Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Breaking the Evidence Bottleneck: A Hybrid Artificial Intelligence and Human Pipeline to Map the over Thirty-Five Years Immunomodulatory Legacy of Lactobacillus rhamnosus GG

Melissa Rocca
University of Milano-Bicocca
Daniele Noviello
University of Milan
Chiara Amoroso
Fondazione IRCCS Cà Granda
Maurizio Vecchi
Fondazione IRCCS Cà Granda
Flavio Caprioli
Fondazione IRCCS Cà Granda
Federica Facciotti
Federica.facciotti@unimib.it
University of Milano-Bicocca

Research Article

Keywords: Lactobacillus rhamnosus GG, Artificial Intelligence, LLM, immunomodulation, Cytokines

Posted Date: September 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10919084/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Lactobacillus rhamnosus GG (LGG) is one of the most studied probiotic strains, with an immunological fingerprint that has evolved over three decades of research. The exponential growth of biomedical literature has created a scalability crisis exceeding human synthesis capacity.

Aim and Study Design

This study aimed to elucidate LGG's immunomodulatory role using a zero-cost, high-speed, hybrid human-AI workflow, leveraging free Large Language Models as a scalable screening tool accessible to any researcher regardless of institutional affiliation.

Methods

An operative pipeline based on four specialized prompts was developed to interrogate three LLM architectures (Gemini 3 Flash, Perplexity Academic, and Claude 4.6 Sonnet), screening 35 years of PubMed literature (1989–2026). Performance was benchmarked against a human-validated Gold Standard using Sensitivity, Specificity, and Spearman Correlation, with a "Semantic Floor" analysis investigating False Negatives.

Results

A paradigm shift emerged: historical studies (pre-2019) focused on mucosal barrier integrity and IgA production, while recent research (2019–2026) emphasizes complex molecular signaling (IL-22, Type-I Interferons, Inflammasome modulation). A consistent core fingerprint persisted, characterized by TNF-α/IL-6 suppression and IL-10/IFNg upregulation. Gemini 3 Flash showed superior sensitivity (0.93–0.97), while Perplexity Academic achieved near-perfect specificity (0.99) due to its Retrieval-Augmented Generation (RAG) architecture which mitigated hallucinations. The Spearman Heatmap revealed a correlation coefficient near zero between Gemini and Perplexity sessions, statistically validating their architectural complementarity.

Conclusions

LGG research has shifted from mucosal barrier protection to complex molecular signaling. Synergy between different LLM architectures allows researchers to act as AI-orchestrators, focusing on critical interpretation rather than mechanical extraction.

Summary

Established knowledge on this subject.

- *Lactobacillus rhamnosus* GG (LGG) is one of the most studied probiotic strains, with a well-documented immunomodulatory profile.
- Biomedical literature is growing exponentially and LLMs are emerging as tools for accelerating evidence synthesis, but their reliability and complementarity in biomedical research remain poorly characterized.

Significant and/or new findings of this study

- We validated a zero-cost, hybrid human-AI pipeline using three distinct LLM architectures to screen 35 years of PubMed literature on LGG.
- We identified a shift in LGG research, from mucosal barrier/IgA focus to a more recent interest in complex molecular signaling (IL-22, Type I IFN, inflammasome).
- A consistent core immunological fingerprint (TNF-α/IL-6 suppression, IL-10/IFN-γ upregulation) persisted across three decades.
- We defined a "Semantic Floor," including a novel "Title Trap" phenomenon, highlighting key limitations requiring human oversight in AI-assisted synthesis.

Introduction

Lactobacillus rhamnosus GG (LGG) is one of the most widely used and documented probiotic strains[1], with over thirty years of clinical history since its patent in 1989 [2]. Yet the massive accumulation of data makes it increasingly difficult to synthesize a clear immunological fingerprint from decades of cytokine and immune-signaling evidence.

Science faces an unprecedented scalability crisis: information now doubles every two days [3], and PubMed indexes over one million papers annually[4], producing information overload. This is worsened by "Publish or Perish" culture, which prioritizes publication quantity over clarity and accessibility [6], inducing professional stress, burnout, and reduced job satisfaction [5] Traditional systematic reviews are becoming an unsustainable bottleneck, as evidence is generated faster than it can be comprehended.

LGG's literature has grown considerably, mirroring microbiome research's evolution, with research questions shifting from early anti-inflammatory characterizations toward nuanced host-microbe interactions and strain-specific mechanisms. This volume and heterogeneity demand a structured, reproducible synthesis of how understanding of LGG's immunological effects has evolved. Artificial Intelligence offers a possible solution, redefining research methodologies [7]. Free, high-speed AI models can democratize research, enabling high-level synthesis without costly software or large teams, shifting the scientist's role toward critical interpretation rather than replacing it.

We therefore validated a hybrid, zero-cost human-AI pipeline mapping LGG's immunological fingerprint (1989-present), deliberately using only free, default-tier LLMs to ensure universal replicability and democratize evidence synthesis.

Materials and methods

Information sources, search and study selection

The dataset comprised 906 papers in two subsets. The recent dataset (n = 409, 2019–2026) was collected from PubMed, EMBASE, and Web of Science using the search string: (Lactobacillus rhamnosus GG OR LGG) AND (Immunomodulation OR cytokines OR interleukines). The historical dataset (n = 497) was derived from the references of Capurso's *Thirty Years of Lactobacillus rhamnosus GG*.[2], a comprehensive review of LGG literature from 1989–2019, and validated through manual cross-referencing across the same three databases to minimize missing articles. Full dataset lists appear in **Supp.** Table 1 (recent) and **Supp. Table 2** (historical).

Table 1
Description of the prompts used in the study

#	Prompt name	Purpose	Full prompt
1	**Contextual Priming Prompt (Initialization)**	Anchor the model to the immunological research domain before analysis	"Hi, I want to give you some context for the tasks that I will give you. I'm a researcher in mucosal immunology and I'm doing a systematic review assisted by AI on the effects of the probiotic and its effects on the immune system modulation, in particular on the production of anti-inflammatory cytokines."
2	**Preliminary Screening Prompt** **(Round1)**	Broad screening of references into a structured summary	"Analyze the list of references that I just gave you and generate a summary table using the following criteria: Disease Classification: Identify the therapeutic area or disease treated (e.g., IBD, Allergies, Obesity, Infections, etc.). Immune Modulation: Indicate with a YES/NO whether the article describes a mechanism of immune system modulation (e.g., APC activation, Treg induction, TLR signaling). Cytokine Measurement: Identify whether the article reports quantitative or qualitative measurements of cytokines (focusing on IL-10, TGF-beta, IL-17, IL-22). Specify which cytokines were measured and by which technique."
3	**Qualitative Analysis of Outliers (FP/FN Investigation)**	Investigate false positives/negatives via relevance scoring	"Please rate the following papers by assigning each a score from 1 to 10 based on its relevance to immunomodulation specifically induced by the Lactobacillus rhamnosus GG (LGG) strain. Use this scale: 9–10: (LGG + direct immune mechanisms); 7–8: (LGG + clinical immune outcomes); 5–6: (mixture of probiotics with LGG); 3–4: (indirect effects/microbiota); 1–2: (other strains or off-topic). Provide the result in a table: Title | Score | Brief justification."
4	**Specialised Refining Prompt (Round 2)**	High-specificity final selection with expert persona	"Act as an immunology expert and systematic literature reviewer. Evaluate the following papers by assigning each a relevance score from 1 to 10 based strictly on the study's ability to demonstrate the immunomodulatory effect of the Lactobacillus rhamnosus GG (LGG) strain. Use this scale: 9–10: (LGG + direct immune mechanisms); 7–8: (LGG + clinical immune outcomes); 5–6: (mixture of probiotics with LGG); 3–4: (indirect effects/microbiota); 1–2: (other strains or off-topic). Give only results with a quality score > 7. Provide the results in a Markdown table with these columns: ID/Title | Immunomodulation (Yes or No) | Disease Treated | Score | Markers Detected | Technique Used | Brief Rationale."

Study selection was conducted March 9–26, 2026, with a final update on April 9, 2026. Inclusion criteria required studies to: (1) be LGG-specific; (2) address an immunomodulatory effect of LGG; (3) provide

direct cytokine measurements; and (4) be a primary study (human, murine in vivo, or in vitro). Studies were excluded if related to Lactobacilli in general or other strains, lacked health-related outcomes or immunological data, or were reviews, meta-analyses, or personal narratives. Full description of prompts used in the study are reported in Table 1.

AI softwares

The study used exclusively free, default-tier configurations via web interfaces, without API access or custom system prompts beyond the domain-specific persona. Model identity was documented by name, tier, and provider at testing time (March 9–26, 2026). Three LLMs were included: Gemini 3 Flash (Google), Perplexity Academic (Perplexity AI), and Claude 4.6 Sonnet (Anthropic). A fourth model, Co-Pilot Smart (Microsoft), was excluded after preliminary assessment revealed reliance on title-level metadata alone.

Study design

The study aimed to reconstruct and update LGG's immunological fingerprint across 30 years of research, using a head-to-head comparison of LLMs. A multistep human-AI pipeline (Fig. 1 **and Supp. Fig.S1**) processed 906 papers, divided into two temporal subsets, through a broad screening phase (Prompt 1) and a high-specificity refining phase (Prompt 2), each applied independently to Gemini 3 Flash and Perplexity Academic. Claude 4.6 Sonnet was then applied as a final validation tool on a targeted subsample of challenging cases. The complete workflow with paper counts per phase is reported in Results **(**Fig. 1).

Human validation occurred at two levels: (i) all papers classified as positive by at least one model were manually reviewed against pre-specified inclusion/exclusion criteria, establishing ground truth for false positive assessment; (ii) a random sample of 50 papers per subset (n = 100 total, Excel-randomized) was drawn exclusively from the model-negative pool and manually validated to estimate the false negative rate and complete the gold standard. This gold standard was built through a two-level review: a primary reviewer with molecular biotechnology expertise assessed all records against inclusion criteria (MR), while ambiguous cases underwent consensus-based adjudication by a senior immunology investigator (CA).

Statistical analyses

Statistical analysis and LLM performance evaluation used an automated Python 3.13.9 pipeline (pandas for data manipulation, scikit-learn for classification metrics, matplotlib/seaborn for visualizations).

Model reliability (Gemini 3 Flash, Claude 4.6 Sonnet, Perplexity Academic) was assessed via: (i) Sensitivity/Recall [TP/(TP + FN)], evaluating ability to identify relevant evidence; (ii) Specificity [TN/(TN +

FP)], measuring correct exclusion of irrelevant papers; (iii) Matthews Correlation Coefficient (MCC), a more robust measure than accuracy accounting for all four confusion-matrix quadrants and performing well even with imbalanced classes; (iv) Cohen's Kappa, measuring AI-human agreement beyond chance.

For Claude 4.6 Sonnet, Kappa interpretation required contextual adjustment: its pilot validation used only a challenging subsample (n = 100) of cases misclassified by Gemini and/or Perplexity, i.e. the corpus's most ambiguous cases. Additionally, the historical subset's highly imbalanced class distribution (n = 4 positives/50) structurally destabilized Kappa estimates regardless of actual model performance.

Data availability statement

The datasets presented in this study can be found in the online Zenodo repository (https://doi.org/10.5281/zenodo.20492569).

Results

A multistep pipeline defined a system's operative workflow

The initial phase leveraged the combined dataset to build a corpus of 906 papers spanning 37 years of LGG research (Fig. 1), to our knowledge the most extensive literature collection assembled on this strain. Model selection prioritized architectural differences over benchmark performance alone, proving essential for complementary analytical coverage. Gemini 3 Flash relies on multidimensional linguistic mapping to extract domain-specific terminology and biomarker profiles [8], while Perplexity Academic validates positive hits through real-time web-based cross-referencing [9]. This complementarity enabled a more robust synthesis of LGG-immune system literature. Both models were initialized with a domain-specific persona (Contextual Priming Prompt, Prompt 1) to anchor them to the immunological domain. The subsequent binary screening tier (Prompt 2) showed that processing papers in batches of n = 20 optimized the sensitivity/specificity trade-off. Second, a two-tier human validation framework covering 41.9% of the corpus (380/906 papers; 32.2% historical, n = 170; 53.8% recent, n = 220) was required to build a reliable gold standard, confirming that AI output alone is insufficient for definitive classification. Third, a continuous quality scoring mechanism (Prompt 3, 1–10 scale) substantially improved precision over binary classification, reclassifying 40 false positives in the recent dataset. Fourth, Claude 4.6 Sonnet was employed as final arbiter on challenging edge cases where Gemini and Perplexity disagreed, resolving the corpus's most semantically ambiguous records (Table 1, Fig. 1, **Suppl Fig. S1**)

The screening phase demonstrates superior discriminative power of Gemini and model-specific performance shifts

The screening phase (Prompt 2) revealed distinct operational profiles between the two LLMs against the gold standard (Fig. 2). In the historical dataset (n = 497), Gemini 3 Flash identified 44 positives (35 TP, 1

FN), while Perplexity Academic identified 18 positives (15 TP, 18 FN), a markedly asymmetric error profile. In the recent dataset (n = 409), both models shifted classification behavior: Gemini yielded 162 positives (98 TP, 7 FN) and Perplexity 176 positives (99 TP, 14 FN) (Fig. 2A, B). In the Historic dataset, Gemini outperformed Perplexity across all metrics (F1 = 0.886, MCC = 0.887 vs. F1 = 0.588, MCC = 0.596), because of Perplexity's higher false-negative rate (FN = 18 vs. 1) despite near-perfect specificity. In the recent dataset, both models decreased in performance due to higher false positives (Gemini FP = 64; Perplexity FP = 77), although Gemini retained superior balance (F1 = 0.734, MCC = 0.627 vs. F1 = 0.685, MCC = 0.551). Notably, Perplexity's sensitivity improved in the historical dataset, at the cost of specificity (Fig. 2C, D; **Suppl. Table ST3**). ROC curve analysis[10] confirmed these trends (Fig. 2D), and the McNemar test confirmed that the difference in classification errors between models was statistically significant (p = 0.02)

Quality check determined the misclassification classes among models

A semantic analysis of misclassified records (FPs/FNs) was implemented to evaluate the screening phase's computational limits by using six error macro-classes: strain ambiguity, structural format errors, title-abstract discrepancy, excluded biomarkers, biological off-target outcomes, and contextual noise (Fig. 3A). In the historical dataset, Gemini 3 Flash generated 8 FPs, mainly linked to strain ambiguity (50%), and a single FN caused by an excluded biomarker. Perplexity showed greater semantic rigidity: 44% of its 18 FNs stemmed from title-abstract discrepancies and 33% from unmentioned biomarkers. In the recent dataset, Gemini's errors were dominated by biological off-target outcomes (45%) and structural format errors (23%), while all 7 FNs resulted from title-abstract discrepancies. Perplexity was more susceptible to strain ambiguity (53% of FPs), and 85% of its 14 FNs again arose from title-abstract discrepancies.

