

Response to Reviewers

A survey of neurosymbolic artificial intelligence:
foundations, advances, and future trajectories

Otto Mättas

Priit Järv

Tanel Tammet

Tracking number 933-1956
Editor: Prof. Dr. Pascal Hitzler

Letter to the editor and reviewers

Dear Prof. Hitzler, dear reviewers,

We thank the editor and the three reviewers for their careful reading and constructive feedback. Please find the responses to individual comments, together with the descriptions of the changes, at the end of this letter. The major changes are:

- Restored visible section numbering and repaired roadmap, table, and worked-example cross-references.
- Clarified the goals-first four-theme lens and the system function roles.
- Sharpened novelty positioning against the closest concurrent surveys and added a source dossier for Zotero-driven bibliography refresh.
- Expanded Ethics and visual reasoning coverage with interface-coded, source-verified examples, and clarified how the main tables relate.

Numbering convention. In this document we use the *OpenReview* reviewer ordering (R1–R3).

Marked-up manuscript. All revised passages are marked in blue in the manuscript and additionally highlighted in the accompanying `nesyai2026-article-changes.pdf` produced with `latexdiff`.

Sincerely,
Otto Mättas, Priit Järv, Tanel Tammet

Reviewer 1 — review-1-20260329.txt

Recommendation: Major revision.

Source: OpenReview Review 1 (Anonymous, 29 March 2026).

Editor's letter label: Review #2.

R1.1 — Clarify “foundations” / “advances” / “future trajectories” in the conclusion

After reading I asked myself whether it is clear from the paper: What is considered as the “foundations” “advances” “future trajectories” I think the authors can emphasize this a bit more in the conclusion.

Response.

Reorganized. Section 6 now opens with a goals-first synthesis sentence, follows with four theme-by-theme recaps, and closes with a trajectory paragraph naming open engineering and evaluation problems.

R1.2 — Contribution (i) — is it the four themes + function roles + interaction patterns?

(i) Foundational anchors: Q: (Q: these are the 4 themes, the function roles, and the interaction patterns?)...My suggestion is to refer to the tables (sections) that belong to the above contributions.

Response.

Yes. The framework is three row-generating dimensions (theme, interface pattern, function role) plus three cell-level dimensions (artifact + operator, evidence tag, trade-off / limitation), defined in Section 2.4; the Contributions list (Section 1.4) now cites Tables 1, 2, and 5–9 per item.

R1.3 — Contribution (ii) — evaluation of the function roles?

(ii) Mapping methods x system functions (+benchmarks and evaluation measures): Q: thus evaluation of the functions (=perception, knowledge, reasoning, planning/control and oversight.)...My suggestion is to refer to the tables (sections) that belong to the above contributions.

Response.

Clarified. Contribution (ii) is a method-to-function mapping with evidence status surfaced per cell rather than a separate evaluation of the function roles themselves. Section 2.4 defines the function-role vocabulary; each per-theme evidence table's evidence column tags every row M / F / C / NE / A.

R1.4 — Contribution (iii) — which table supports it?

(iii) evaluate each theme (evaluation measures, benchmarks, reproducibility signals). (Which table??)...My suggestion is to refer to the tables (sections) that belong to the above contributions.

Response.

Added. Contribution (iii) is supported by the four per-theme evidence tables (Tables 5–9); the Contributions list names them explicitly.

R1.5 — Contribution (iv) — trade-offs vs pitfalls?

(iv) theme analysis of system-design pitfalls (eg. Cost vs. guarantees) Do the authors mean trade-offs instead of pitfalls? (Described via interface patterns)...My suggestion is to refer to the tables (sections) that belong to the above contributions.

Response.

Confirmed. The Contributions list now reads “theme analysis of system-design trade-offs (e.g. cost vs. guarantees), described via interface patterns” and points to the limitation columns of Tables 5–9, which carry the per-row codes (cost, guar., artifact, deploy, risk).

R1.6 — Contribution (v) — evaluation criteria for functions or themes?

(v) future directions / challenges (concrete evaluation criteria and test considerations) Do the authors mean evolution criteria for functions or themes?...My suggestion is to refer to the tables (sections) that belong to the above contributions.

Response.

Clarified. Evaluation criteria are stated per theme but anchored to functions: Section 5 now pairs each theme criterion with the function role it applies to (e.g. for Performant, measured cost / latency under workload; for Understandable, faithfulness with human protocols).

