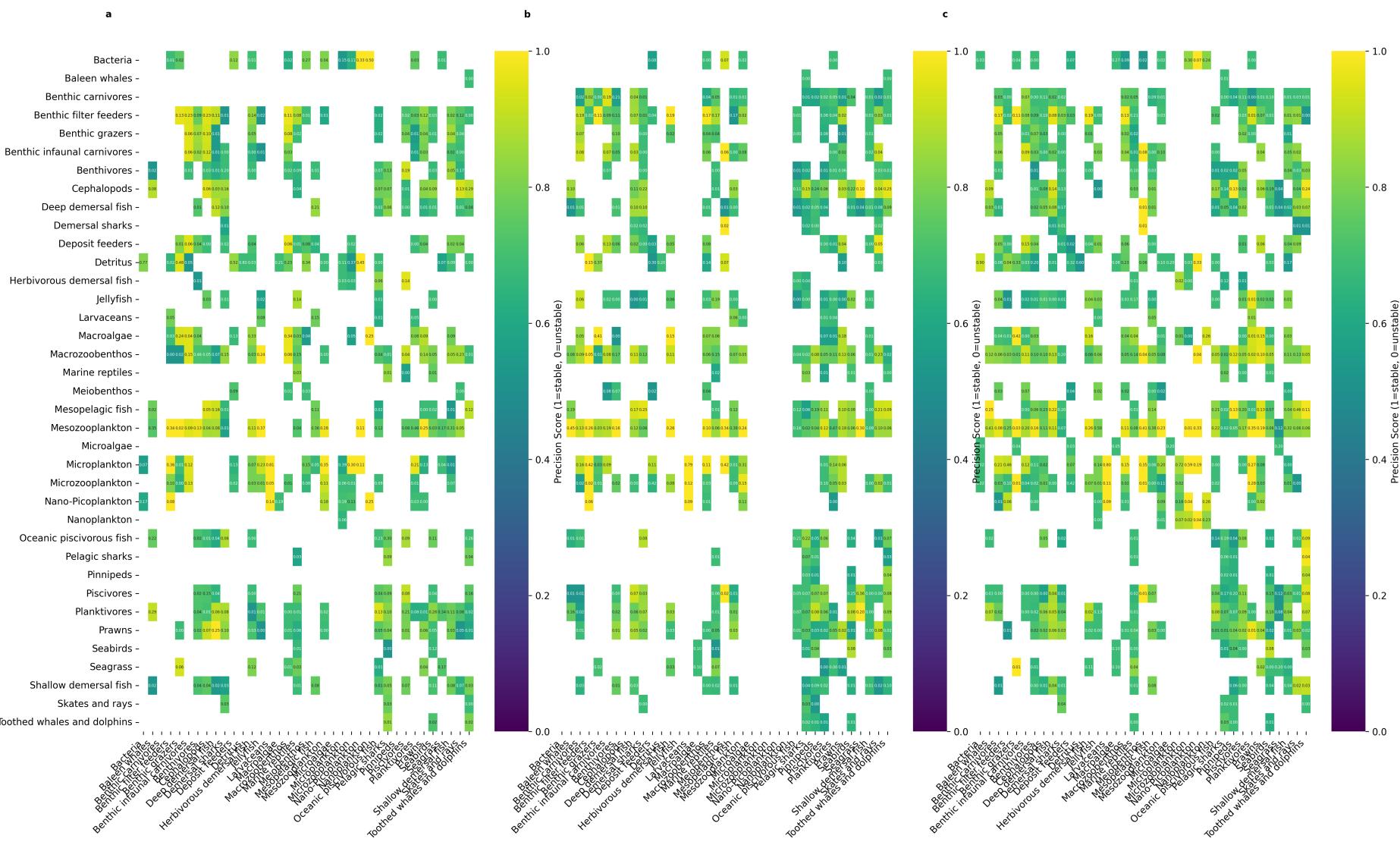


Figure S1: Detailed diet matrix consistency across five iterations for each geographic region. Column names represent predator groups and row names represent their prey groups. Numbers in each cell indicate the mean diet proportions across five iterations, while cell colors indicate the stability score (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).

