



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Reproducibility Signals for Review Support

*The paper has a GitHub, the GitHub has a README, **the README has nothing***



Federico Bolelli*



Davide Santoli*



Kevin Marchesini



Luca Lumetti



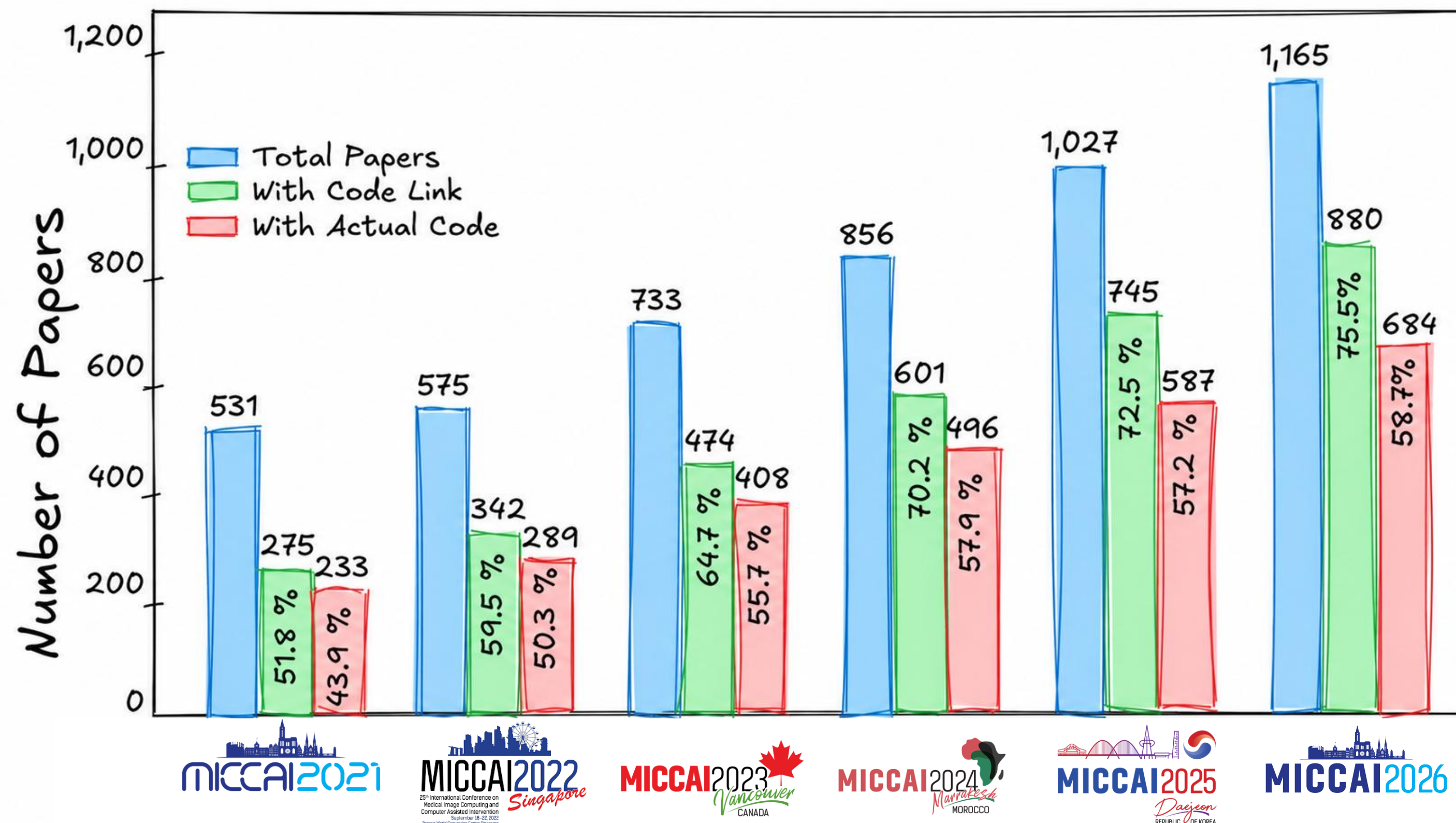
Costantino Grana

AlmageLab · University of Modena and Reggio Emilia, Italy

federico.bolelli@unimore.it

** Equal contribution*

Code sharing is rising. **Verifiable code is not keeping up proportionally.**



MICCAI explicitly lists **reproducible research policies**. Verifying them is left to the reviewer.

1

Model and Algorithms

- Settings
- Framework used
- ...

2

Datasets

- Statistics
- Annotation Protocol
- ...

3

Code Release

- Training code + README
- Pre-trained models
- ...

4

Experimental Results

- Hyperparameters
- Statistical analysis
- ...

conferences.miccai.org

6. Reproducible Research

MICCAI is committed to reproducible research. At MICCAI 2026, we strongly encourage authors to improve the reproducibility of their work by providing open data, open implementations, and appropriate evaluation design and reporting. Where possible, we invite authors to use code available for open access by other researchers. We will instruct reviewers and Area Chairs to take reproducibility of the submission.

Authors are strongly encouraged to consider the following rules for scientific reproducibility.

- **For all models and algorithms, please include**
 - A description of the mathematical setting, algorithm, and/or model.
 - An explanation of any assumptions.
 - A declaration of what software framework and version you used.
- **For all datasets used, please include:**
 - The relevant statistics, such as the number of examples.
 - Description of the study cohort.
 - For existing datasets, citations as well as descriptions if they are not publicly available.
 - For new data collected, a complete description of the data collection process, such as descriptions of the experimental acquisition parameters, subjects/objects involved, instructions to annotators, and methods for quality control.
 - A link to a downloadable version of the dataset (if public).
 - Whether ethics approval was necessary for the data.
- **For all code related to this work that you have made available or will release if this work is accepted, please include:**
 - Specification of dependencies.
 - Training code.