R1.7 — Cite the supporting tables/sections from each contribution

My suggestion is to refer to the tables (sections) that belong to the above contributions.

Response.

Done. The Contributions list (Section 1.4) now carries an explicit section / table pointer per item (mapping under R1.2).

R1.8 — Provenance of the five function roles

Where do those five function roles in an AI system (perception, knowledge, reasoning, planning/control and oversight) come from?

Response.

Added. Section 2.4 now opens with role-provenance: perception, reasoning, and planning are classical (Russell & Norvig; cognitive-architecture lines); knowledge is a system function in KR-centric AI work (Hogan et al.; Oltramari et al.); and oversight is added to give the ethics and reliability evidence an explicit accountability locus.

R1.9 — “Knowledge” is a strange type compared to the other roles

Two comments on those five chosen function roles: (1) Knowledge seems a strange “type” compared with the other roles

Response.

Confirmed. Knowledge is kept as a function role because in a deployed system a defined component holds the structured knowledge (KG, ontology, rule base, learned KB) and exposes it to the other roles; Section 2.4 now states this explicitly.

R1.10 — Should “explaining” be an additional function role?

(2) “explaining” would that not also be a function role?

Response.

Clarified. Explaining is treated as an outcome of perception, reasoning, planning, or oversight rather than as a separate role, because no surveyed system implements explanation as a standalone module; Section 2.4 states this choice and Section 3.2 records the resulting evidence per (interface, function) cell.

R1.11 — “Structured traces” as a reasoning operator

How do the authors consider the “structures traces” as a reasoning operator?

Response.

Removed. Structured traces are the artifact a reasoning operator emits, not the operator itself; Section 2.4 now lists operators directly (DL inference, constraint propagation, ILP, differentiable logic) and treats traces as the inspectable artifact they expose.

R1.12 — Background section — parenthetical “(In the classical sense...)”

Section Background: (In the classical sense....). Decide whether it is important what is written in brackets. If it is important then remove the brackets (include it in the main text, and make two sentences), or delete (...).

Response.

Done. The parenthetical was removed and the sentence rewritten as two declaratives in Section 1.

R1.13 — Problem statement — make the need for benchmarking measures explicit

Problem Statement: Make more explicit that there is a need for benchmarking measures.

Response.

Done. The Problem Statement now states explicitly that the neurosymbolic literature lacks shared benchmarking measures across themes and function roles; Section 2.6 specifies the M / F / C / NE / A protocol that operationalises this need.

R1.14 — Page 6 — disambiguate “AI subdomains”

Page 6: “AI subdomains” Do the authors mean “the function roles in AI systems” or do they mean a domain like medical or subdomains like NLP, vision.

Response.

Clarified. The sentence now reads “across function roles in AI systems (perception, knowledge, reasoning, planning / control, oversight) and across domains such as NLP, vision, knowledge graphs, planning, control, RL, robotics, and verification.”

R1.15 — Page 6 — explicit difference vs (Bubeck 2023, Rezazadegan 2024) and forward-reference Table 11

Page 6: What is the difference between this survey and the surveys from (Bubeck 2023, Rezazadegan 2024)? Do they have a different focus, what are the main takeaways from those surveys? (Later I saw the table 11). Might be good to add a sentence here (page 6) about the main difference.

Response.

Added. Section 1.3 now contrasts this survey paragraph-by-paragraph against the three most-overlapping recent surveys (Colelough & Regli, Michel-Deletie & Sarker, Bhuyan et al.) and forward-references Table 11, where Bubeck (2023) and Rezazadegan (2024) sit as broader comparators.

R1.16 — Page 6 — define “end-to-end implications”; themes vs functions for evaluation measures

Page 6: What are end-to-end implications? “A consistent mapping to system functions and evaluation measures”. Those evaluation measures are for the four themes or for the system functions? I would expect for the four themes.

Response.

Removed. “End-to-end implications” is gone; evaluation measures are now associated with themes (the survey question is goal-oriented), with each theme takeaway naming its dominant measures (e.g. latency / cost for Performant; faithfulness for Understandable). Section 1.5 states this.