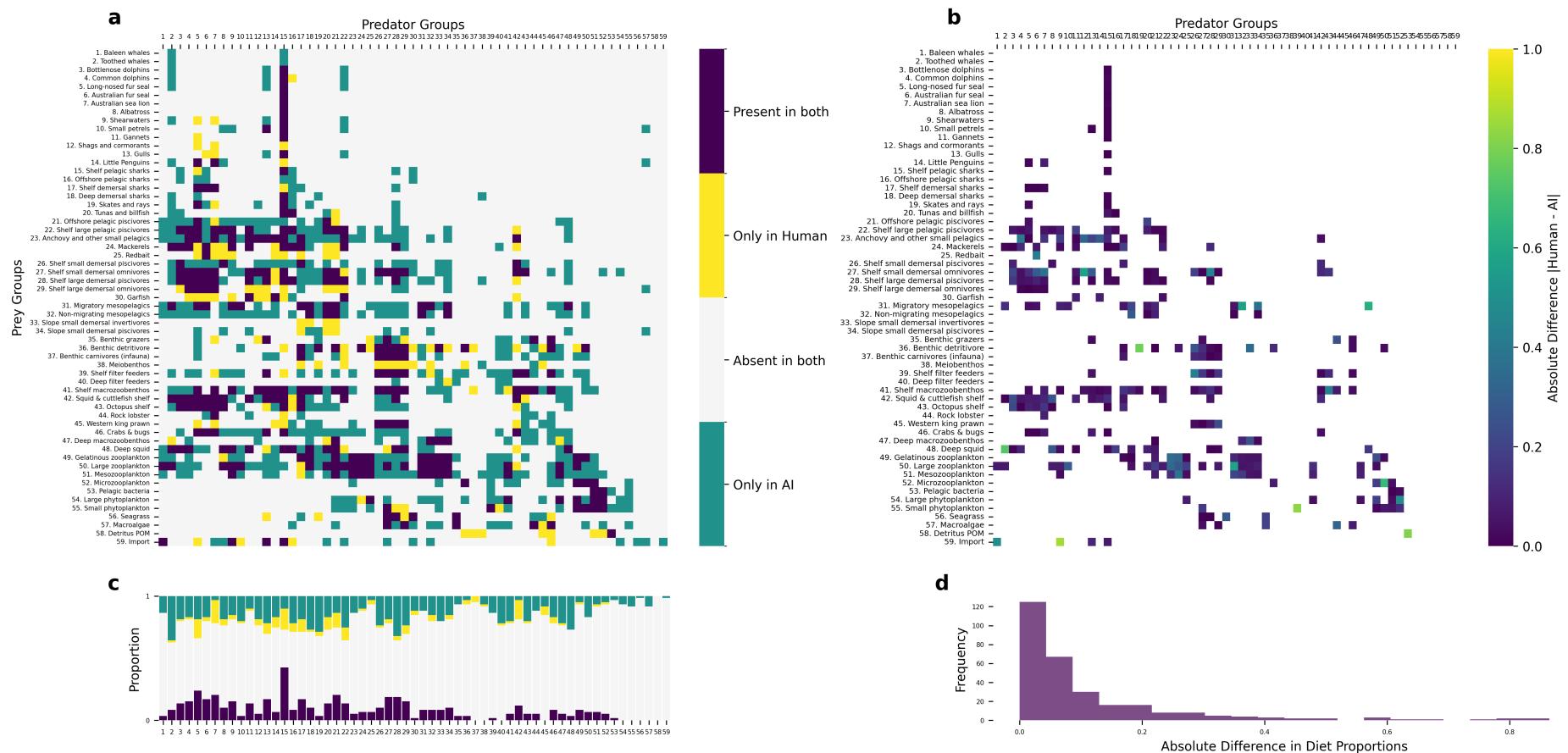


Figure S2: Detailed comparison of diet matrix elements between expert-derived and AI-generated matrices for the Great Australian Bight ecosystem. Panel (a) shows the complete diet matrix with color-coded interaction types: dark purple indicates interactions present in both matrices, yellow shows expert-only 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