MICCAI explicitly lists **reproducible research policies**. Verifying them is left to the reviewer.

1 Model and Algorithms

- Settings
- Framework used
- ...

2 Datasets

- Statistics
- Annotation Protocol
- ...

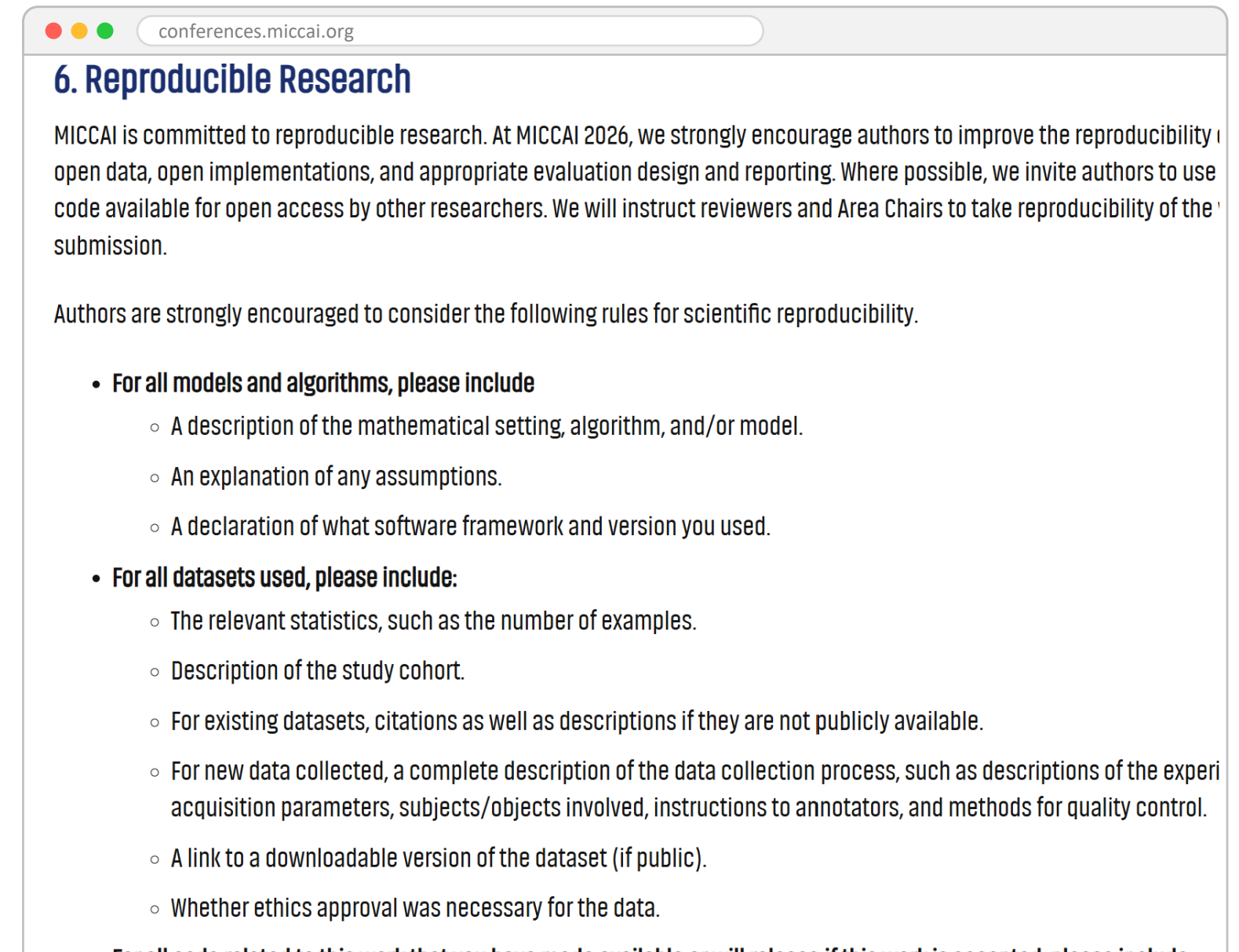
3 Code Release

- Training code + README
- Pre-trained models
- ...

4 Experimental Results

- Hyperparameters
- Statistical analysis
- ...

Many review-time questions are **verifications**,
not judgments!



conferences.miccai.org

6. Reproducible Research

MICCAI is committed to reproducible research. At MICCAI 2026, we strongly encourage authors to improve the reproducibility of their work by providing open data, open implementations, and appropriate evaluation design and reporting. Where possible, we invite authors to use code available for open access by other researchers. We will instruct reviewers and Area Chairs to take reproducibility of the submission.

Authors are strongly encouraged to consider the following rules for scientific reproducibility.

- For all models and algorithms, please include
 - A description of the mathematical setting, algorithm, and/or model.
 - An explanation of any assumptions.
 - A declaration of what software framework and version you used.
- For all datasets used, please include:
 - The relevant statistics, such as the number of examples.
 - Description of the study cohort.
 - For existing datasets, citations as well as descriptions if they are not publicly available.
 - For new data collected, a complete description of the data collection process, such as descriptions of the experimental acquisition parameters, subjects/objects involved, instructions to annotators, and methods for quality control.
 - A link to a downloadable version of the dataset (if public).
 - Whether ethics approval was necessary for the data.

For all code related to this work that you have made available or will release if this work is accepted, please include:

Can we leverage modern AI tooling to perform **bounded checks** and present the results as **decision support**, not an automated gatekeeper, in a transparent, evidence-grounded report?