To mitigate these errors, a quality-score threshold calibration (Prompt 3, 1–10 scale) was introduced. In the historical dataset, Gemini reclassified 4 FPs as non-pertinent, whereas Perplexity recovered 11 FNs and reduced FPs to two residual cases. In the recent dataset, Gemini reclassified 40 FPs and recovered 5 FNs, while Perplexity reclassified 65 FPs, leaving 11 residual false positives mainly related to inflammation without direct cytokine measurements or to non-LGG probiotics (Fig. 3B and **Fig. S2**)

Model Agreement and Refining Performance Across the Two-Tier Classification Pipeline

Cohen's Kappa quantified model-human agreement across prompts and datasets (**Suppl. Table ST4**). During the screening (Prompt 2), Gemini achieved almost perfect agreement in the historical dataset (κ = 0.876) and substantial agreement in the recent dataset (κ = 0.614), whereas Perplexity showed only moderate agreement (κ = 0.568 and 0.524), consistent with its lower sensitivity and higher false-negative rate (**Suppl. Table ST3**). In the refining phase (Prompt 3), Perplexity outperformed Gemini in the

historical dataset (κ = 0.749 vs. 0.489), while both converged to moderate agreement in the recent dataset (κ = 0.460–0.548).

Prompt 4 introduced a quality-score threshold (≥ 8) to address strain ambiguity, format errors, and unmentioned biomarkers (**Suppl. Table ST5**). Specificity remained high, but sensitivity diverged: Perplexity performed best in the historical dataset (MCC = 0.763) before declining in the recent one (MCC = 0.515), whereas Gemini remained stable (MCC = 0.563 in both) (Fig. 4A-B). ROC analysis confirmed Gemini's stronger ranking ability (AUC = 0.946 historical) compared with Perplexity's more strict binary classification behavior (AUC = 0.657 historical; 0.700 recent) (Fig. 4C-D).

Internal stability and reliability of the two LLM was tested across 3 independent sessions

Internal stability of Gemini 3 Flash and Perplexity Academic was assessed across three independent sessions on separate days (**Suppl. Tables ST6-ST7**, Fig. 5). Gemini showed fair-to-moderate agreement (historical κ = 0.367; recent κ = 0.467), while Perplexity showed higher consistency (historical κ = 0.600; recent κ = 0.667) (**Suppl. Table ST6**). A two-way random-effects ICC2 model indicated that Perplexity's grading was poorly reliable across both subsets, while Gemini reached moderate reliability in the recent one (all $p < 0.01$; Fig. 5A, **Suppl. Table ST7**). Spearman correlation (ρ) confirmed this pattern (Fig. 5B-C and **Suppl. Table ST8**). Boxplot analysis (Fig. 5D) showed no outliers for Gemini, reflecting its wide score distribution, whereas Perplexity's recent scores were strongly downward-skewed, with scores above 0.5 flagged as statistical outliers.

Overall, temporal context introduced measurable stochasticity, particularly in Gemini, though this did not alter model ranking performance, suggesting however that identical queries submitted at different times to the same free-tier model may yield non-identical outputs.

Claude as a mitigation tool

To resolve systematic boundary errors and cross-platform disagreements, Claude 4.6 Sonnet was tested on a representative sub-corpus (n = 100: 50 historical, 50 recent) targeting the most ambiguous cases where Gemini and Perplexity conflicted or failed. Unlike Perplexity's sensitivity decay in recent literature (Fig. 4A,D), Claude showed consistent sensitivity across both datasets (**Suppl Fig. S5A,B; Suppl Table ST9**), achieving higher precision and specificity in the recent dataset (0.895), effectively resolving Gemini's over-inclusion bias. Its F1-score (0.642) and MCC (0.360) in the recent dataset reflect a balanced refining performance, further supported by ROC analysis (AUC = 0.688 recent vs. 0.587 historical; **Suppl Fig. S5C**). Within the study's constraints, Claude showed slight agreement in the historical(κ = 0.070) but fair agreement in the recent dataset (κ = 0.300), the latter reflecting its strong precision in correcting the primary models' over-inclusion bias on emerging literature (**Table S10)**.

Cross model failures analysis and semantic floor analysis.

In the recent dataset errors were driven mainly by abstract-full text discrepancies and biochemical complexity. Many FNs occurred because key inclusion criteria (cytokines, precise strain) appeared only in the full text ("The Full-Text Gap")[12–14]. Gemini's high exploratory sensitivity (0.933) supports its use as a screening tool, whereas more conservative models like Claude and Perplexity, relying strictly on explicit abstract evidence, often failed here. Several shared failures (IDs #6, #26, #72, #76, #128, #194) involved derived components (p75 proteins, extracellular vesicles), engineered strains, or delivery systems ("Biochemical and engineering complexities struggle"), revealing that LLMs often struggle to independently judge whether a biological derivative or modified component is pertinent (defined as “Semantic Floor”).

When titles used generic terms (e.g., "probiotic") with the specific strain named only in the abstract, models often failed to preserve this semantic link (IDs #156, #365; "The Title Trap"). Claude specifically failed on ID #42, where LGG and IL-6 appeared in title/abstract respectively, while Perplexity correctly identified it. A shared FP (#406, a non-GG-specific *L. rhamnosus* isolate)[15] and shared FN (#199) [12] confirm that Strain Ambiguity remains a key hurdle: LLMs consistently conflate *Lactobacillus rhamnosus* with the LGG strain when not explicitly differentiated, causing persistent precision noise and over-inclusion errors across all platforms.

LGG acts as an immunomodulator, upregulating regulatory and suppressing pro-inflammatory cytokines

Finally, it was mapped LGG's immunomodulatory effects extracted from the Historical (n = 47 positive papers) and Recent (n = 111 positive papers) datasets (https://doi.org/10.5281/zenodo.20492569) across three levels: its core cytokine fingerprint, the shifts in research focus over time, and evidence distribution across experimental models.

Cross-referencing both subsets identified five primary cytokines as pillars of LGG-mediated immunomodulation: TNF-α, IL-10, IL-6, IL-1β, and IFN-γ (Fig. 6A, **Suppl. Table ST11**), clustering into two functionally distinct axes consistent with their known biology. The first is a pro-inflammatory suppression axis (TNF-α, IL-6, IL-1β). TNF-α was the most frequently studied cytokine overall (n = 19 historical, n = 49 recent), with a dominant down-regulation trend (84% historical, 90% recent), confirming LGG as a consistent antagonist of systemic and localized inflammation via NF-κB and TLR signaling inhibition. Up-regulation was instead observed in cancer, viral infections, or liver injuries [16–22], likely reflecting disease-specific immune activation rather than a direct probiotic effect. IL-6, frequently co-reported with TNF-α, was reduced across 41 studies (n = 5 historical, n = 43 recent), further supporting NF-κB inhibition. IL-1β emerged predominantly in recent literature (n = 36), where its consistent down-regulation points to NLRP3 inflammasome modulation as an emerging mechanistic target [23], virtually absent in the older literature.

The second axis comprises IL-10 and IFN-γ (Fig. 6A). IL-10 was the primary mediator of immune resolution associated to LGG functions, being up-regulated across 30 studies in both datasets (n = 9

historical, n = 27 recent) and acting through Treg induction and IL-10R2 upregulation [24–27]. IFN-γ, identified in 29 studies (n = 13 historical, n = 16 recent), showed dual modulation: up-regulated in viral infections, cancer, or allergies [28–30], and down-regulated in inflammatory conditions, highlighting LGG's role not merely as an anti-inflammatory agent but as an immune-fitness enhancer adapting to the host's pathological state.

The temporal subsets reveal a clear evolution in LGG's scientific framing (Fig. 6B,C). Historical literature (1989–2019) centered on humoral protection and mucosal barrier integrity, with IgA production (n = 7) and classical pro-inflammatory markers (IL-6, TNF-α, IFN-γ) as dominant endpoints, reflecting LGG's established role in intestinal barrier defense [31,32]. Recent literature (2019–2026) introduced specialized markers. IL-22 upregulation (n = 4) points to emerging tissue-repair signaling via the AhR/ILC3 pathway [33–36], while Type I IFN upregulation documents a newly characterized antiviral axis [37,38]. Most recently, research has shifted toward bioengineered strains and postbiotic derivatives (p75 proteins, extracellular vesicles) [14,19,28,39–41]

Finally, the corpus spanned three experimental levels (Fig. 6D; **Suppl. Table ST11**): in vitro, murine/in vivo, and human clinical trials. Core fingerprint markers (TNF-α, IL-6, IL-10) were consistently represented across all three levels in both temporal subsets. Emerging markers showed instead a markedly different distribution: IL-22 upregulation (via AhR/ILC3) and Type I IFN induction (via cGAS-STING and RIG-I) were concentrated almost exclusively in murine models, with minimal representation in human trials, indicating that while the translation of mechanistic evidence for inflammasome and antiviral pathway modulation remains largely understudied. IFN-γ, despite its dual modulation pattern, is similarly underrepresented in human trials relative to animal models.

Discussion

Lactobacillus rhamnosus GG's (LGG) scientific footprint has grown considerably over three decades, paralleling the expansion of microbiome research. Our analysis shows that LGG’s role evolved from early, descriptive anti-inflammatory characterizations toward more complex host-microbe interactions and strain-specific signaling pathways. This evolution occurred against remarkable consistency: across historical (1989–2019) and recent (2020–2026) literature, LGG maintained a stable dual-action fingerprint: suppression of TNF-α and IL-6 and upregulation of IL-10, and context-dependent modulation via IFN-γ.

The distribution of positive studies across different clinical areas further underscores LGG's evolving relevance. Historical research concentrated on IBD/Colitis, Allergy/Atopic conditions, and Liver Disease. Recent literature shows diversification towards Oncology and Sepsis/Liver Injury, joined by Neurological/Behavioral disorders, Bone/Joint/Cardiovascular disease, and Gut Barrier/Microbiome research.

From a practical point of view, retrieving relevant information amidst today's scientific publication overload demands alternatives to traditional systematic review. This work proposes freely available LLMs as a solution. Copilot was initially considered but relied solely on titles, yielding an unacceptably high false-negative rate. Gemini 3 Flash emerged as the primary screening tool, ensuring minimal missed evidence, while Perplexity Academic and Claude 4.6 Sonnet acted as high-precision filters during refining and final validation [42]

This complementarity is well explained by architectural differences. Gemini's Sparse Mixture-of-Experts architecture favors rapid inference and high sensitivity [43],whereas Perplexity's multistage RAG pipeline contributes to its stability and specificity [44]. Claude's Constitutional AI framework reinforces rule-based rigor and caution, producing distinct model "personalities": conservative on historical literature and more flexible on recent literature.

We mapped three failure types limiting full automation, collectively named the "Semantic Floor", or the threshold beyond which AI precision falters without human intervention: i) the Title Trap, where LGG was absent from titles despite appearing in abstracts; ii) Strain Ambiguity, reflecting difficulties distinguishing Lactobacillus rhamnosus from the GG strain; and iii) Biological Complexities, including borderline cases such as postbiotic derivatives or engineered strains. These findings underscore the continued necessity of human oversight.

Our study has specific limitations. Screening and reclassification relied on public web interfaces rather than dedicated APIs, limiting reproducibility since commercial platforms undergo continuous updates. In addition, the human-validated gold standard, although rigorously cross-checked, retains subjective interpretation biases. AI screening was largely limited to metadata and abstracts, potentially missing relevant information contained in full texts or supplementary materials. Future work using frozen API endpoints, multi-expert consensus standards, and full-text processing architectures will further benchmark LLM consistency in biomedical curation.

We also considered AI's environmental impact. While LLMs consume energy and water resources, these costs should be weighed against efficiency gains and reductions in human-hour requirements. Beyond environmental considerations, this study's ethical core lies in Open Science. By using free LLMs instead of costly review software, we propose a methodology accessible to researchers in resource-limited settings. LLMs can therefore function as scientific equalizers, providing a low-cost, high-speed framework for democratized evidence synthesis.

Conclusions

This study demonstrates that synergy between human expertise and AI can effectively break the information bottleneck, validating a dataset of over 900 papers spanning nearly 40 years of research. Through AI-driven research, we decoded LGG's immunological fingerprint, revealing a transition from mucosal barrier protection to complex molecular signaling.

Declarations

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was made possible thanks to the financial support of Associazione Italiana per la Ricerca sul Cancro (Investigator Grant-IG 2025 32374 to FF). This work has received funding from the Fondazione Cariplo (Supporting young Italian talents in the European Research Council competitions 2023) to FF. This work has been partially supported by the Italian Ministry of Health with 5 × 1000 research award and Ricerca Corrente from Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico (to F.C.).

Author Contribution

MR: Conceptualization, Formal analysis, Writing –original draft. DN: Investigation, Writing – review & editing. CA: Writing – review &editing, Conceptualization, FC: Writing – review & editing, Writing –original draft. MV: Writing – review & editing. FF: Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgement

We thank Ing. Maurizio Acrì for the fruitful discussions that gave origin to this work.

Data Availability

The datasets presented in this study can be found in the online Zenodo repository (https://doi.org/10.5281/zenodo.20492569).