R1.17 — Table 2 is confusing — benchmarks missing, interface patterns unclear, mismatch with Table 4

Comments wrt. Table 2: Page 7: confusing remarks about table 2: [1] (ii) table 2 is a mapping from methods to system functions (with benchmarks, evaluation measures) [2] In the text: “The summary matrix of Table 2 provides a consistent thread for mapping advances to goals, placing results within a practical system setting, and clarifying where knowledge and explanations originate.” [3] Caption table 2: survey theme, system function, interface pattern, evaluation levers. Where are the benchmarks from [1] in the table? Wrt [2]. The goals are the functions or the themes? What about the Knowledge and Explanations? Which interface patterns do you consider? This is not clear. The “typical methods (example)” are the interface patterns? (It seems they are not always match with the interaction patterns in table 4).

Response.

Reorganized. The earlier overlapping summary tables (former Tables 1, 2, 4) are consolidated into three vocabulary tables (Tables 1, 2, 3) plus four per-theme evidence tables (Tables 5–9). Benchmarks now appear as scope information attached to each evidence-table cell rather than as a separate column.

R1.18 — Page 8 — section numbers in the overview are broken

Page 8: Overview of the paper. Section numbers do not make any sense.

Response.

Fixed. Visible section numbering is restored after `\maketitle`; the Overview (Section 1.5) now uses `\ref` to live labels, and all “Section ??” references are resolved.

R1.19 — Page 9 — add the NeSy Journal to the list of relevant journals

Page 9: Source selection. The authors submit a survey to the Neurosymbolic Journal, but they do not mention the journal as “journals relevant to Neurosymbolic AI and KR”.

Response.

Added. Section 2 now names *Neurosymbolic Artificial Intelligence* alongside the other relevant venues.

R1.20 — Appendix — motivate source-dependent queries

Appendix: Can you motivate why the queries in the table are dependent of the source?

Response.

Done. The paragraph preceding Table 12 now explains that the search vocabulary is shared but each platform exposes a different boolean / quoting syntax, which is why per-source rows are useful; the table itself was converted from query-string examples to keyword sets per source, with separate filter and date columns.

R1.21 — Page 9 — which tags; predefined or derived from the papers?

Page 9: Items were tagged with primary and optional with secondary tag. Which “tags” have been used. Are they based on the papers, or did you start with a set of tags and then adapted the set based on the papers?

Response.

Clarified. Tags were derived from the papers, not predefined: the set was bootstrapped from the first batch of paper-level coding and refined as new (theme, interface pattern, function role) combinations appeared; Section 2.6 now states this provenance.

R1.22 — Critical analysis — “problem abstractions” vs “system functions”

Critical analysis: (i) Problem abstractions and integration patterns (ii) evaluation designs, datasets, evaluation measures (iii) limitations and threats to validity Q: Do the authors mean with the problem abstractions and the system functions the same?

Response.

Reworded. “Problem abstraction” and “system function” were the same concept under two names; the manuscript now uses “function role” consistently throughout (Section 2.4).

R1.23 — Is the reproducibility aspect new vs the page-8 overview?

Then they summarize again what to expect: System functions - evaluation per theme - reproducibility (including ablations for robustness) - discussion & future work. New is the reproducibility aspect? Earlier (overview of the paper page 8), they mentioned: Theme organization: Problem framing, representative advances, Evaluation/benchmarks, Limitations, Takeaway.

Response.

Yes. Reproducibility is a leg of the theme structure (problem framing, advances, evaluation / benchmarks, reproducibility signals, limitations, takeaway); the Overview now lists it explicitly so it no longer appears later as a surprise.

R1.24 — Table 1 — “cost profile” is not an interface pattern; clarify dimensions

Table 1 gives the possible interface pattern of the neurosymbolic systems? I think “neuro → symbolic” is an interface pattern. However the “cost profile” does not seem to me an interface pattern. What are the dimensions in table 1? Or are those the tags?

Response.

Agreed. Table 2 now contains only the I0–I8 interface patterns plus an explicit crosswalk to Kautz’s six integration patterns in the caption; cost profile is moved out and recorded in the limitation columns of the per-theme evidence tables.

R1.25 — Citation roles (Spine, pattern exemplar, ...) — where are they used? Is Table 1 “Spine”?

Page 11,12: A number of citation roles are given: Spine, pattern exemplar, evidence citation, context/background, position/opinion. Q1: where are those citation roles used? Table 1 is of the type “Spine”? If so, add this in the caption of table 1.

Response.

Removed. The citation-role taxonomy (Spine, Pattern exemplar, Evidence, Context, Position) is dropped from the methods; after consolidation into the per-theme evidence tables every cited paper is located as an evidence row or as adjacent context, and the Methods section now describes only the evidence-tag protocol.