01

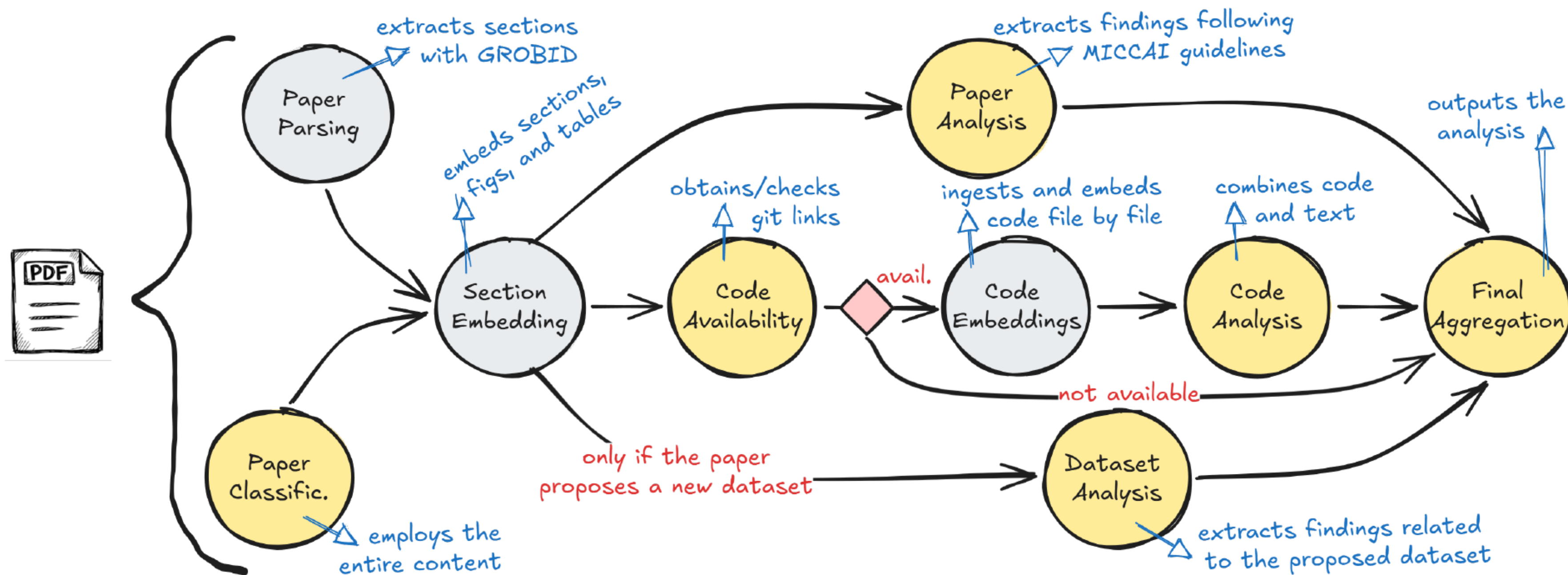
02

03

Scalable and auditable

The system never executes untrusted code or attempts GPU-heavy reproduction

*Many papers run concurrently,
exposed failures, intermediate
artifacts kept for auditing*



01

Parsing

GROBID turns the PDF into structured TEI/XML: section-wise text plus figures, tables and captions.

GROBID → TEI/XML

03

Section embedding

Sections are split into overlapping spans — 2k characters, 20 % overlap — embedded and stored in a reusable vector index.