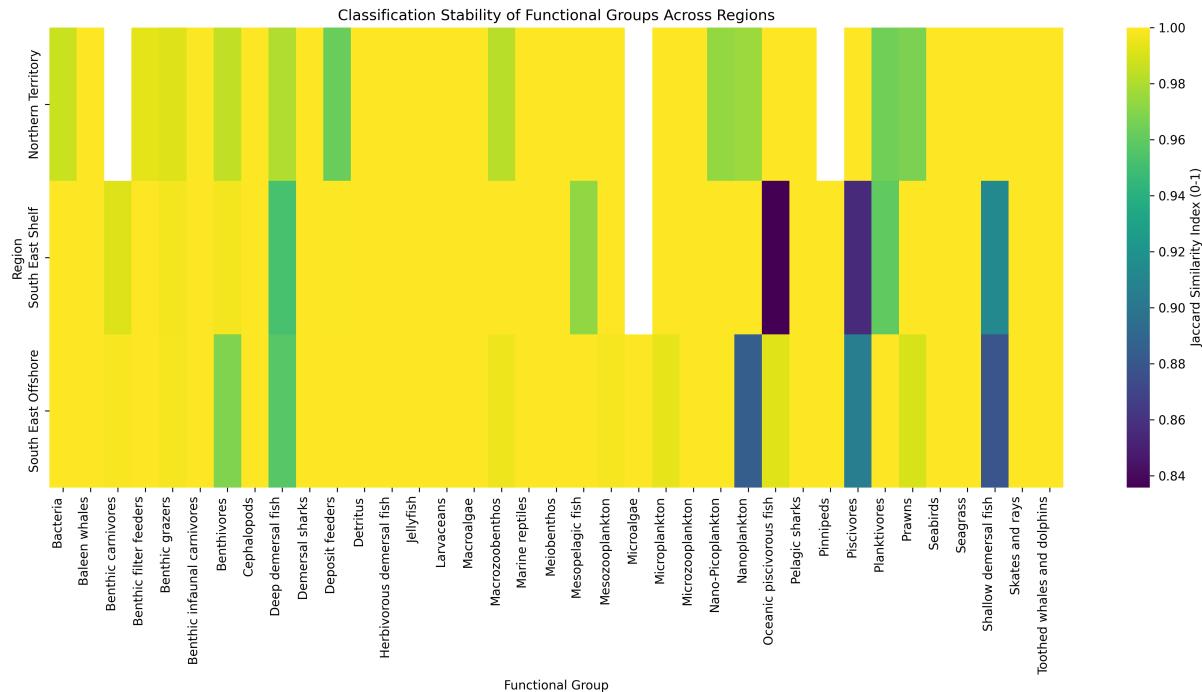


Figure S3: Heatmap showing the stability of functional group classifications across regions. Each cell displays the Jaccard similarity score (ranging from 0.975 to 1.000) between consecutive framework iterations, where 1.000 indicates perfect consistency in species assignments. Yellow colors represent 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	Carrión and freshly discarded material from fisheries activities
Detritus	Labile components of natural death and waste

887 S4.2. Retrieval-Augmented Generation Implementation

888 We implement a retrieval-augmented generation system using ChromaDB for vector
 889 storage and document management. Document processing begins with LlamaParse
 890 conversion of source materials to markdown format, preserving structural elements
 891 while enabling consistent text extraction across document types. We segment docu-
 892 ments using a token-aware chunking strategy with a 2000-token maximum size, de-
 893 termined through empirical testing to balance context preservation with model limita-
 894 tions.

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

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

What do [group] eat?

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

908 We implement comprehensive error handling throughout the pipeline. The system
 909 employs exponential backoff retry logic for API interactions, with configurable param-
 910 eters including maximum retries (10), initial delay (1 second), and maximum delay
 911 (300 seconds). For model interactions, we utilise LlamaIndex’s query engine with zero-
 912 temperature sampling to ensure deterministic responses. The system supports multi-
 913 ple language model backends including Claude-3 Sonnet (200k token context), GPT-4,
 914 and AWS Claude, enabling flexible deployment based on availability and performance
 915 requirements.