References

1. Segers ME, Lebeer S (2014) Towards a better understanding of Lactobacillus rhamnosus GG - host interactions. Microb Cell Fact 13:S7
2. Capurso L (2019) Thirty Years of Lactobacillus rhamnosus GG. J Clin Gastroenterol 53:S1–S41
3. Siegler MG (2010) Every 2 Days We Create As Much Information As We Did Up To 2003
4. Landhuis E (2016) Scientific literature: Information overload. Nature 535:457–458

5. Arnold M, Goldschmitt M, Rigotti T (2023) Dealing with information overload: a comprehensive review. Front Psychol 14
6. Teixeira da Silva JA, Nazarovets S (2026) The publish or perish, publish and perish, publish then perish, and now retract and perish cultures in academia. Naunyn Schmiedebergs Arch Pharmacol 399:3115–3131
7. Ge L et al (2024) Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. Syst Rev 13:269
8. Gemini, Team et al (2025) Gemini: A Family of Highly Capable Multimodal Models
9. Lewis P et al (2021) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks
10. Çorbacıoğlu ŞK, Aksel G (2023) Receiver operating characteristic curve analysis in diagnostic accuracy studies. Turk J Emerg Med 23:195–198
11. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 15:155–163
12. Mata Forsberg M et al (2019) Extracellular Membrane Vesicles from Lactobacilli Dampen IFN-γ Responses in a Monocyte-Dependent Manner. Sci Rep 9:17109
13. Fong FLY, El-Nezami H, Mykkänen O, Kirjavainen PV (2022) The Effects of Single Strains and Mixtures of Probiotic Bacteria on Immune Profile in Liver, Spleen, and Peripheral Blood. Front Nutr 9
14. Wu C, Ouyang S, Lai J, Xue Z, Song C (2026) Impact of continuous probiotic supplementation on intestinal barrier function and hepatic biomarkers in fulminant liver failure models. Front Cell Infect Microbiol 16
15. Rago L et al (2026) Heat-Treated Lacticaseibacillus rhamnosus Strains Modulate Inflammatory and Metabolic Processes in In Vitro Systems Relevant to Canine Osteoarthritis. Cells 15:336
16. Kandasamy M et al (2011) *Lactobacillus rhamnosus* GG secreting an antigen and Interleukin-2 translocates across the gastrointestinal tract and induces an antigen specific immune response. Microbiol Immunol 55:704–714
17. Zhang Z et al (2013) Protective effects of *Lactobacillus rhamnosus* GG against human rotavirus-induced diarrhoea in a neonatal mouse model. Pathog Dis 67:184–191
18. Wigg AJ (2001) The role of small intestinal bacterial overgrowth, intestinal permeability, endotoxaemia, and tumour necrosis factor α in the pathogenesis of non-alcoholic steatohepatitis. Gut 48:206–211
19. Tong L et al (2021) Lactobacillus rhamnosus GG Derived Extracellular Vesicles Modulate Gut Microbiota and Attenuate Inflammatory in DSS-Induced Colitis Mice. Nutrients 13:3319
20. Qiu X et al (2023) <scp> *Lactobacillus rhamnosus* </scp> *GG* alleviates colitis caused by chemotherapy via biofilm formation. J Gastroenterol Hepatol 38:1158–1169
21. Guo Y, Du X, Bian Y, Wang S (2020) Chronic unpredictable stress-induced reproductive deficits were prevented by probiotics. Reprod Biol 20:175–183

22. Moludi J et al (2022) The effects of co-administration of probiotics and prebiotics on chronic inflammation, and depression symptoms in patients with coronary artery diseases: a randomized clinical trial. Nutr Neurosci 25:1659–1668
23. Schultz M, K. H-JLNL (2003) Z. J. G. W. F. and J. S. Immunomodulatory consequences of oral administration of Lactobacillus rhamnosus strain GG in healthy volunteers. J Dairy Res 165–173
24. Chen D et al (2023) Vitamin D3 and *Lactobacillus rhamnosus* GG/p40 Synergize to Protect Mice From Colitis by Promoting Vitamin D Receptor Expression and Epithelial Proliferation. Inflamm Bowel Dis 29:620–632
25. Vorobieva OA, Shikh EV, Drozdov VN, Shikh NV (2023) The results of the use of a combined probiotic (Lactobacillus rhamnosus GG and Bifidobacterium animalis spp. lactis BB-12) in children with gastrointestinal and skin manifestations of food allergy. Probl Nutr 92:79–86
26. Mirpuri J et al (2012) Lactobacillus rhamnosus (LGG) Regulates IL-10 Signaling in the Developing Murine Colon through Upregulation of the IL-10R2 Receptor Subunit. PLoS ONE 7:e51955
27. Ding L et al (2019) Lactobacillus rhamnosus GG Ameliorates Liver Injury and Hypoxic Hepatitis in Rat Model of CLP-Induced Sepsis. Dig Dis Sci 64:2867–2877
28. Jiang Y et al (2025) Engineered Probiotics Remodel the Tumor Immune Microenvironment. ACS Nano 19:39303–39319
29. Yu J et al (2022) Design of a self-driven probiotic-CRISPR/Cas9 nanosystem for sono-immunometabolic cancer therapy. Nat Commun 13:7903
30. Szyller K et al (2025) Immune Modulation Through Long-Term *Lacticaseibacillus rhamnosus* Therapy in Home Mechanically Ventilated Patients. *Microbiologyopen* 14
31. Wang Y et al (2017) An LGG-derived protein promotes IgA production through upregulation of APRIL expression in intestinal epithelial cells. Mucosal Immunol 10:373–384
32. Yan F et al (2017) Neonatal colonization of mice with LGG promotes intestinal development and decreases susceptibility to colitis in adulthood. Mucosal Immunol 10:117–127
33. Wang J et al (2024) AhR ligands from *LGG* metabolites promote piglet intestinal ILC3 activation and IL-22 secretion to inhibit PEDV infection. J Virol 98
34. Ge Y et al (2022) The amelioration of alcohol-induced liver and intestinal barrier injury by *Lactobacillus rhamnosus* Gorbach-Goldin (LGG) is dependent on Interleukin 22 (IL-22) expression. Bioengineered 13:12650–12660
35. Gu Z et al (2021) Exosome-Like Nanoparticles From Lactobacillus rhamnosus GG Protect Against Alcohol-Associated Liver Disease Through Intestinal Aryl Hydrocarbon Receptor in Mice. Hepatol Commun 5:846–864
36. Maragno P et al (2026) The Assessment of Multidimensional Clinical, Biological and Patient-Reported Outcomes to Evaluate the Efficacy of Add-On Lactobacillus rhamnosus GG Supplementation in Mild Ulcerative Colitis: A Randomized Pilot Trial. Nutrients 18:1329

37. Si W et al (2022) *Lactobacillus rhamnosus* GG induces cGAS/STING- dependent type I interferon and improves response to immune checkpoint blockade. Gut 71:521–533
38. Ji J et al (2021) Probiotics protect against RSV infection by modulating the microbiota-alveolar-macrophage axis. Acta Pharmacol Sin 42:1630–1641
39. Kang S-J, Jun J-S, Hong K-W (2022) Transcriptome Analysis Reveals Immunomodulatory Effect of Spore-Displayed p75 on Human Intestinal Epithelial Caco-2 Cells. Int J Mol Sci 23:14519
40. Qiu X, Zou Z, Lin T, Guo C, Lin D (2025) Engineered *Lactobacillus rhamnosus* Producing 3-Hydroxybutyrate: A Dual-Action Therapeutic Strategy for Colon Cancer Cachexia. Biotechnol Bioeng 122:1574–1589
41. Diao W et al (2025) DGGcm Loaded with LGG for Promoting Diabetic Infected Wound Healing. Macromol Biosci 25
42. Deshmukh R, Raut R, Bhavsar M, Gurav S, Patil Y, Optimizing Human -AI Interaction: Innovations in Prompt Engineering. in (2025) *3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)* 1240–1246 (IEEE, 2025). 10.1109/IDCIOT64235.2025.10914815
43. Xu Y et al (2026) Mixture of Experts for Decentralized Generative AI and Reinforcement Learning in Wireless Networks: A Comprehensive Survey. IEEE Commun Surv Tutorials 28:4051–4085
44. Bai Y et al (2022) Constitutional AI: Harmlessness from AI Feedback
45. Ren S, Tomlinson B, Black RW, Torrance AW (2024) Reconciling the contrasting narratives on the environmental impact of large language models. Sci Rep 14:26310

Figures

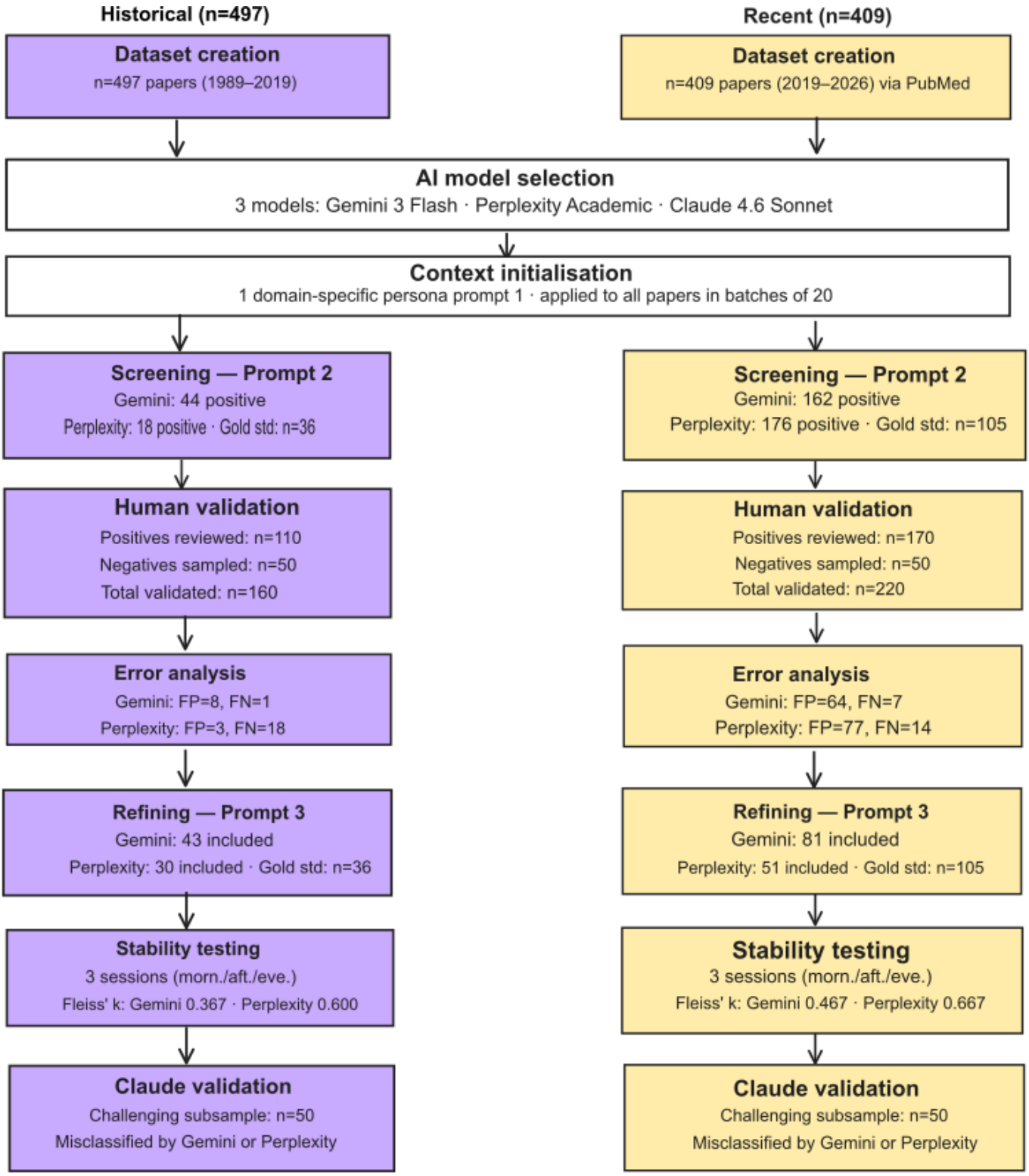


Figure 1

Operative workflow with paper counts split by Historical and Recent dataset.

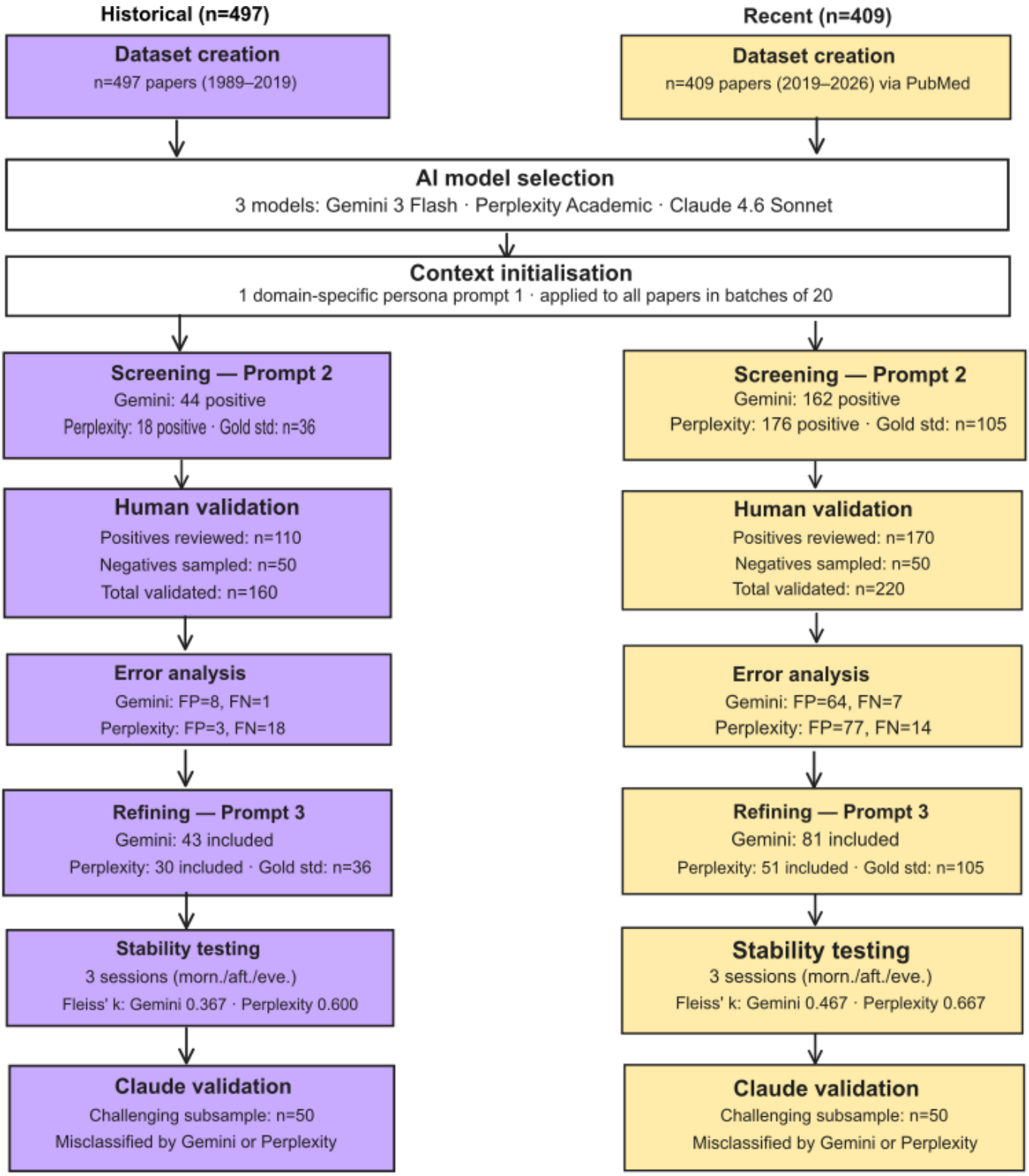


Figure 1

Operative workflow with paper counts split by Historical and Recent dataset.