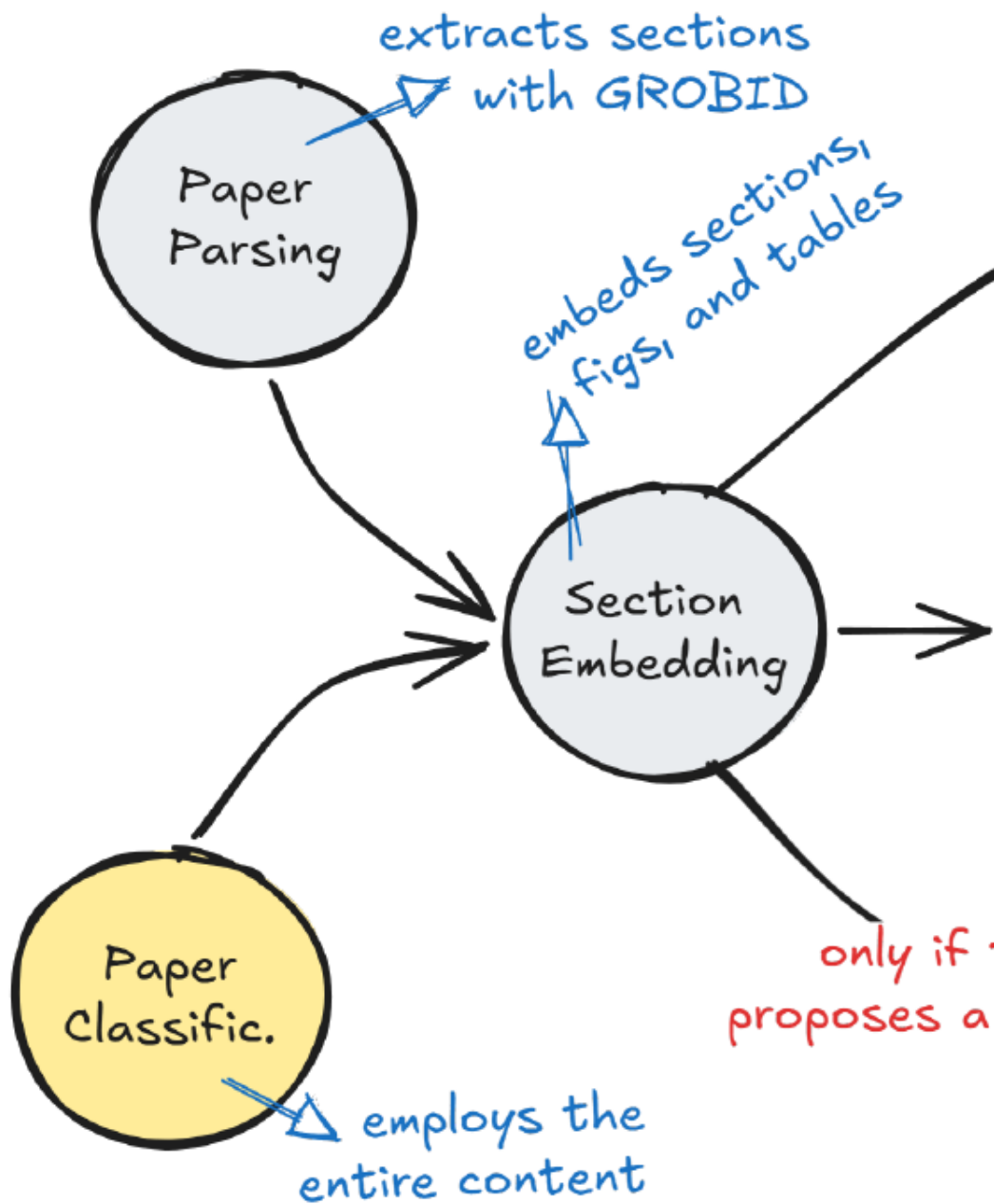
REUSABLE INDEX

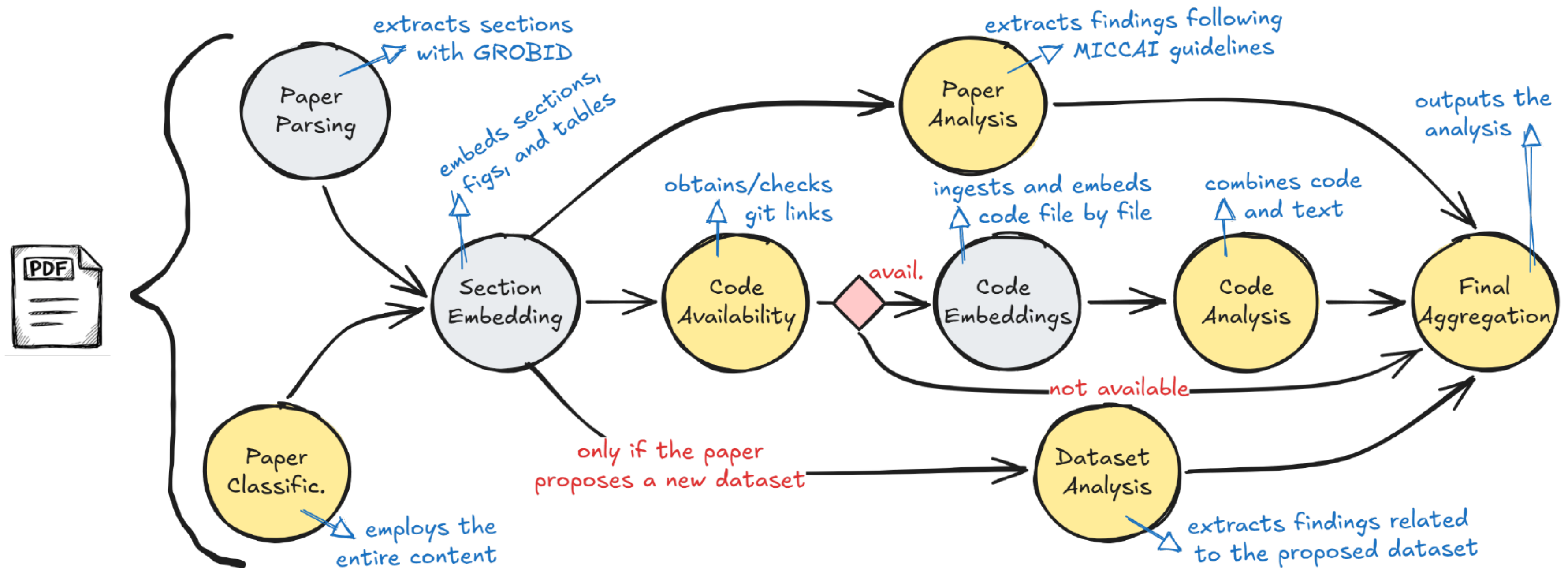
02

Classification

An LLM labels the paper dataset, method, both, or purely theoretical. The label gates which branches apply and sets the weights.

4 PAPER TYPES





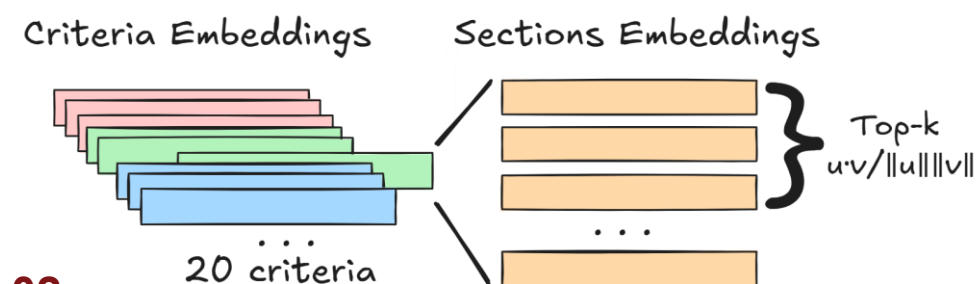
LLMs extract the evidence. The arithmetic stays deterministic.