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

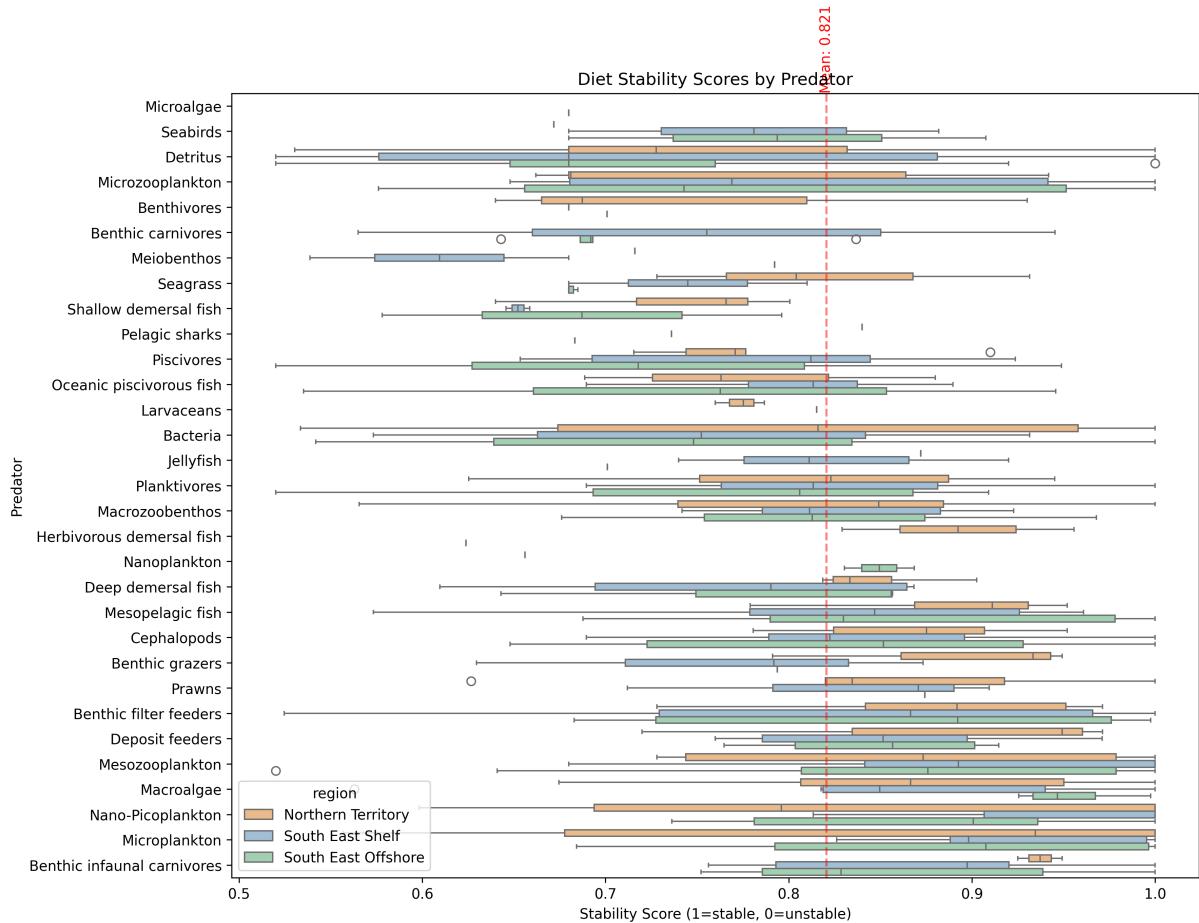


Figure S4: Diet stability scores for substantial interactions (those comprising more than 5% of a predator's diet) grouped by predator, ordered by median stability. Box plots show the distribution of stability scores for each predator's diet across regions (colored by region). The red dashed line indicates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

918 between knowledge bases while enabling efficient parallel processing. We track docu-
919 ment citations throughout the retrieval process, maintaining provenance information
920 for all retrieved content. The complete implementation, including embedding genera-
921 tion, chunking algorithms, and query processing functions, is available in the project
922 repository.