R1.26 — Page 12 — broken cross-reference “Table 3”

Page 12: Last sentence: “This table illustrates the protocol in Table 3” Which table?

Response.

Fixed. The dangling “Table 3” is now Section 2.5; the surrounding sentence was rewritten so the reader reaches the protocol without a forward-reference jump.

R1.27 — Page 13 — clarify the non-comparability caveat

Page 13: Can the authors clarify “it does not imply that all listed systems are directly comparable or that any dimension transfers across tasks”.

Response.

Removed. The sentence was redundant with the table caption and raised more confusion than it prevented; the cross-task comparability caveat is now stated once at Section 2.5.

R1.28 — Page 13 — resolve “this table” and “Problem setting” = functional role

Note on page 13: “this table” → table 4. “Problem setting” = functional role.

Response.

Fixed. “This table” now reads “Table 5” (or the relevant theme table) at every occurrence, and “problem setting” has been replaced throughout by “function role”.

R1.29 — Table 4 — expected function roles, not “problem settings”; explain relation of Tables 1, 2, 4

Table 4: I expected here the function roles (=perception, knowledge, reasoning, planning/control and oversight.) Where do those “problem settings” come from? I would like to suggest to explain the relation between table 1, table 2 and table 4 more explicitly.

Response.

Reorganized. The function-role vocabulary is now in Table 3 and is the second axis of every per-theme evidence table; the former Tables 1, 2, 4 are consolidated into three vocabulary tables (Tables 1, 2, 3) plus four per-theme evidence tables (Tables 5–9).

R1.30 — Page 14 — align theme heading to “Performance”

page 14: use the same name of the themes, so first section is “Performance”. (“Performant AI”: ...)

Response.

Done. The first theme heading now reads “Performance”; the four theme names (Performance, Understandable, Reliable, Ethical) are used consistently between heading and prose throughout.

R1.31 — Page 14 — refer to Table 5 in the text

Page 14: Refer in the text to Table 5.

Response.

Added. The Performance-theme opener now cites Table 5 and explains how to read it (evidenced interface-function rows, M / F / C / NE / A in the evidence column, dominant trade-off in the limitations column); the same template is used for all four themes.

R1.32 — Page 21 — “AI reasoning” vs “Neuro”

Page 21: “we review works that make AI reasoning more interpretable”. Do the authors mean “AI” or “Neuro”?

Response.

Clarified. The sentence now reads “we review works that make neurosymbolic reasoning more interpretable” (“AI reasoning” was too broad); the broader contrast with neural-only reasoning is handled in Section 4.

R1.33 — Page 27 — which running example?

Page 27: “In the running example, ...” Eh which running example?

Response.

Removed. The running-example device is gone (the manufacturing copilot was too domain-specific to anchor a cross-domain survey); concrete instantiation is now carried by the rows of the per-theme evidence tables (Section 2.5).

R1.34 — Page 29 — which running example?

Page 29: “In the running example, ...” Which running example?

Response.

Removed. Same fix as R1.33.

R1.35 — Pages 30–34 need rewriting

Page 30-34: this part needs to be improved.

Response.

Removed. The former §4 Synthesis section is deleted; its citations are absorbed under more specific themes, and the per-theme evidence tables now carry the synthesis role directly.

R1.36 — Pages 30–34 — missing storyline; long sentences ending with references

Page 30-34: What is the story line of this section, and how fits this in the previous mappings (interfacing patterns, themes, function roles)? Those pages are a bad read. They seem rather separate sentences, often long sentences, and each sentence ends with a reference.

Response.

Removed. Same fix as R1.35; the cross-theme storyline is now carried by the per-theme evidence tables (Tables 5–9), the Discussion (Section 4), and the rewritten Conclusion opener (Section 6).

R1.37 — Page 32 — broken cross-reference “section ??”

Page 32: “section ??”

Response.

Fixed. The paragraph was rewritten in interface-pattern terms and the dangling cross-reference is resolved.

R1.38 — Page 32 — copilot introduced too late

Page 32: “same running scenarios used throughout this paper: a manufacturing maintenance copilot (section 8)”?? This is the first time that copilot is mentioned.

Response.

Removed. Same fix as R1.33; the manufacturing copilot scenario and the worked-examples subsection are gone, and no forward-references to the copilot remain.