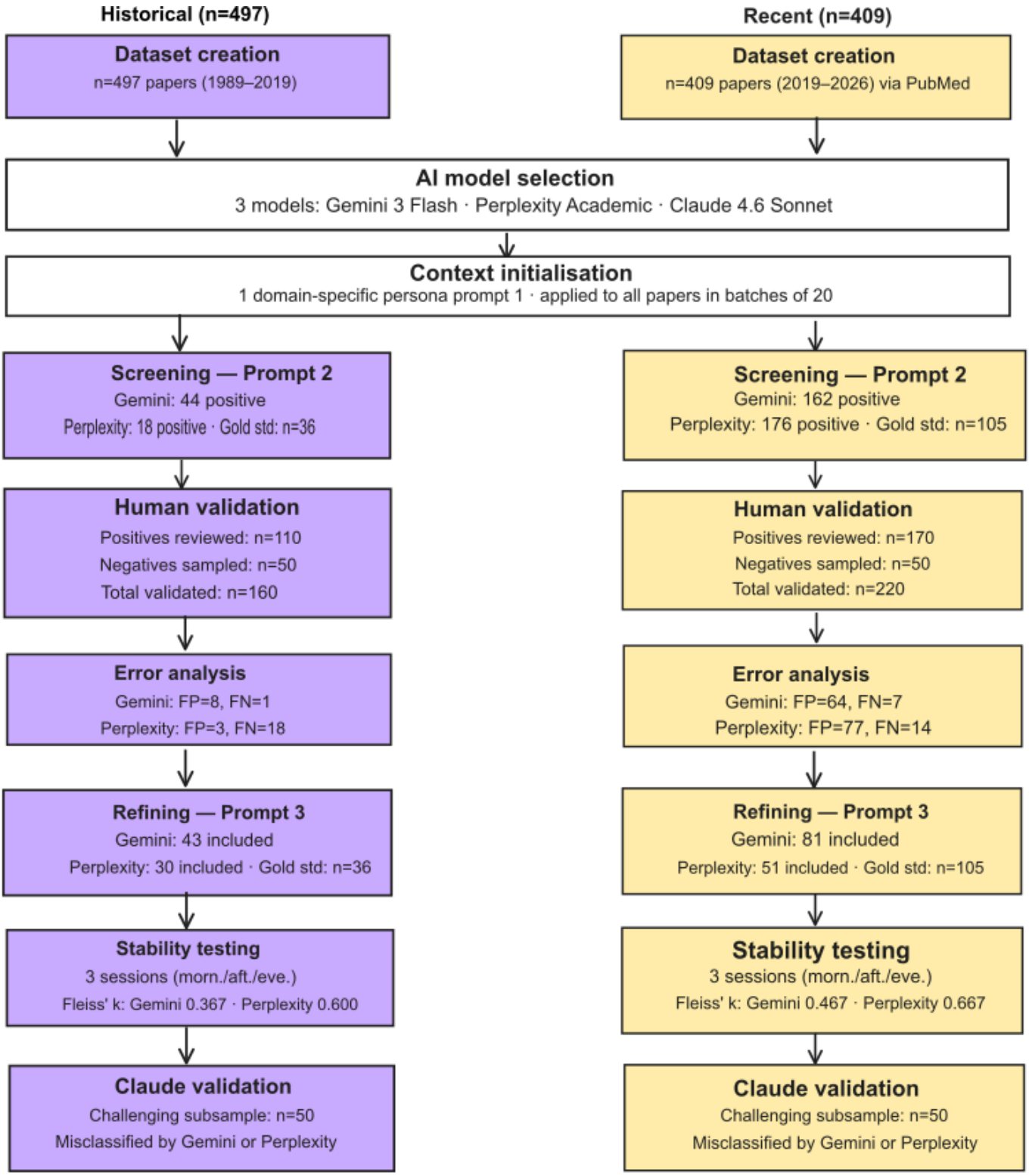


Figure 1

Operative workflow with paper counts split by Historical and Recent dataset.

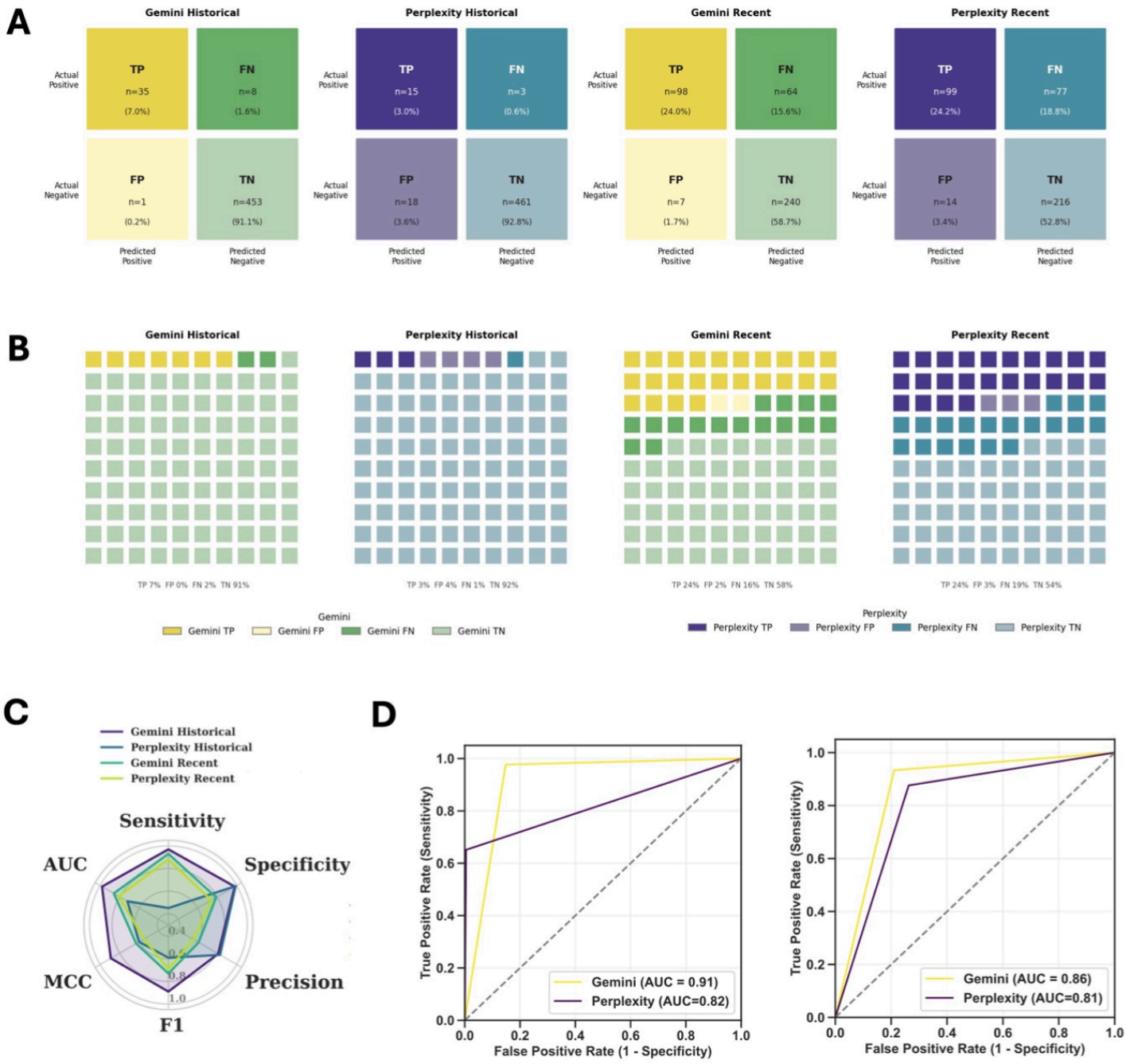


Figure 2

Evaluated classification performance and statistical metrics of Large Language Models (LLMs) in the literature screening pipeline. ***(A)*** *Confusion matrices for the screening phase (Prompt 2): heatmaps display the absolute counts and percentages of True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN) for Gemini 3 Flash and Perplexity Academic, shown separately for the historical dataset (n = 497) and the recent dataset (n = 409).* ***(B)****Waffle plots providing a proportional visual representation of the same TP/FP/FN/TN distributions shown in (A), for each model and dataset.* ***(C)****Comprehensive performance profiles: radar chart summarizing comparative evaluation metrics for Gemini and Perplexity across both chronological datasets. Performance is mapped across six standard*

statistical axes: Sensitivity, Specificity, Precision, F1-Score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC), normalized on a 0.0–1.0 scale to illustrate model-specific classification behavior and trade-offs. ***(D)****ROC curves for the screening phase (Prompt 2), comparing Gemini and Perplexity for the historical (left) and recent (right) datasets, with corresponding AUC values.*

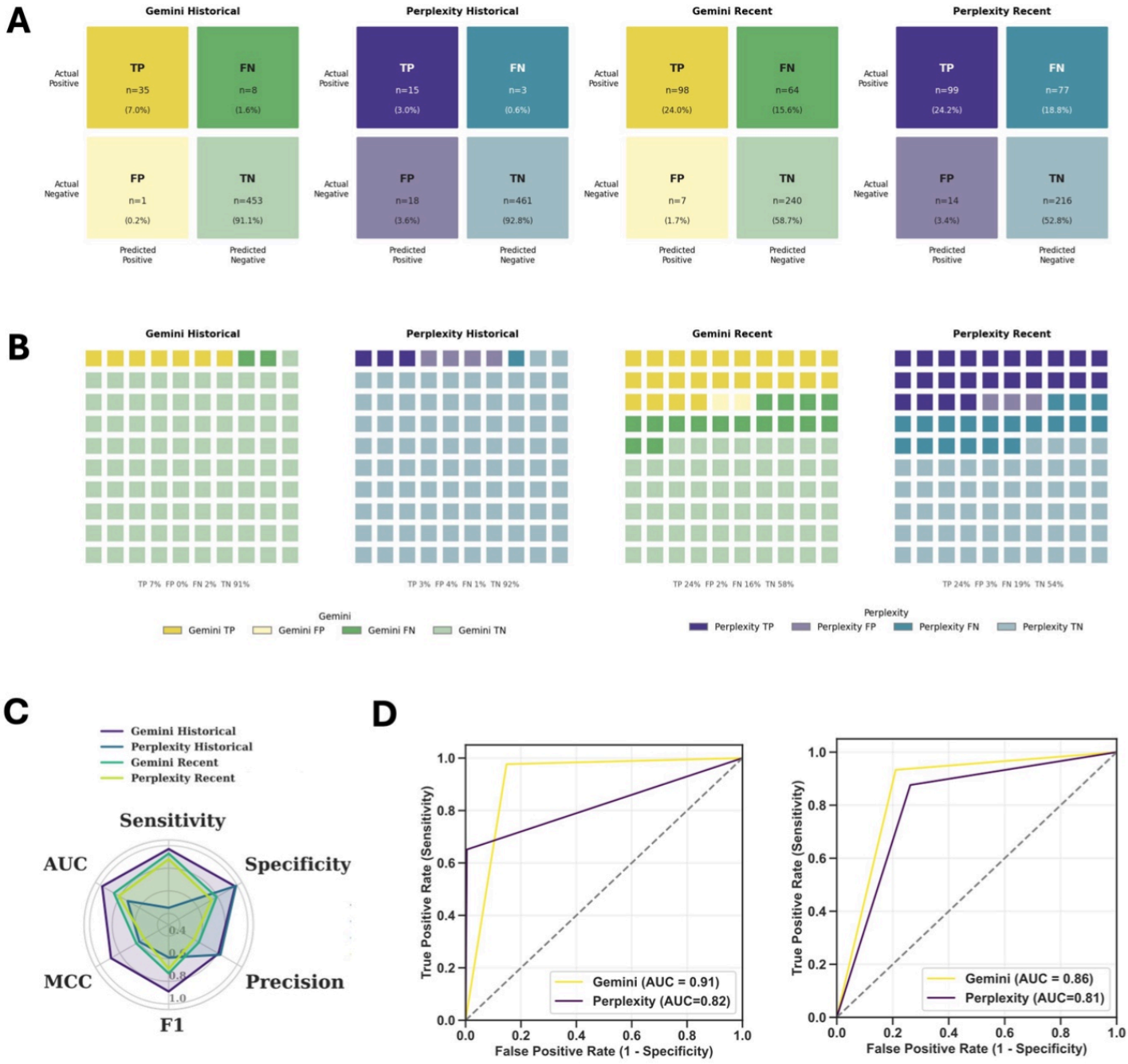


Figure 2

Evaluated classification performance and statistical metrics of Large Language Models (LLMs) in the literature screening pipeline*.* ***(A)*** *Confusion matrices for the screening phase (Prompt 2): heatmaps display the absolute counts and percentages of True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN) for Gemini 3 Flash and Perplexity Academic, shown separately for the historical*

dataset (n = 497) and the recent dataset (n = 409). ***(B)****Waffle plots providing a proportional visual representation of the same TP/FP/FN/TN distributions shown in (A), for each model and dataset.* ***(C)****Comprehensive performance profiles: radar chart summarizing comparative evaluation metrics for Gemini and Perplexity across both chronological datasets. Performance is mapped across six standard statistical axes: Sensitivity, Specificity, Precision, F1-Score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC), normalized on a 0.0–1.0 scale to illustrate model-specific classification behavior and trade-offs.* ***(D)****ROC curves for the screening phase (Prompt 2), comparing Gemini and Perplexity for the historical (left) and recent (right) datasets, with corresponding AUC values.*