01

Criteria

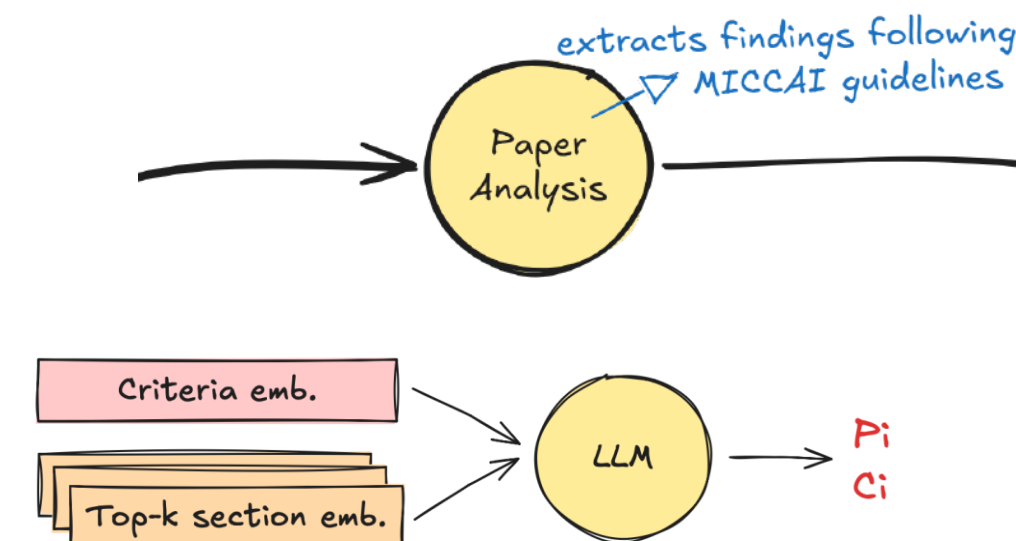
20 MICCAI criteria grouped by 3 categories: models, datasets, experiments. Each with a precomputed embedding.



02

Retrieval

For each criterion embedding, cosine similarity picks the top-k most similar section embeddings.



03

LLM check

One criterion at a time, LLM check its presence (p_i) in the section embeddings and the confidence (c_i).

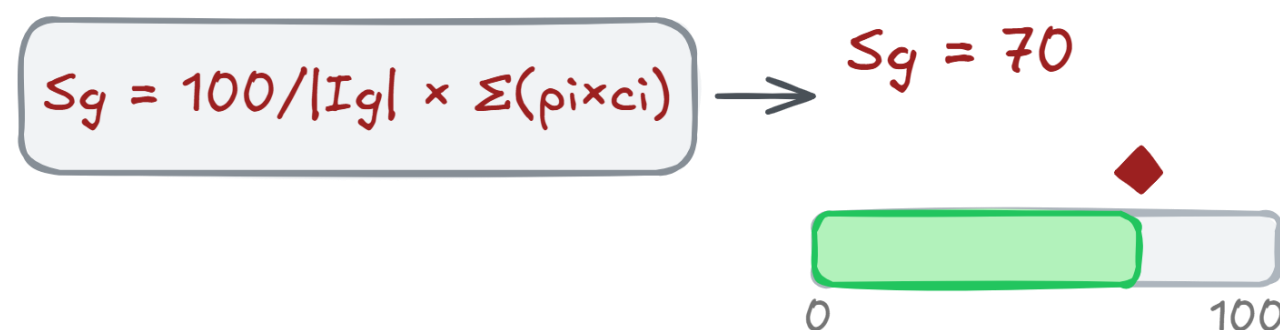
04

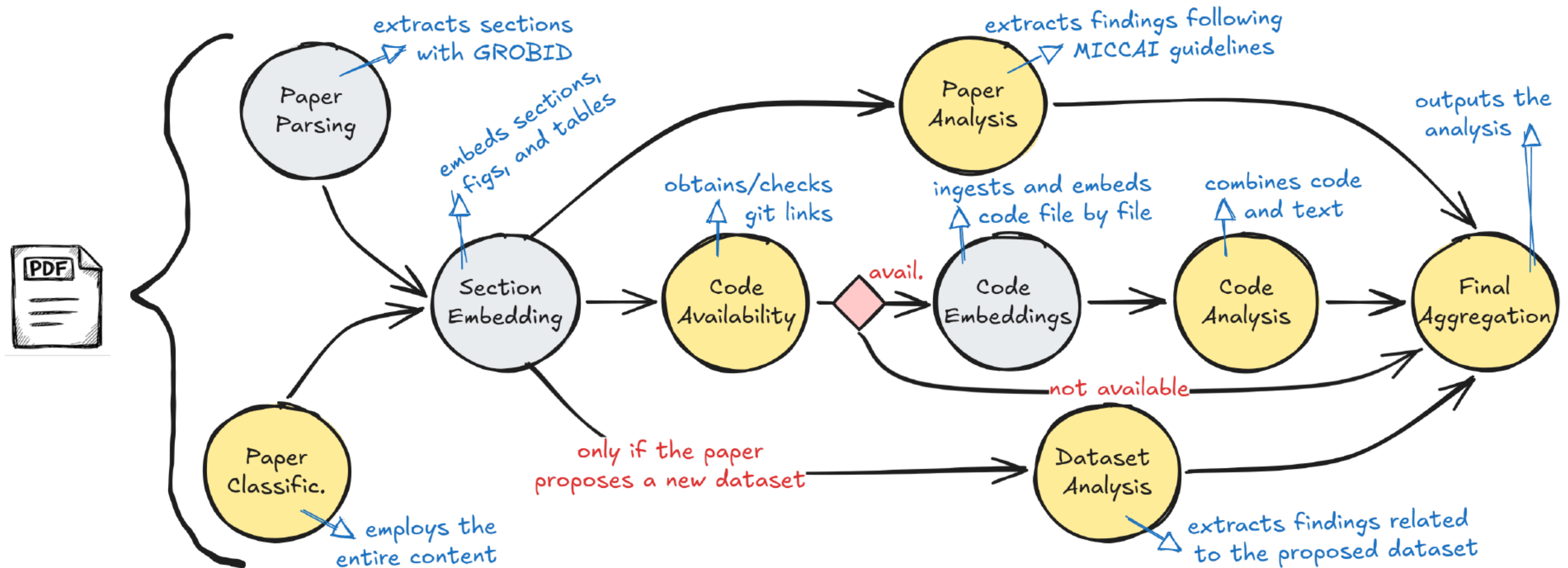
Paper analysis output

Structured JSON + Written Summary



Three scores, one for each category (S_g)





01

Code availability

Links are resolved from the manuscript and checked if they are reachable, non-empty, and not a placeholder.