R1.39 — Page 39 — remove the “Outlook” paragraph; clarify “dependable systems”

Page 39: I would remove the “Outlook” paragraph. The authors have already a good section Future Directions and Challenges. What do the authors mean with the field can move toward systems that are dependable systems??

Response.

Removed. The “Outlook” paragraph and the vague “dependable systems” phrase are gone; Section 6 now closes with a concrete trajectory paragraph (scalable differentiable reasoning and KR backends, platform-aware co-design, meta-cognitive control, open assets) pointing forward to Section 5.

R1.40 — Suggestion — include quantitative insights from the survey

Suggestion: including some quantitative insides from your work would be interesting as well.

Response.

Added. The synthesis aggregates **313 accepted paper-level rows from 152 distinct papers into 65 evidenced (theme, interface, function) combinations**, with the per-theme breakdown 23 (Table 5) / 20 (Table 7) / 20 (Table 8) / 2 (Table 9, with 5 cited systems across the two rows). Each row carries the chronologically-sorted reference list, the M / F / C / NE / A evidence-tag mix, and the dominant trade-off codes; the row-count asymmetry (performant-and reliable-heavy, ethics-sparse) is itself a quantitative finding documented in Section 4.3.

Reviewer 2 — review-2-20260413.txt

Recommendation: Major revision. **Source:** OpenReview Review 2 (Md Kamruzzaman Sarker, 13 April 2026). **Editor’s letter label:** Review #3.

R2.1 — Novelty constrained by concurrent surveys; positioning needs stronger argument

the survey’s novelty is constrained by the fact that several recent surveys (Colelough & Regli 2025; Michel-Deletie & Sarker 2025; Bhuyan et al. 2024; DeLong et al. 2025) discusses similar themes. While the authors make reasonable differentiation arguments in Table 11, the incremental positioning relative to these concurrent works could be argued more forcefully.

Response.

Reorganized. Section 1.3 now contrasts this survey one paragraph at a time against the closest concurrent surveys (Colelough & Regli, Michel-Deletie & Sarker, Bhuyan et al., DeLong et al., Khan et al.), followed by a synthesis paragraph stating the distinct contribution as a goal-theme \times interface-pattern \times function-role synthesis with per-row evidence tags. Table 11 now serves as supporting detail.

R2.2 — Computer-vision coverage beyond VQA is sparse

The coverage of neurosymbolic approaches in computer vision beyond VQA settings (e.g., scene understanding, autonomous driving perception) appears sparse relative to their activity in the literature.

Response.

Added. Section 3.2 now names four CV-beyond-VQA sub-areas with 14 evidenced perception papers across them: (i) indoor and outdoor 3D scene understanding, (ii) autonomous-driving perception, (iii) spatio-temporal scene-graph generation and multimodal grounding, and (iv) domain-specific perception; each is grounded in (interface pattern, function role) rows of Tables 5, 7, and 8. The cross-domain scope still constrains per-paper depth, but the gap is materially narrower.

R2.3 — Bibliography — distinguish preprints from published work; confirm venues for arXiv entries

the bibliography would benefit from more explicit versioning of preprints versus published work, since several arXiv entries appear without journal/conference venue confirmation.

Response.

Done. Preprint-only entries are now explicitly marked in the bibliography; where a preprint was later published, we cite the venue and drop or annotate the preprint.

R2.4 — Document the arXiv paper selection process

More information about the selection process of the arxiv paper would be helpful.

Response.

Added. Section 2 now states the arXiv inclusion criteria explicitly: (i) cs.AI / cs.CL / cs.LG subject classification, (ii) a neurosymbolic keyword from Appendix A, and (iii) an explicit symbolic representation and operator-level coupling.

R2.5 — Section numbering inconsistent and opaque; roadmap references do not resolve

Section numbering is inconsistent and opaque. The roadmap in the Introduction references “Section 9,” “Section 14,” “Section 21,” “Section 25,” “Section 28,” “Section 30,” “Section 34,” “Section 36,” and “Section 37,” but readers of the submitted manuscript cannot verify these numbers as the visible section headers use descriptive titles without matching numbers.

Response.

Fixed. Visible section numbering is restored after `\maketitle`; the roadmap in Section 1.5 now uses `\ref` to live labels, and all “Section ??” references are resolved.