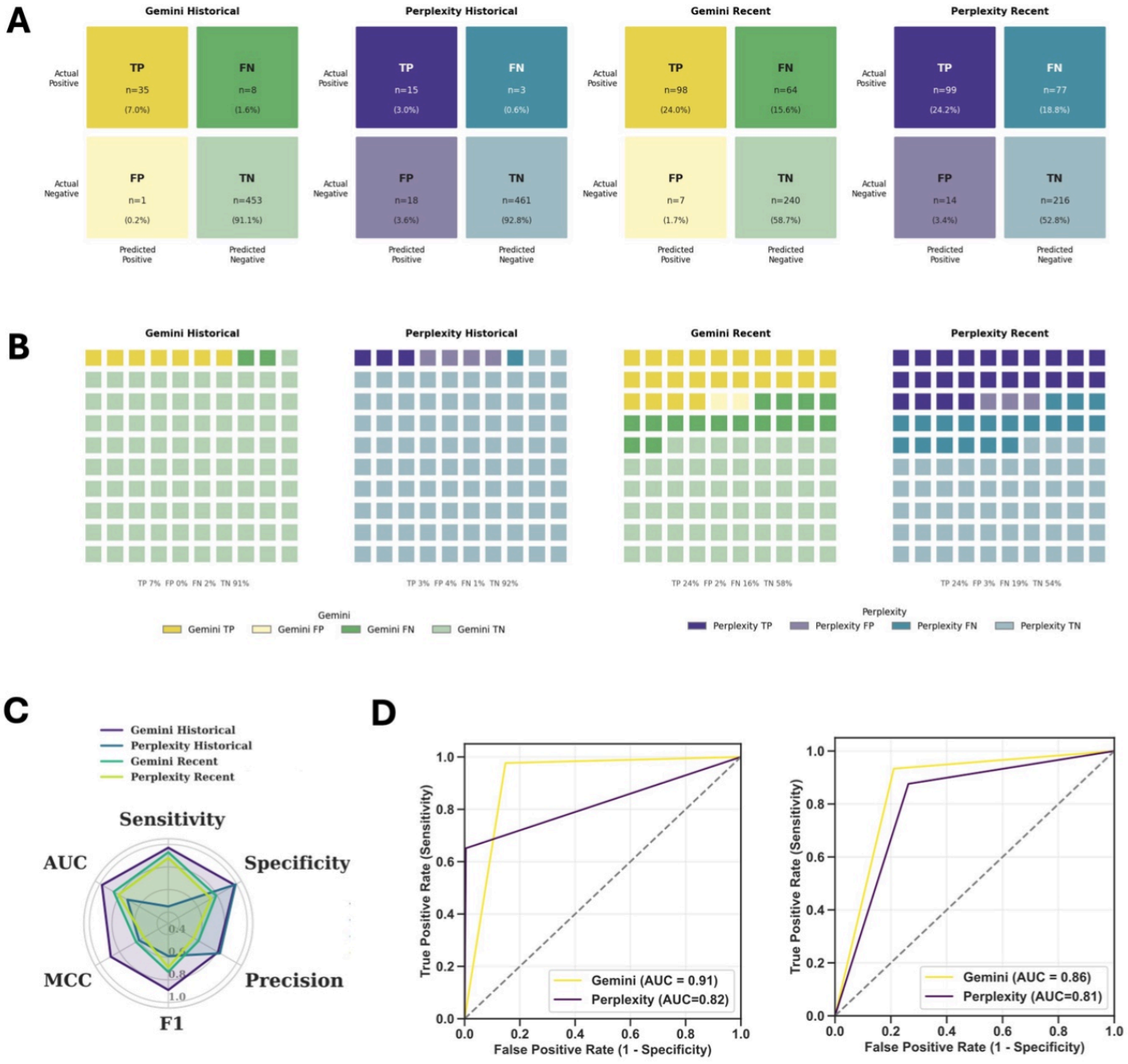


Figure 2

Evaluated classification performance and statistical metrics of Large Language Models (LLMs) in the literature screening pipeline. ***(A)*** *Confusion matrices for the screening phase (Prompt 2): heatmaps display the absolute counts and percentages of True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN) for Gemini 3 Flash and Perplexity Academic, shown separately for the historical dataset (n = 497) and the recent dataset (n = 409).* ***(B)****Waffle plots providing a proportional visual representation of the same TP/FP/FN/TN distributions shown in (A), for each model and dataset.* ***(C)****Comprehensive performance profiles: radar chart summarizing comparative evaluation metrics for Gemini and Perplexity across both chronological datasets. Performance is mapped across six standard statistical axes: Sensitivity, Specificity, Precision, F1-Score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC), normalized on a 0.0–1.0 scale to illustrate model-specific classification behavior and trade-offs.* ***(D)****ROC curves for the screening phase (Prompt 2), comparing Gemini and Perplexity for the historical (left) and recent (right) datasets, with corresponding AUC values.*

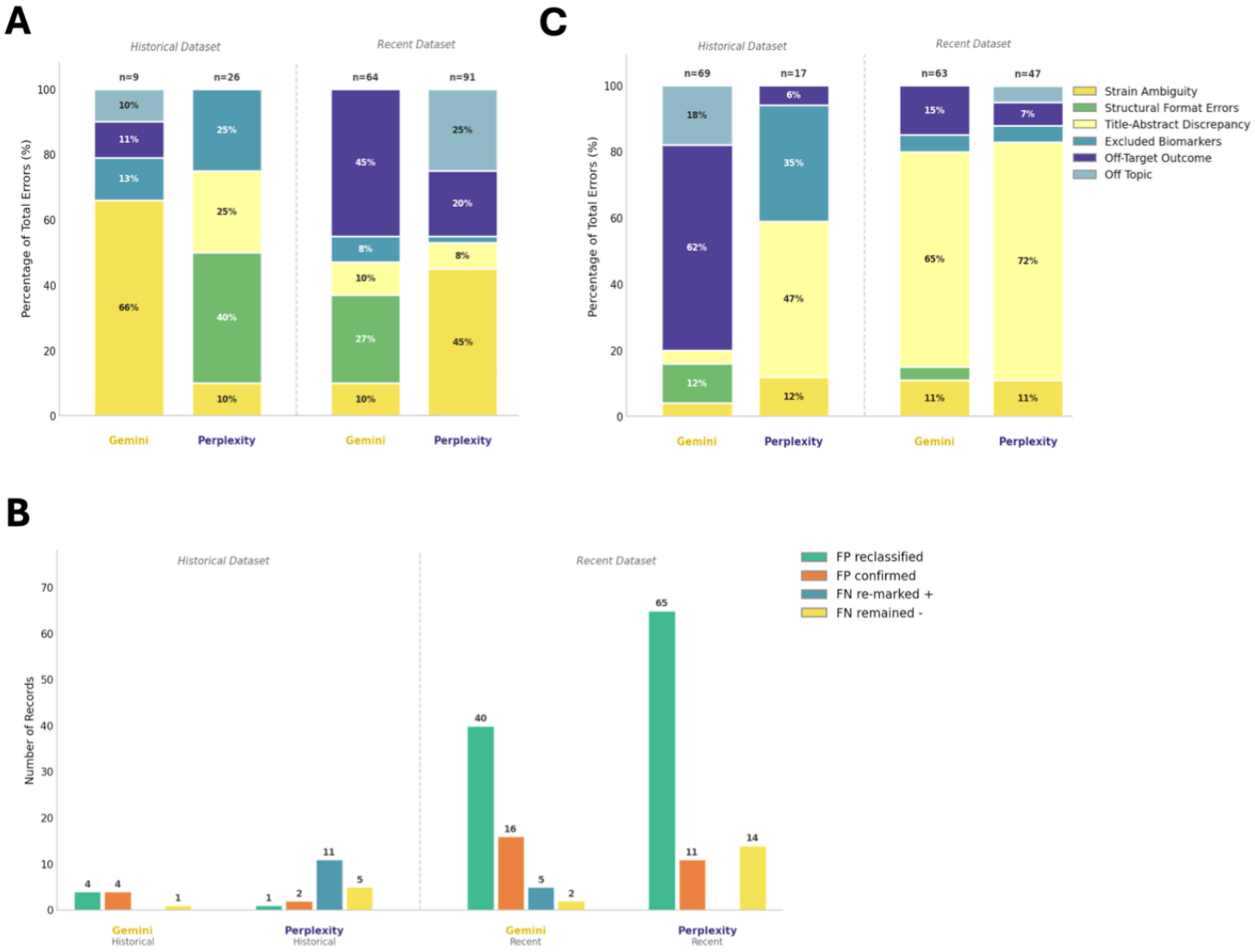


Figure 3

Qualitative error profiling and impact of quality score reclassification on LLM screening outcomes. *(A,B) Qualitative error analysis: 100% stacked bar chart illustrating the relative distribution of error types encountered during the screening phase (A) and the refinement phase (B) across models and chronological subsets. Total errors are classified into six distinct qualitative categories: Strain Ambiguity, Structural format errors, Title-Abstract Discrepancy, Excluded Biomarkers, Off-Target Outcome, and Off topic. (C) Quality score reclassification dynamics: Distribution bar plot demonstrating the quantitative correction effect when applying a stringent quality score threshold (threshold< 0.8). The chart tracks the transition of misclassified records across four categories: False Positives reclassified (FP reclassified), False Positives confirmed (FP confirmed), False Negatives re-marked as positive (FN re-marked +), and False Negatives that remained negative (FN remained -) for Gemini 3 Flash and Perplexity Academic across both the historical and recent literature datasets.*

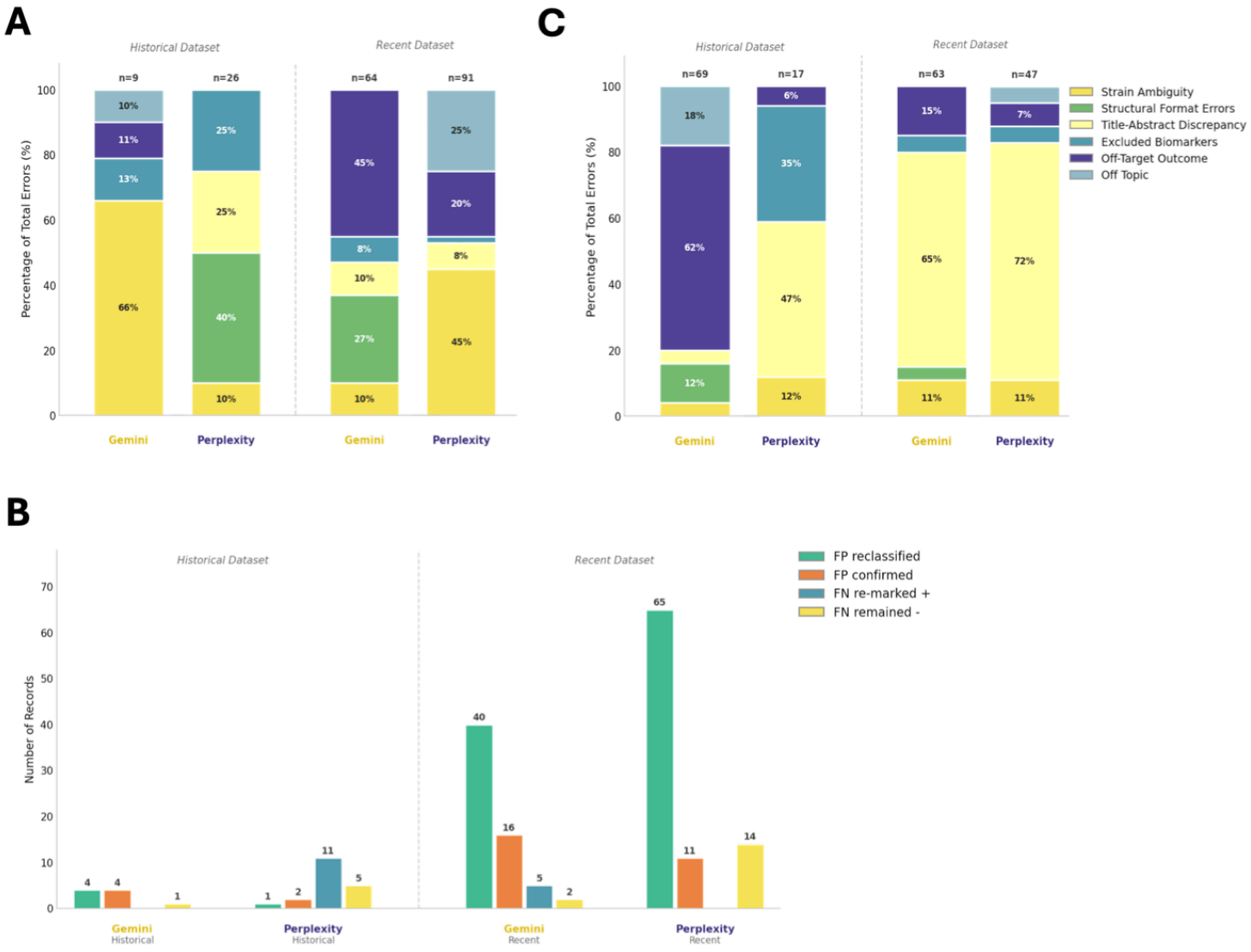


Figure 3

Qualitative error profiling and impact of quality score reclassification on LLM screening outcomes. *(A,B) Qualitative error analysis: 100% stacked bar chart illustrating the relative distribution of error types*

encountered during the screening phase (A) and the refinement phase (B) across models and chronological subsets. Total errors are classified into six distinct qualitative categories: Strain Ambiguity, Structural format errors, Title-Abstract Discrepancy, Excluded Biomarkers, Off-Target Outcome, and Off topic. (C) Quality score reclassification dynamics: Distribution bar plot demonstrating the quantitative correction effect when applying a stringent quality score threshold (threshold< 0.8). The chart tracks the transition of misclassified records across four categories: False Positives reclassified (FP reclassified), False Positives confirmed (FP confirmed), False Negatives re-marked as positive (FN re-marked +), and False Negatives that remained negative (FN remained -) for Gemini 3 Flash and Perplexity Academic across both the historical and recent literature datasets.