02

Two-pass ingestion

Pass one reads the README and file tree; an LLM returns an ordered include list.

03

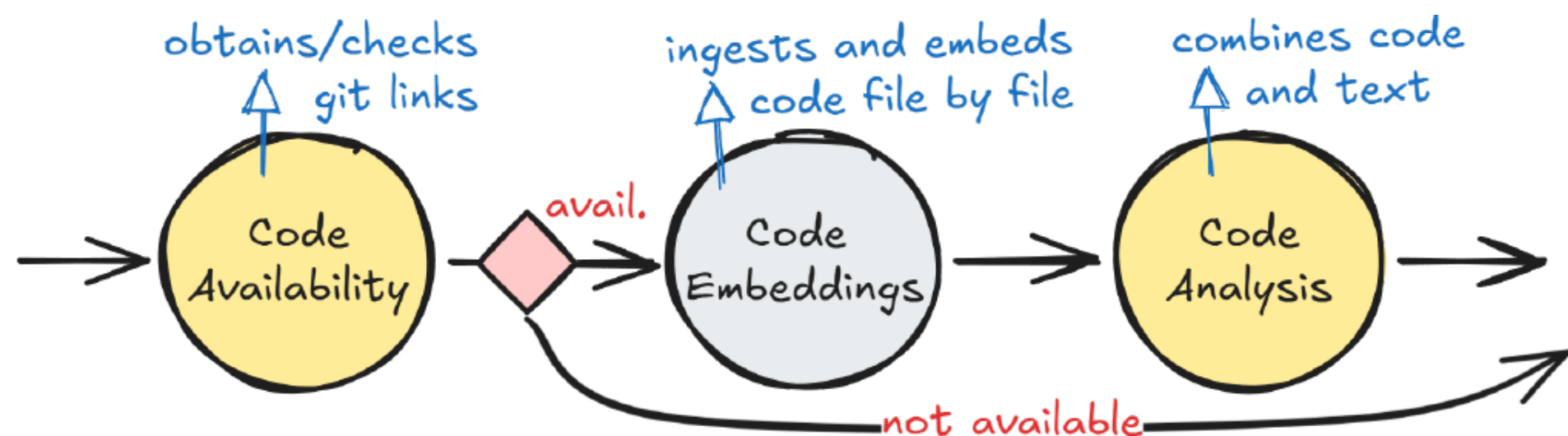
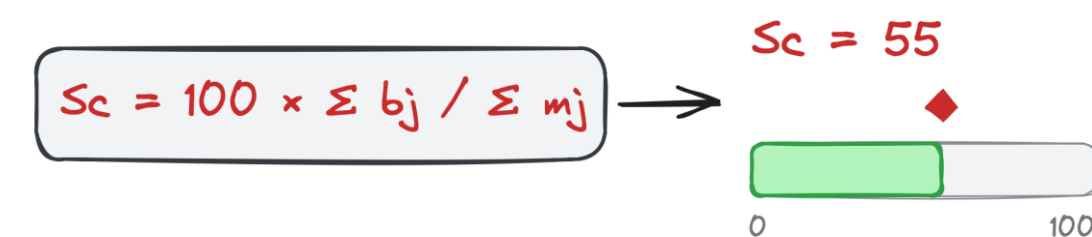
Chunk & fingerprint

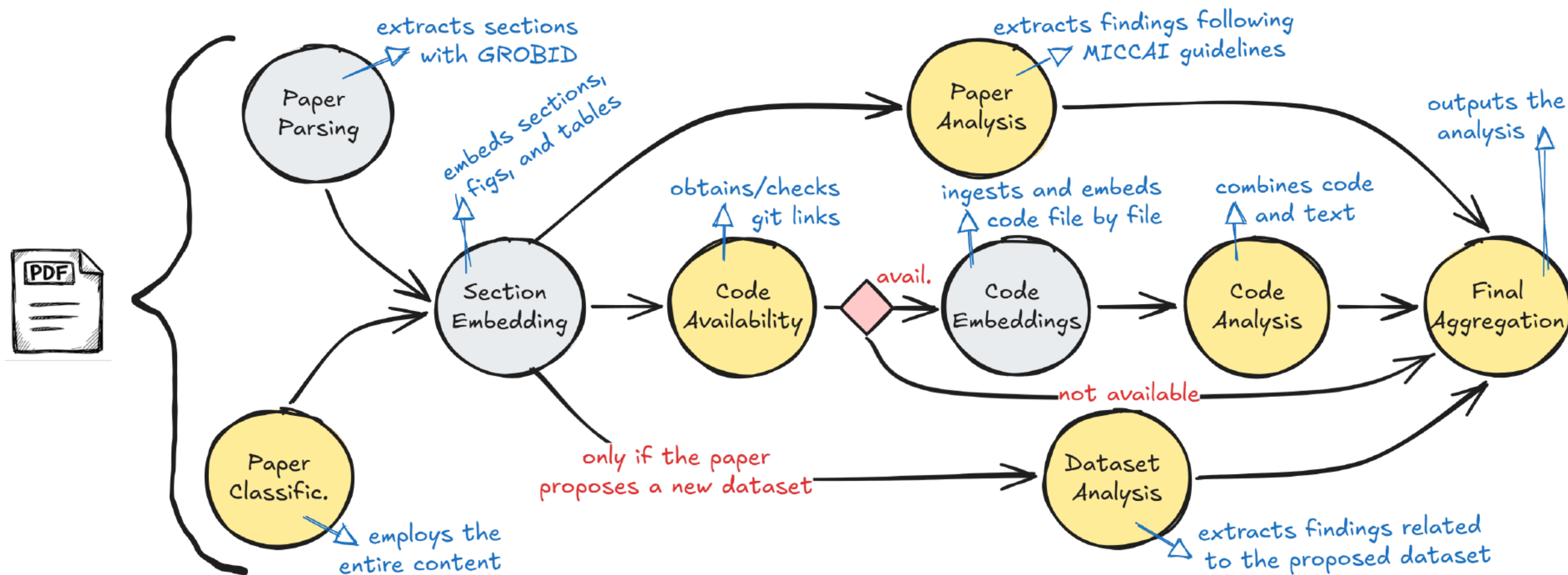
Files are chunked into ~20k-character blocks with path and index kept, and a per-file SHA-256 for cache safety.