R2.6 — Novelty subsection should contrast against the 2–3 most directly competing surveys

the “Novelty” subsection, while informative, would benefit from a sharper contrast with the two or three most directly competing concurrent surveys rather than the broader field.

Response.

Done. Section 1.3 now opens with three one-paragraph contrasts (Colelough & Regli; Michel-Deletie & Sarker; Bhuyan et al.) followed by a synthesis paragraph stating the unique contribution; the broader Table 11 remains in the Discussion as supporting detail.

R2.7 — Move “Comparison with Previous Works and Theories” to the beginning

Comparison with Previous Works and Theories can be written at the beginning of the paper, rather in the last.

Response.

Partially addressed. The direct contrast with the three closest concurrent surveys is now in the Introduction (Section 1.3); the broader Comparison subsection and Table 11 stay in the Discussion (Section 4), forward-referenced from the Introduction.

R2.8 — “What we do not claim” reads as a defensive addendum; integrate caveats throughout

The Conclusion is well-structured in its four-part thematic recap, but the “What we do not claim” paragraph, while valuable, reads as a defensive addendum rather than an integrated conclusion. Incorporating these caveats more organically throughout the thematic sections would strengthen the overall argumentation.

Response.

Removed. The standalone “What we do not claim” block is gone; caveats are now integrated theme by theme via the evidence-tag and limitation columns of Tables 5–9 and the per-theme takeaways, and Section 6 closes with those takeaways plus a forward pointer to Future Directions.

R2.9 — Table 4 readability

Table 4 is not easily readable, should be adjusted for better viewing/readability.

Response.

Reorganized. The old Table 4 is replaced by the four per-theme evidence tables (Tables 5–9); each uses a `longtable` layout with fixed column widths, repeating header / footer rows, line-break-friendly slashes in compound terms, chronologically-sorted references per cell, and a caption that lists the M / F / C / NE / A and limitation-code legends in place.

R2.10 — Explicit broken reference “Section ??” and non-existent section numbers

The manuscript contains at least one explicit broken reference (“Section ??”) and a pattern of section numbers that doesn’t exist in the paper. Perhaps the sections is the page number or someting?

Response.

Fixed. Visible section numbering is restored after `\maketitle`; every “Section ??” instance is resolved by adding the missing label or rewriting the sentence to use a structural reference.

R2.11 — Ethical AI section is underdeveloped relative to the other three themes

The Ethical AI section is substantially shorter and less technically detailed than the other three themes. While ethics is acknowledged as harder to operationalize empirically, the section would benefit from additional representative systems with interface-centric coding comparable to Tables 6–8.

Response.

Reorganized. Section 3.4 now mirrors the other themes’ structure (problem framing, advances, evaluation / benchmarks, reproducibility signals, limitations, takeaway) and is grounded in Table 9, which carries 5 evidenced systems across two (interface, function) rows: I2/planning-control with *Ahmed et al. 2023* (training-time injection of normative constraints) and I8/oversight with *Elia et al. 2024* (accountability assessment for autonomous driving), *Bonfanti et al. 2025* (KG-driven oversight workflow), *Sunny & Anwer 2026* (neuro-symbolic agent with traceability), and *van Hurne et al. 2026* (ontological foundation for accountability). Table 10 extends the I8/oversight prose with three deployer-facing ethics families (executable normative artifacts, governance workflows over audit artifacts, trustworthy-AI reviews).

Reviewer 3 — review-3-20260317.txt

Recommendation: Reject and resubmit. **Source:** OpenReview Review 3 (Anonymous, 17 March 2026). **Editor’s letter label:** Review #1.

R3.1 — Scope too wide

the scope of the paper is too wide: the area of neurosymbolic AI spans such a variety of totally different topics that it is hard to follow the paper, as every of these many different research lines is mentioned only quite shortly.

Response.

Confirmed. The cross-domain scope is intentional (the survey’s contribution is a goals-first, cross-domain synthesis), but Section 2.4 introduces a single shared vocabulary (theme \times interface pattern \times function role) that places every surveyed research line at a fixed coordinate in Tables 5–9; the synthesis aggregates **313 accepted rows from 152 papers into 65 evidenced (theme, interface, function) combinations**, so recurring lines appear as multi-paper rows rather than one-cite anecdotes. Section 4.3 states the temporal scope (2020 to mid-2026) and the audit trail (Appendix A).