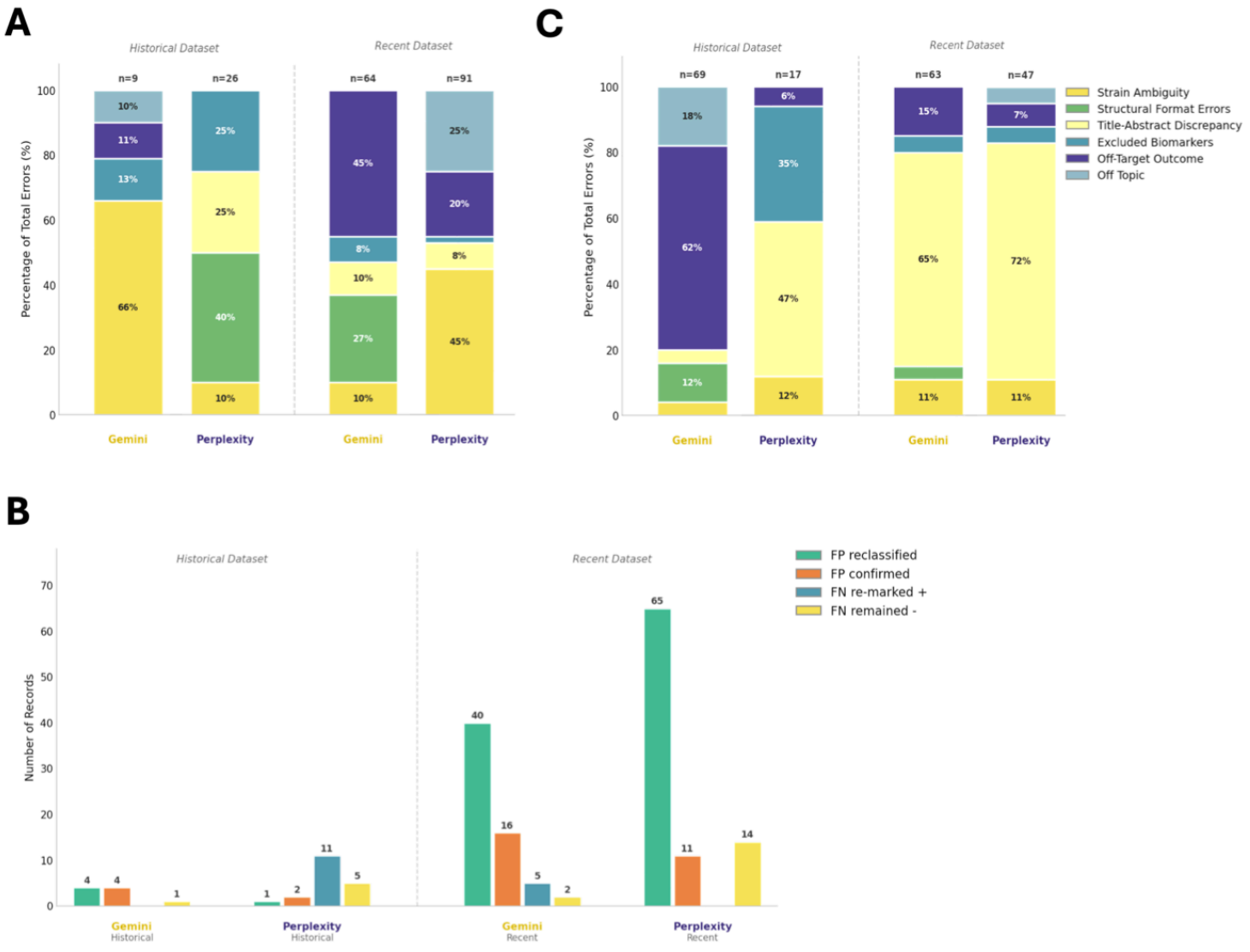


Figure 3

Qualitative error profiling and impact of quality score reclassification on LLM screening outcomes. *(A,B) Qualitative error analysis: 100% stacked bar chart illustrating the relative distribution of error types encountered during the screening phase (A) and the refinement phase (B) across models and chronological subsets. Total errors are classified into six distinct qualitative categories: Strain Ambiguity, Structural*

format errors, Title-Abstract Discrepancy, Excluded Biomarkers, Off-Target Outcome, and Off topic. (C) Quality score reclassification dynamics: Distribution bar plot demonstrating the quantitative correction effect when applying a stringent quality score threshold (threshold< 0.8). The chart tracks the transition of misclassified records across four categories: False Positives reclassified (FP reclassified), False Positives confirmed (FP confirmed), False Negatives re-marked as positive (FN re-marked +), and False Negatives that remained negative (FN remained -) for Gemini 3 Flash and Perplexity Academic across both the historical and recent literature datasets.

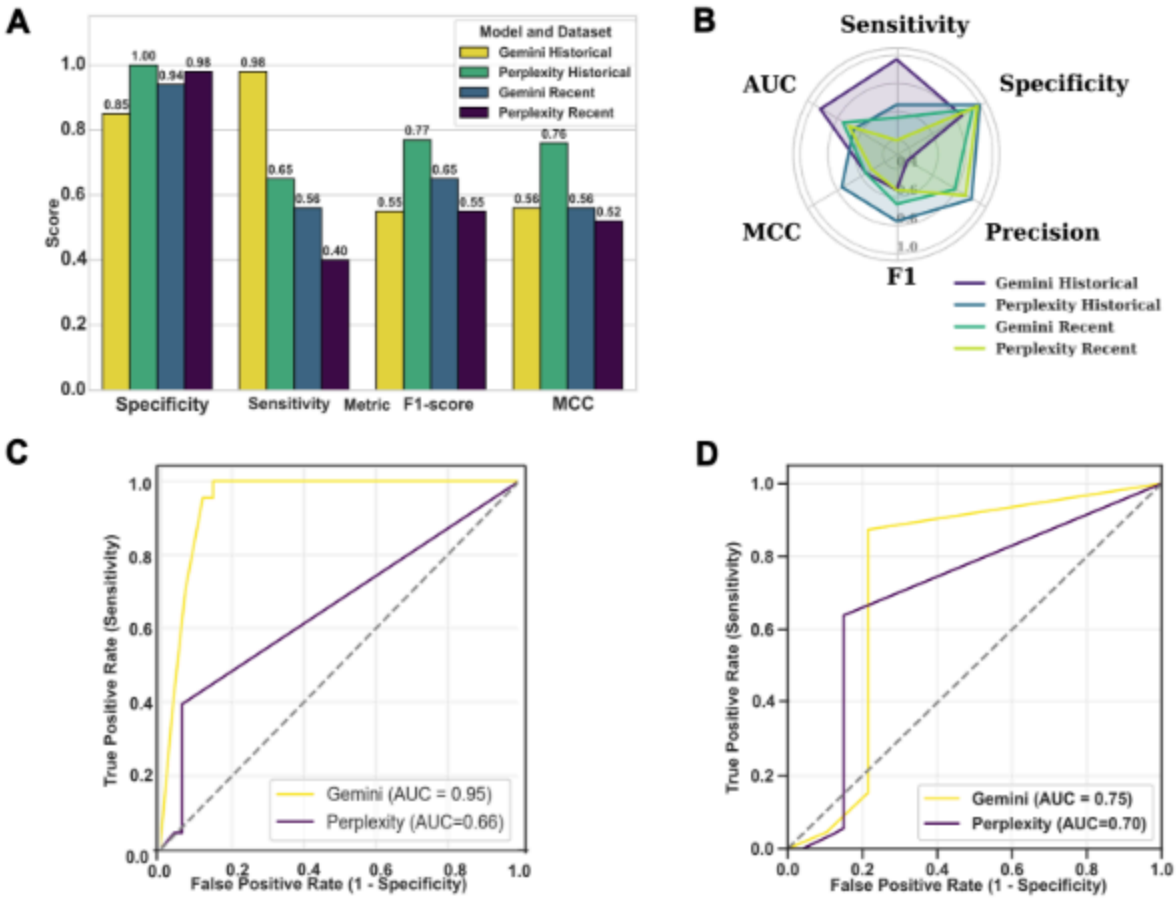


Figure 4

Performance benchmarking of the multi-stage LLM pipeline. *(A) Absolute screening distribution: Bar plot comparing the performance of Gemini 3 Flash and Perplexity Academic during the refining phase (Prompt 3). Quantitative distributions of True Positives (TP) and False Negatives (FN) are shown across both the historical literature dataset (n = 497) and the recent literature dataset (n = 409). (B) Comprehensive performance profiles: Radar chart summarizing the comparative evaluation metrics across both chronological datasets. Performance is mapped across six standard statistical axes: Sensitivity, Specificity, Precision, F1-Score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC). Values are normalized on a scale from 0.0 to 1.0 to illustrate model-specific classification behavior and trade-offs (C-D) ROC curves demonstrating the performance uplift achieved through the refining stage (Prompt 3).*

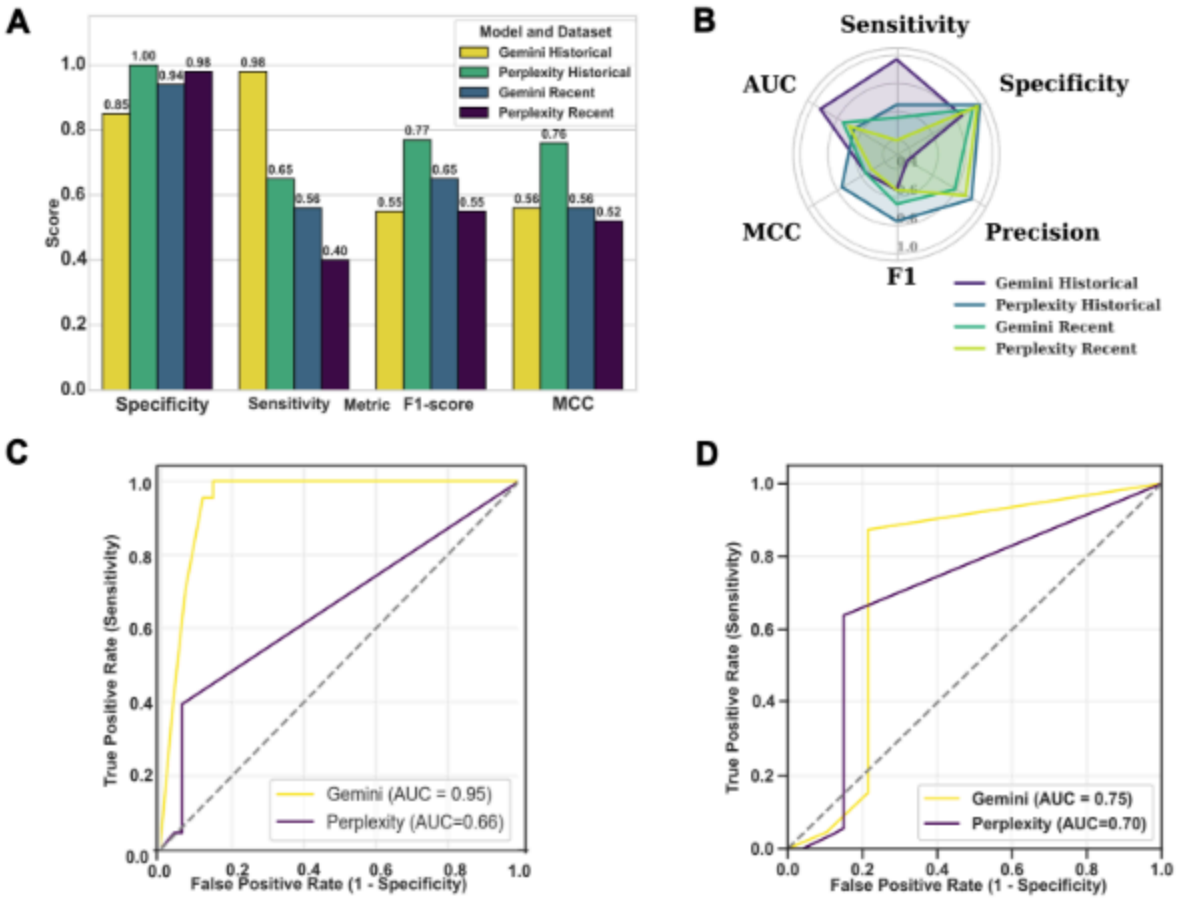


Figure 4

Performance benchmarking of the multi-stage LLM pipeline. *(A) Absolute screening distribution: Bar plot comparing the performance of Gemini 3 Flash and Perplexity Academic during the refining phase (Prompt 3). Quantitative distributions of True Positives (TP) and False Negatives (FN) are shown across both the historical literature dataset (n = 497) and the recent literature dataset (n = 409). (B) Comprehensive performance profiles: Radar chart summarizing the comparative evaluation metrics across both chronological datasets. Performance is mapped across six standard statistical axes: Sensitivity, Specificity, Precision, F1-Score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC). Values are normalized on a scale from 0.0 to 1.0 to illustrate model-specific classification behavior and trade-offs (C-D) ROC curves demonstrating the performance uplift achieved through the refining stage (Prompt 3).*

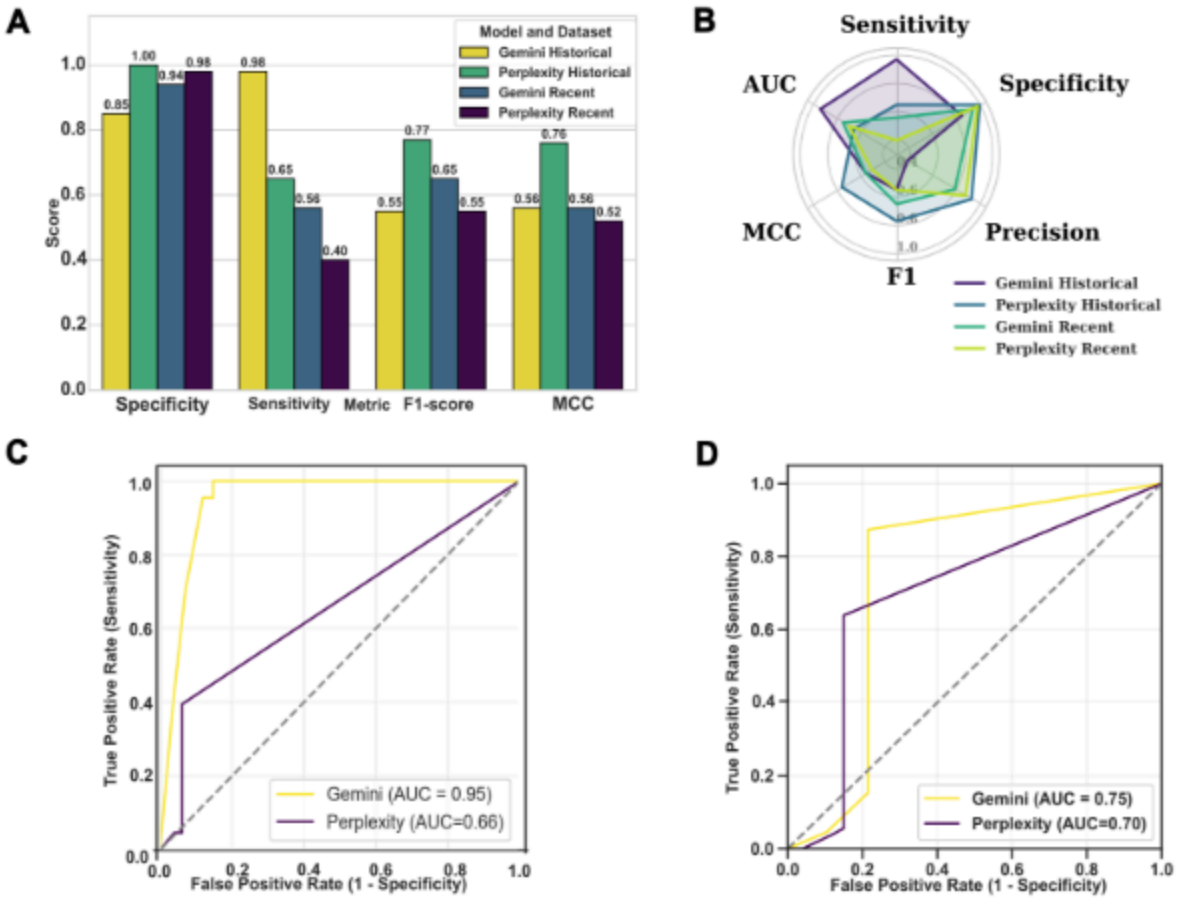


Figure 4

Performance benchmarking of the multi-stage LLM pipeline. *(A) Absolute screening distribution: Bar plot comparing the performance of Gemini 3 Flash and Perplexity Academic during the refining phase (Prompt 3). Quantitative distributions of True Positives (TP) and False Negatives (FN) are shown across both the historical literature dataset (n = 497) and the recent literature dataset (n = 409). (B) Comprehensive performance profiles: Radar chart summarizing the comparative evaluation metrics across both chronological datasets. Performance is mapped across six standard statistical axes: Sensitivity, Specificity, Precision, F1-Score, Matthews Correlation Coefficient (MCC), and Area Under the ROC Curve (AUC). Values are normalized on a scale from 0.0 to 1.0 to illustrate model-specific classification behavior and trade-offs (C-D) ROC curves demonstrating the performance uplift achieved through the refining stage (Prompt 3).*