04

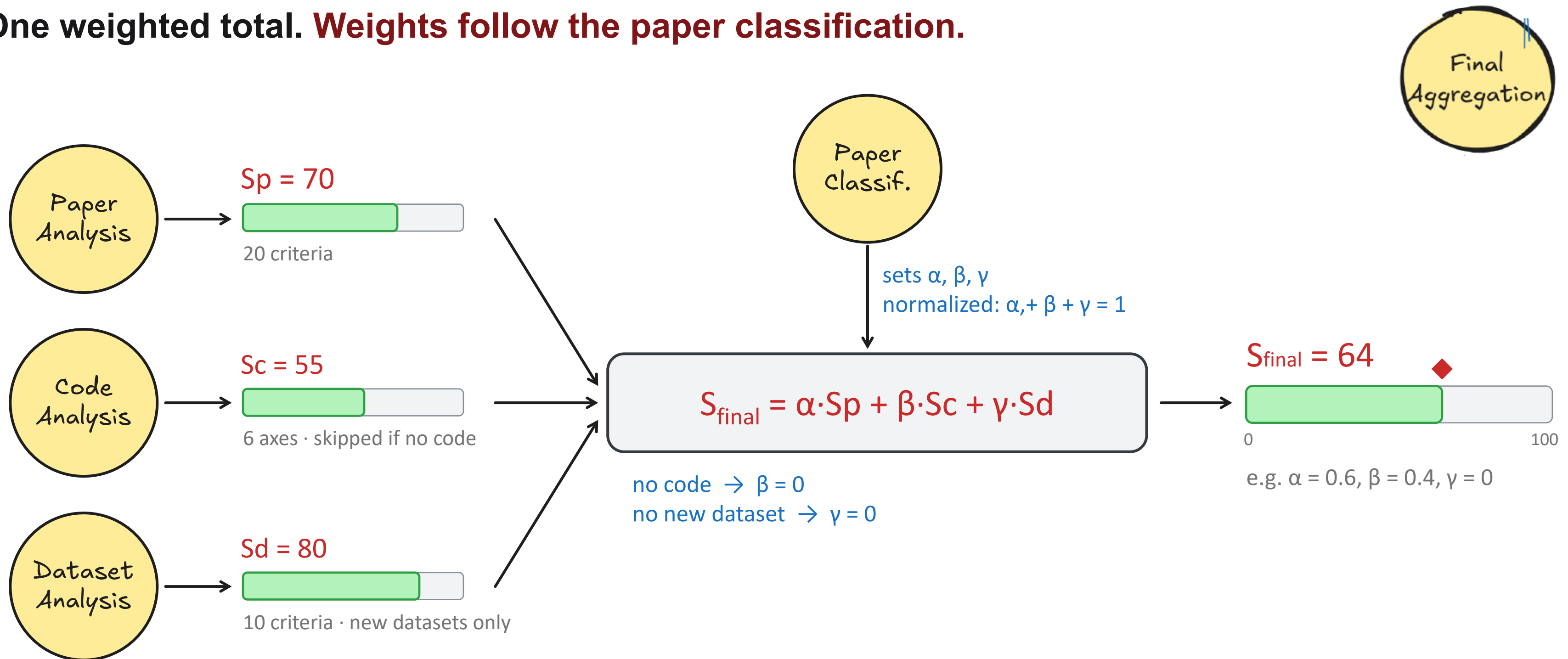
Code analysis output

Sum of six components: a superset of MICCAI code-related requirements.





One weighted total. **Weights follow the paper classification.**



Human – algorithm agreement validated on 100 papers from MICCAI 2025, randomly sampled, «stratified» over the three categories: dataset, method, both.

ANALYSIS	GPT-4o		GPT-5	
	AGREEMENT	MCC	AGREEMENT	MCC
Code	90.7%	0.81	84.8%	0.68
Paper	77.5%	0.57	66.3%	0.38
Dataset	72.2%	0.29	65.0%	0.18
Global	80.4%	0.60	70.9%	0.43