R3.2 — Consider Kautz-style categorization inside themes

this thematic organization leads to the fact that there are many different approaches in each theme. That makes it hard to understand regularities between the approaches. There, it would be possibly easier to use one of the standard classifications of Nesy-approaches, e.g., by Kautz or others, and do such a thematic organization inside the different categories. Then, it would have been easier to understand commonalities and differences of the approaches.

Response.

Added. Each theme’s evidence-table rows are now indexed by interface-pattern code (I0–I8) with an explicit crosswalk to Kautz’s six integration patterns in Table 2; goal-themes remain the primary axis but Kautz-style regularities are now visible inside each theme.

R3.3 — Four themes are not disjoint; justify and discriminate

the four themes are not disjoint at all: many of the approaches aim at increasing the result quality in several themes. E.g., an explainable and interpretable approach mostly also improves ethical aspects and allows for safety proofs. These four categories need a better justification (especially with a focus on why exactly these categories have been chosen), definition and especially need a better discrimination to related themes.

Response.

Agreed. The themes are now treated as a goals-first synthesis lens with first-class overlap, not as a partition. Section 2.4 defines each theme as a deployer goal-question (Performant: “utility within budget?”; Understandable: “can a human follow why?”; Reliable: “does it hold under stress?”; Ethical: “does it respect rules, fairness, accountability?”), anchored to the human-compatible-AI line (Russell, *Human Compatible*); papers carry a primary theme tag plus zero or more secondary tags.

R3.4 — Inconsistent terminology: explainability vs understandable vs interpretable

Especially the second theme is sometimes called “explainability”, in the heading it is called “understandable” and in the text “interpretable”. However, these three themes, though occur often together, are different.

Response.

Fixed. The second theme is consistently named *understandable* throughout; Section 3.2 defines it as the umbrella goal (“can a human follow why?”) and treats explainability (a generated artifact) and interpretability (mechanistic inspection of the model) as two distinct sub-strategies inside the theme.

R3.5 — Overall structural ask

either a different sorting scheme or a more detailed categorization would be needed to get a thorough overview of nesy-AI.

Response.

Added. Table 2 defines the I0–I8 interface-pattern vocabulary and its crosswalk to Kautz’s six integration patterns; each per-theme evidence table groups rows by interface pattern, so the survey can be read either by goal (theme) or by integration pattern (interface).

R3.6 — Running example — mark up and reference consistently

The idea of a running example seems to be good, however, it needs to be mentioned more often, ideally in an example environment and there should be a reference to the introduction of this example everytime it is mentioned. Otherwise, the running example is hard to understand.

Response.

Removed. The running-example device is dropped (a single example does not anchor a cross-domain survey); the concrete-instantiation role is now played by the rows of the per-theme evidence tables (Tables 5–9), each of which is itself a worked (theme, interface, function) instantiation.

R3.7 — Tables — columns too small, row boundaries unclear (Table 6)

The tables are hard to understand, as the columns are too small. Consider Table 6 as an example, there especially in the scope-column, it is not clear where one row ends and the other starts.

Response.

Reorganized. Table 6 is replaced by the four per-theme evidence tables (Tables 5–9), which use a `longtable` layout with explicit horizontal rules between rows for clear boundaries; the M / F / C / NE / A and limitation-code legends are listed in each caption.

R3.8 — Redundant “Tag” column (always “measured”)

The “Tag” on the other hand is always “measured” and thus does not need to be added there.

Response.

Partially addressed. The single-value Tag column was removed; in the new per-theme evidence tables (Tables 5–9) the evidence column is non-trivial (every row carries a mix of M, F, C, NE, and A tags), so it is kept as part of the methodology (Section 2.6).

R3.9 — Tables should surface commonalities, not only differences

the idea of having those tables seems promising, however, it would be helpful to work out commonalities and differences in the table better, as for now, it is hard to gain insights from it that go over the fact that these approaches are different.

Response.

Reorganized. The per-theme evidence tables are now organised by interface pattern \times function role, so each row is one recurring (interface, function) pattern that multiple papers instantiate; each theme takeaway names the dominant interface patterns explicitly.

R3.10 — Broken section references throughout

all references to sections are wrong, as there are no section numbers (see, e.g., the last line of the caption of table 6 but also throughout the paper).

Response.

Fixed. Visible section numbering is restored after `\maketitle`; every “Section ??” instance is resolved by adding the missing label or rewriting the sentence to use a structural reference, and the Table 5 caption no longer carries a dangling reference.