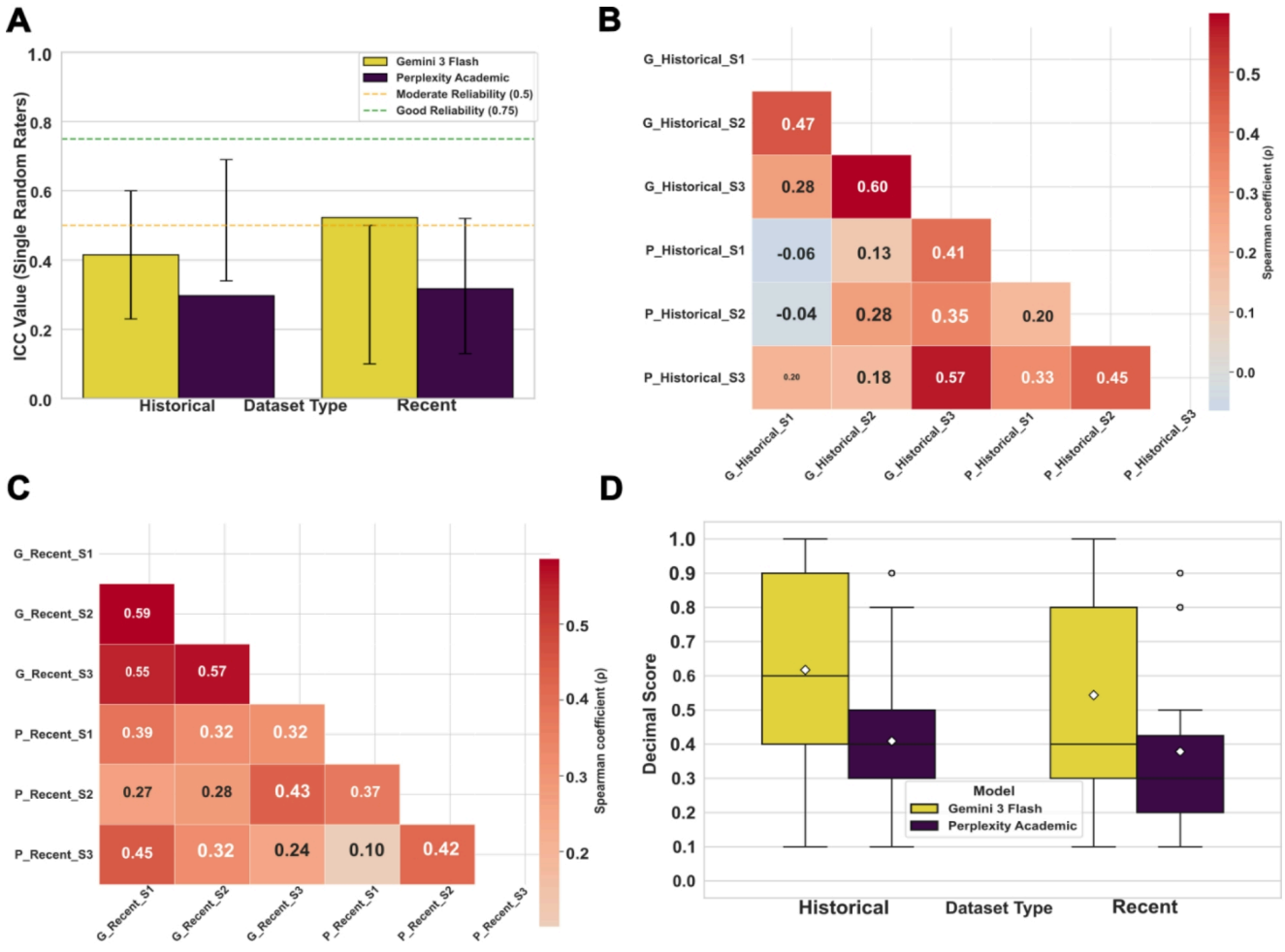


Figure 5

Statistical consistency and inter-model reliability analysis. *(A-B) Spearman correlation heatmaps comparing Gemini 3 Flash and Perplexity Academic outputs; the low correlation values statistically validate the architectural complementarity of the models. (C) Intraclass Correlation Coefficient (ICC2) results demonstrating high stability across autonomous sessions. (D) Distribution of AI-driven Quality Scores across the validated Lactobacillus rhamnosus GG corpus.*

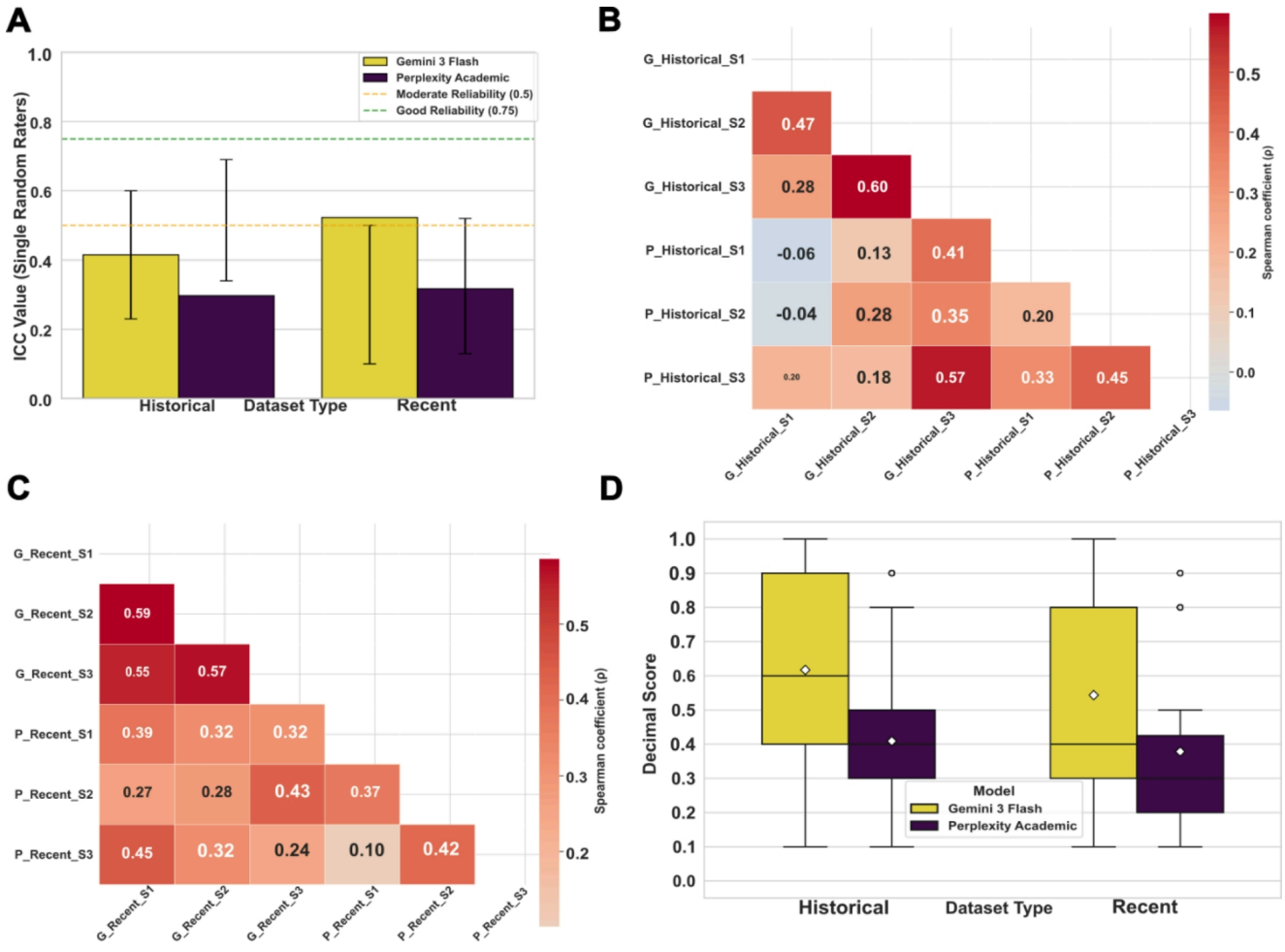


Figure 5

Statistical consistency and inter-model reliability analysis. *(A-B) Spearman correlation heatmaps comparing Gemini 3 Flash and Perplexity Academic outputs; the low correlation values statistically validate the architectural complementarity of the models. (C) Intraclass Correlation Coefficient (ICC2) results demonstrating high stability across autonomous sessions. (D) Distribution of AI-driven Quality Scores across the validated Lactobacillus rhamnosus GG corpus.*

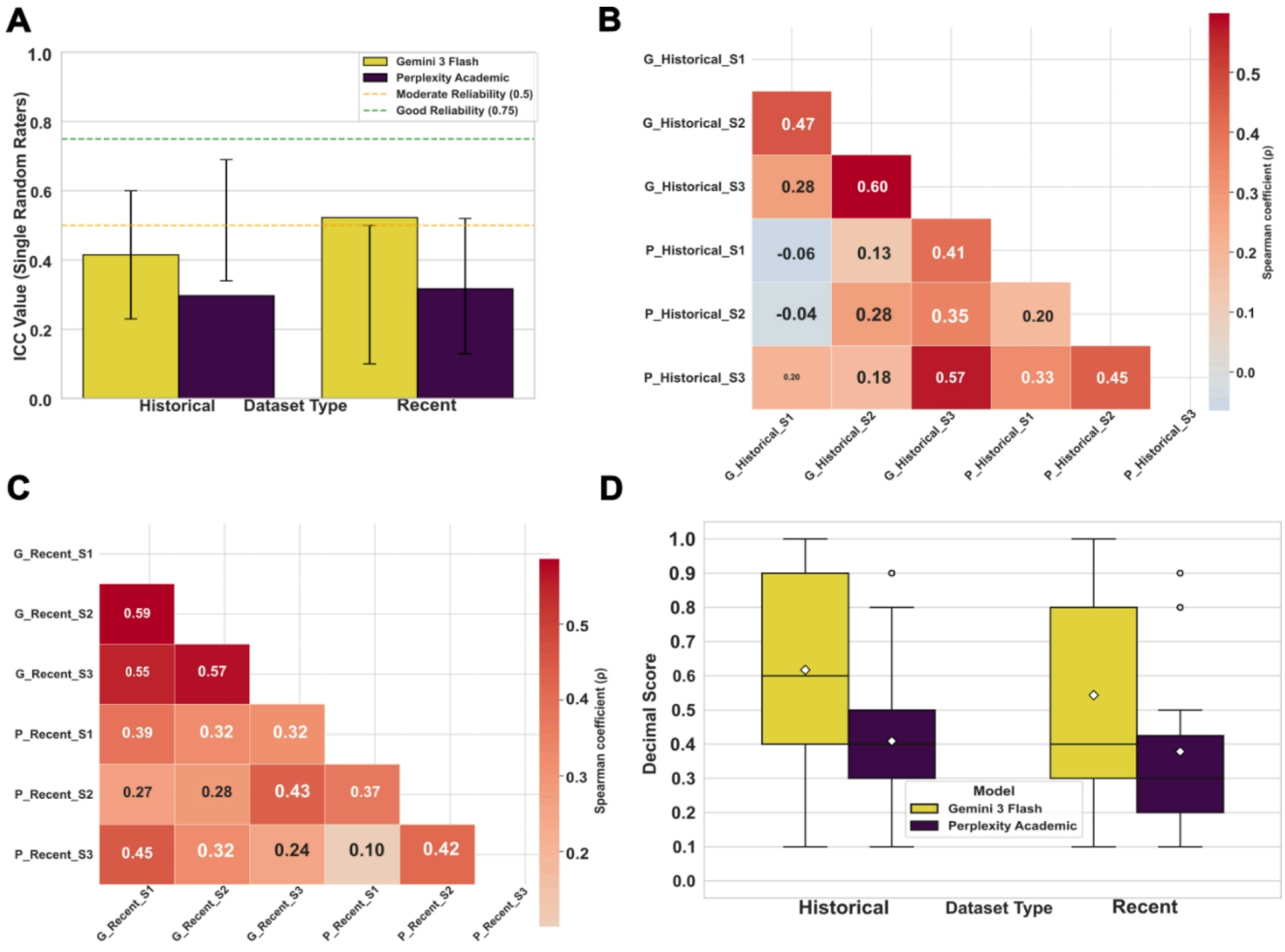


Figure 5

Statistical consistency and inter-model reliability analysis. *(A-B) Spearman correlation heatmaps comparing Gemini 3 Flash and Perplexity Academic outputs; the low correlation values statistically validate the architectural complementarity of the models. (C) Intraclass Correlation Coefficient (ICC2) results demonstrating high stability across autonomous sessions. (D) Distribution of AI-driven Quality Scores across the validated Lactobacillus rhamnosus GG corpus.*