GLOBAL ACCURACY

+ 9.5

GPT-4o 80.4% vs GPT-5 70.9%

GLOBAL MCC

+ 0.17

GPT-4o 0.60 vs GPT-5 0.43

COST PER PAPER

– \$0.20

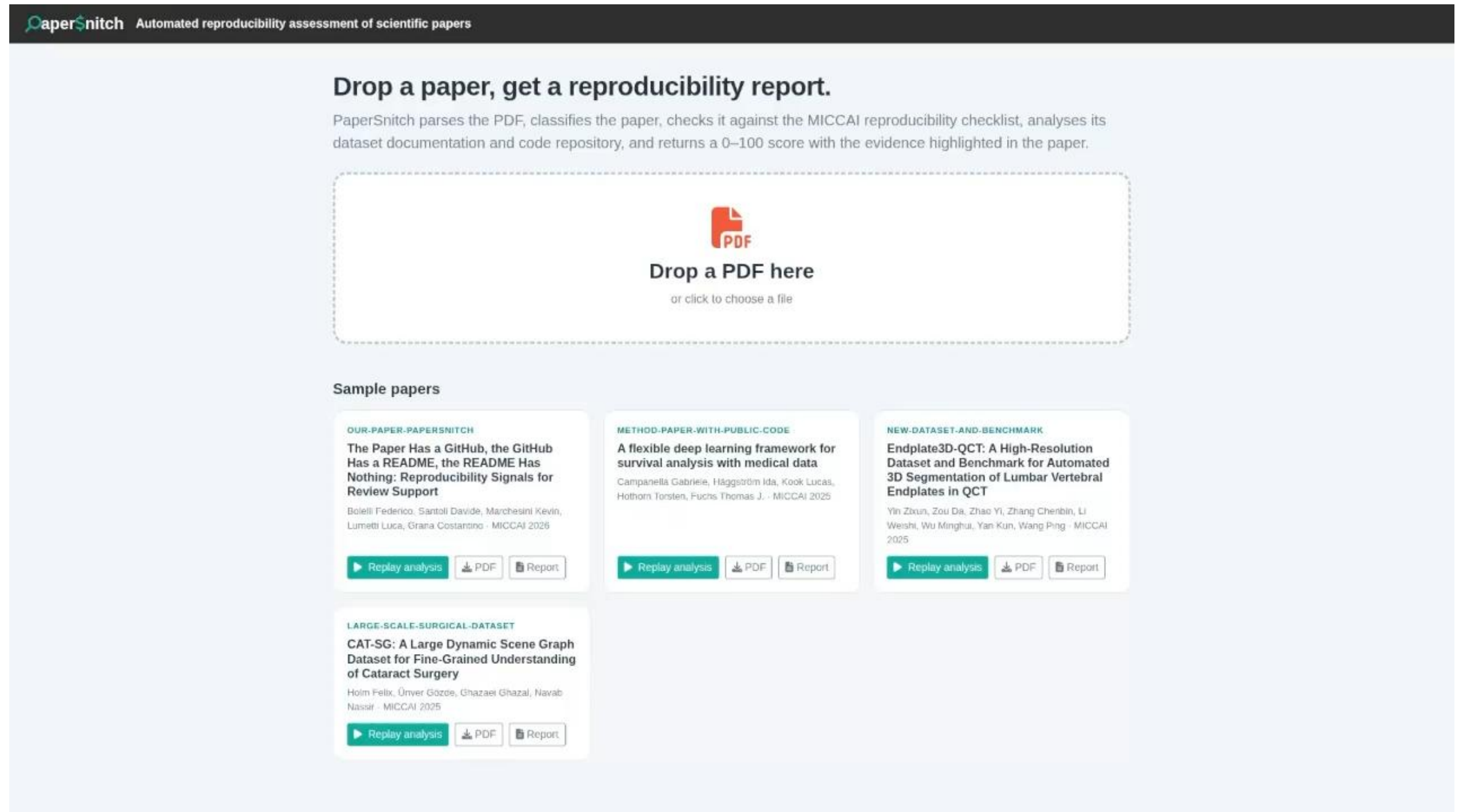
GPT-4o \$0.23 vs GPT-5 \$0.43

Evidence Found

- Evidence-grounded screening
- Supports reviewers, does not replace their judgment
- Tune weights and criteria per venue

Still Under Investigation

- Binary labels miss partial releases and restricted-access data
- Prompt injection (e.g., in the code or in the paper)
- LLM cost



The screenshot shows the PaperSnitch website, which provides automated reproducibility assessments for scientific papers. The header includes the logo and the tagline "Automated reproducibility assessment of scientific papers". The main content area features a large dashed box with a PDF icon and the text "Drop a PDF here" and "or click to choose a file". Below this, there is a section titled "Sample papers" displaying four paper cards. Each card includes a title, a brief description, the authors, and the venue. At the bottom of each card are three buttons: "Replay analysis", "PDF", and "Report".

PaperSnitch Automated reproducibility assessment of scientific papers

Drop a paper, get a reproducibility report.

PaperSnitch parses the PDF, classifies the paper, checks it against the MICCAI reproducibility checklist, analyses its dataset documentation and code repository, and returns a 0–100 score with the evidence highlighted in the paper.

Drop a PDF here
or click to choose a file

Sample papers

- OUR-PAPER-PAPERSNITCH**
The Paper Has a GitHub, the GitHub Has a README, the README Has Nothing: Reproducibility Signals for Review Support
Boilelli Federico, Santoli Davide, Marchesini Kevin, Lumetti Luca, Grana Costantino - MICCAI 2026
[Replay analysis] [PDF] [Report]
- METHOD-PAPER-WITH-PUBLIC-CODE**
A flexible deep learning framework for survival analysis with medical data
Campanella Gabriele, Häggström Ida, Kook Lucas, Hothorn Torsten, Fuchs Thomas J. - MICCAI 2025
[Replay analysis] [PDF] [Report]
- NEW-DATASET-AND-BENCHMARK**
Endplate3D-QCT: A High-Resolution Dataset and Benchmark for Automated 3D Segmentation of Lumbar Vertebral Endplates in QCT
Yin Zibun, Zou Da, Zhao Yi, Zhang Chenbin, Li Weishi, Wu Minghui, Yan Kun, Wang Ping - MICCAI 2025
[Replay analysis] [PDF] [Report]
- LARGE-SCALE-SURGICAL-DATASET**
CAT-SG: A Large Dynamic Scene Graph Dataset for Fine-Grained Understanding of Cataract Surgery
Holm Felix, Ünver Gözde, Ghazaei Ghazal, Navaab Nassir - MICCAI 2025
[Replay analysis] [PDF] [Report]

IAPR TECHNICAL COMMITTEE

TC-22

Reproducible Research in Pattern Recognition



MICCAI WORKSHOP

R2MI 2026

Reproducible Research in Medical Imaging



IAPR TECHNICAL COMMITTEE

TC-22

Reproducible Research in Pattern Recognition



MICCAI WORKSHOP

~~R2MI 2026~~ 2027?

Reproducible Research in Medical Imaging





UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Thank
you!

Reproducibility Signals for Review Support



*The paper has a GitHub, the GitHub has a README, **the README has nothing***



Federico Bolelli*



Davide Santoli*



Kevin Marchesini



Luca Lumetti



Costantino Grana

AlmageLab · University of Modena and Reggio Emilia, Italy

federico.bolelli@unimore.it

** Equal contribution*

P. Session 5
W-PM-022