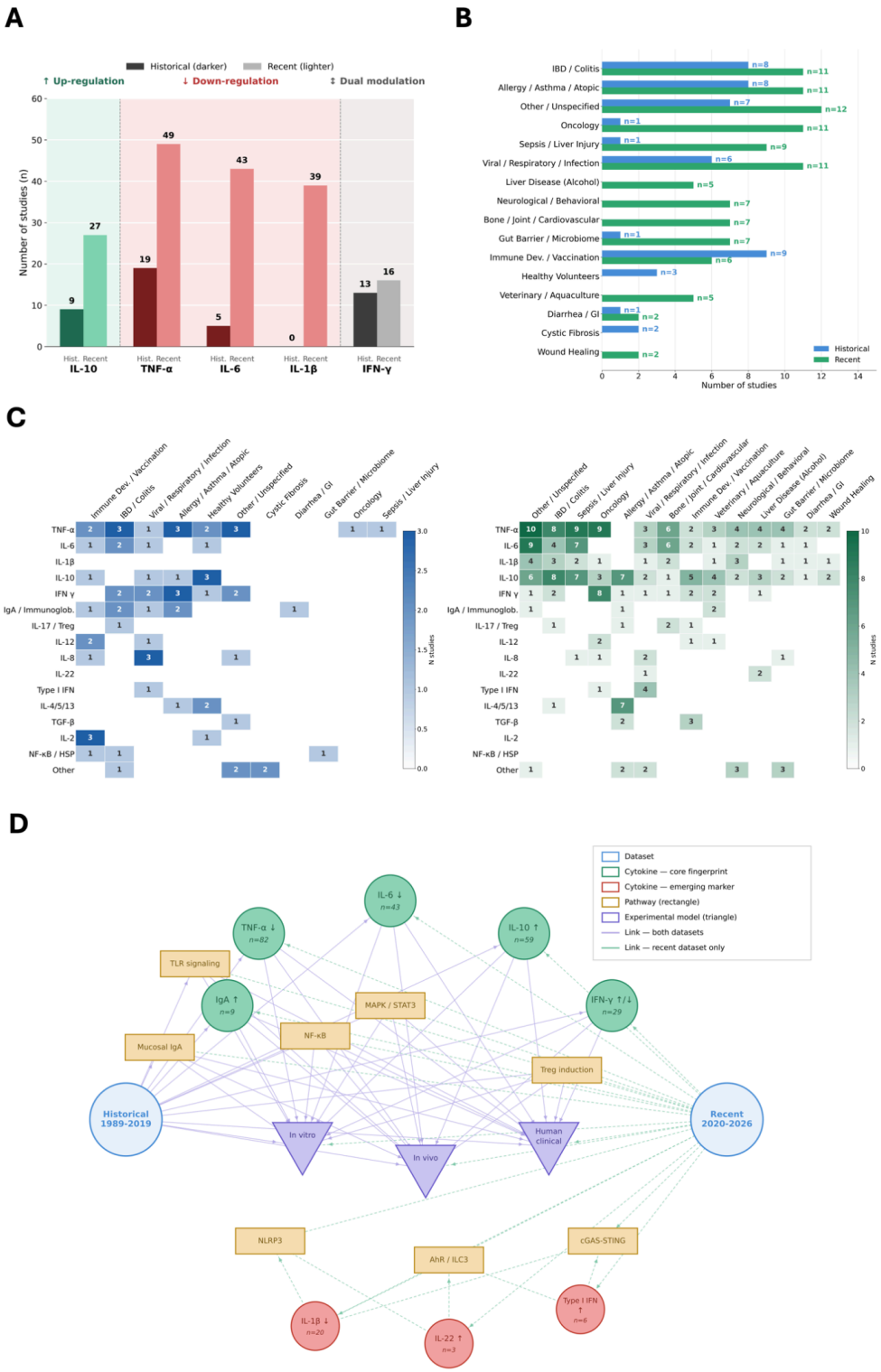


Figure 6

LGG's immunomodulatory fingerprint across historical (1989-2019) and recent (2020-2026) literature. (A) *Number of studies reporting the five core cytokines (IL-10, TNF-α, IL-6, IL-1β, IFN-γ) in the historical (darker bars) versus recent (lighter bars) datasets, color-coded by modulation direction: up-regulation (green), down-regulation (red), and dual modulation (grey).* ***(B)*** *Distribution of positive studies across clinical/therapeutic areas in the historical (blue) and recent (green) datasets, illustrating the diversification*

of LGG research beyond its classical gastrointestinal and atopic focus. ***(C)*** *Heatmaps showing the number of studies reporting each cytokine/marker across clinical areas, for the historical (left, blue) and recent (right, green) datasets.* ***(D)*** *Network diagram linking datasets (blue circles), core-fingerprint cytokines (green circles) and emerging markers (red circles) to their associated signaling pathways (orange rectangles) and experimental models (purple triangles: in vitro, in vivo, human clinical). Purple edges indicate associations present in both datasets; teal edges indicate associations found only in the recent dataset, highlighting the emergence of new mechanistic pathways (e.g., NLRP3, AhR/ILC3, cGAS-STING) alongside the persistent core signature.*

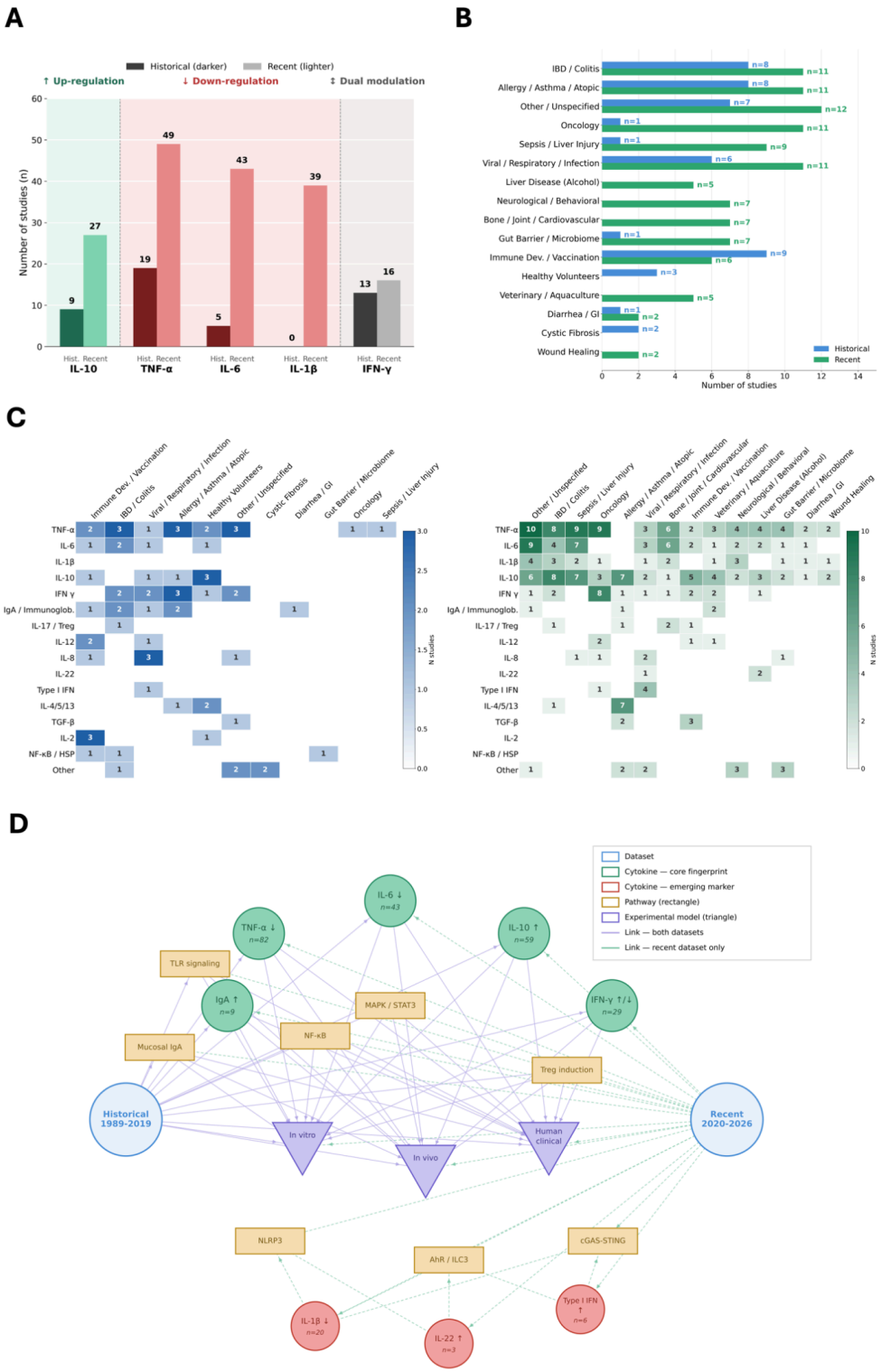


Figure 6

LGG's immunomodulatory fingerprint across historical (1989-2019) and recent (2020-2026) literature. (A) *Number of studies reporting the five core cytokines (IL-10, TNF-α, IL-6, IL-1β, IFN-γ) in the historical (darker bars) versus recent (lighter bars) datasets, color-coded by modulation direction: up-regulation (green), down-regulation (red), and dual modulation (grey).* ***(B)*** *Distribution of positive studies across clinical/therapeutic areas in the historical (blue) and recent (green) datasets, illustrating the diversification*

of LGG research beyond its classical gastrointestinal and atopic focus. ***(C)*** *Heatmaps showing the number of studies reporting each cytokine/marker across clinical areas, for the historical (left, blue) and recent (right, green) datasets.* ***(D)*** *Network diagram linking datasets (blue circles), core-fingerprint cytokines (green circles) and emerging markers (red circles) to their associated signaling pathways (orange rectangles) and experimental models (purple triangles: in vitro, in vivo, human clinical). Purple edges indicate associations present in both datasets; teal edges indicate associations found only in the recent dataset, highlighting the emergence of new mechanistic pathways (e.g., NLRP3, AhR/ILC3, cGAS-STING) alongside the persistent core signature.*

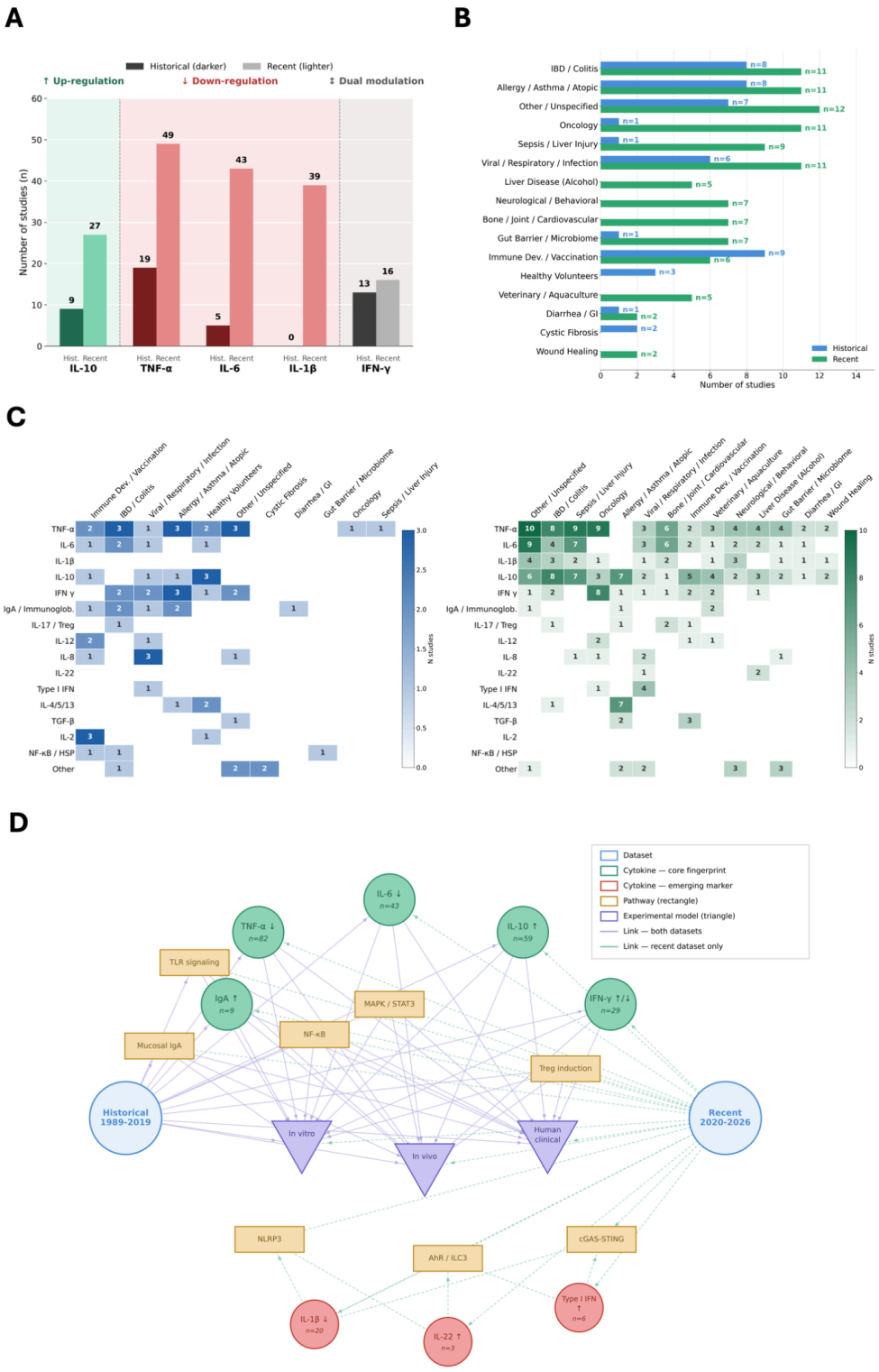


Figure 6

LGG's immunomodulatory fingerprint across historical (1989-2019) and recent (2020-2026) literature. (A) *Number of studies reporting the five core cytokines (IL-10, TNF-α, IL-6, IL-1β, IFN-γ) in the historical (darker bars) versus recent (lighter bars) datasets, color-coded by modulation direction: up-regulation (green), down-regulation (red), and dual modulation (grey).* ***(B)*** *Distribution of positive studies across clinical/therapeutic areas in the historical (blue) and recent (green) datasets, illustrating the diversification*

of LGG research beyond its classical gastrointestinal and atopic focus. ***(C)*** *Heatmaps showing the number of studies reporting each cytokine/marker across clinical areas, for the historical (left, blue) and recent (right, green) datasets.* ***(D)*** *Network diagram linking datasets (blue circles), core-fingerprint cytokines (green circles) and emerging markers (red circles) to their associated signaling pathways (orange rectangles) and experimental models (purple triangles: in vitro, in vivo, human clinical). Purple edges indicate associations present in both datasets; teal edges indicate associations found only in the recent dataset, highlighting the emergence of new mechanistic pathways (e.g., NLRP3, AhR/ILC3, cGAS-STING) alongside the persistent core signature.*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graphicalabstract.docx
- Graphicalabstract.docx
- Graphicalabstract.docx
- RoccaetalSupplementaryMaterial.docx
- RoccaetalSupplementaryMaterial.docx
- RoccaetalSupplementaryMaterial.docx