

Messai.io

MICROBIAL ELECTROCHEMICAL SYSTEMS · AI PLATFORM

— WHITE PAPER

AI-assisted infrastructure for cross-study comparison, *reproducibility*, and prediction in microbial electrochemical systems

Companion to the EU-ISMET 2026 abstract — an open AI-powered platform for systematic parameter intelligence across 10,829 MES publications.

SAM FRONS · MESSAI.IO · SAM@MESSAI.IO

WHO THIS IS FOR

MES researchers — people who build reactors, grow biofilms, run polarization curves, and review each other's papers. It is **not** written for software engineers. Wherever a method comes from machine learning, this paper explains it in the language of the bench. The recurring question is practical: how can an electrochemist use AI to do better science, faster — and what does our platform hand you for free?

01 • EXECUTIVE SUMMARY

A field that has outgrown its ability to compare itself

The MES literature has passed 10,000 publications, but the field cannot quantitatively compare its own results. Reported power densities span five orders of magnitude (0.003 to 10,144,200 mW/m²; coefficient of variation ~1,285%), and most of that spread is not biology — it is missing metadata. Across the papers we scored, the average study reports fewer than **1 in 4** (23.1%) of the parameters another lab would need to reproduce or normalize it. Three of the most decision-relevant parameters — electrode spacing (~15% of papers), hydraulic retention time (~8%), and measurement method (~8%) — are almost never stated.

10,829

MES PUBLICATIONS INDEXED

~1,285%

POWER-DENSITY COV

23.1%

AVG REPORTING COMPLETENESS

687

ONTOLOGY SLOTS • UNIT-AWARE

MESSAI is a free, open platform that uses AI to turn that fragmented literature into a structured, queryable, provenance-tagged knowledge base. We extracted full text from 2,405 PDFs with an open OCR model, ran six domain-specific extractors over them, and mapped every value into a 687-slot, unit-aware parameter ontology (extraction F1 = 94.2%, Cohen's κ = 0.91, unit accuracy 99.7%). On top of that data sit cross-study comparison, reproducibility scoring, calibrated performance prediction, and a 100+ endpoint open API.

This paper describes how the data process actually works, what it reveals about the field, the 5-parameter minimum-reporting checklist we propose as a community standard, and — most importantly — **five concrete ways you can apply the same AI techniques in your own lab**. Everything is at messai.io and github.com/Messai-io.

02 • THE PROBLEM

A field that cannot compare itself

Imagine asking a simple question: "For acetate-fed MFCs with carbon-felt anodes, how does electrode spacing affect volumetric power density?" Today, that is a multi-month manual literature review, and at the end of it you still

cannot trust the comparison — because half the papers do not report electrode spacing, a third do not say whether power density was normalized to anode area or reactor volume, and many do not state whether the headline number is a polarization-curve maximum or a steady-state value.

This is not a knock on any individual author. It is a **systemic reporting problem** that compounds every year. Peer review accepts papers that omit the parameters needed to normalize them, so the comparability debt grows with the literature. The five-orders-of-magnitude spread in reported power density is the visible symptom; the cause is that the field has no shared, machine-readable vocabulary and no minimum-reporting floor.

The consequence is concrete and expensive: **we re-run experiments that the literature already answers, and we cannot tell redundant work from genuine knowledge gaps**. A graduate student spends eight months building a reactor whose performance envelope was, in effect, already bounded by twenty prior papers — if only those papers could be compared.

03 · THE VISION

AI as shared infrastructure, not a black box

Our thesis is simple. The literature we have all published is the substrate. What the field lacks is a shared **electrode** — a common piece of infrastructure that turns that substrate into something every lab can draw current from. AI makes that infrastructure buildable for the first time:

1. **Reading at scale.** Modern OCR and language models extract structured experimental parameters from thousands of full-text PDFs in days, not the person-years a manual meta-analysis would take.
2. **Normalizing meaning.** Language models map "anode surface area," "projected electrode area," and "geometric area of the working electrode" to a single ontology slot with explicit units — the step that makes cross-study math possible.
3. **Reasoning under uncertainty, honestly.** Bayesian and conformal methods produce a prediction *with a calibrated confidence interval* rather than a single number, so the tool tells you when it does not know.

We treat AI as an **assistant that shows its work** — every extracted value links back to the exact page, figure, and table it came from, carries a confidence score, and has a one-click "flag incorrect" button. Every prediction ships with units, a confidence interval, and a source. **No value in the system is unattributable, and no prediction is a bare point estimate**. That is the design principle that separates infrastructure a scientist can trust from a chatbot they cannot.

04 · HOW THE DATA PROCESS WORKS

Five stages, each with a method and a quality gate

4.1 · Corpus assembly

We index 10,829 MES publications (1932–2025): MFC 3,847; MES 4,842; BES 1,096; MEC 966; MDC 73. The broader crawl touches ~23,000 works; the 10,829 pass automated MES-relevance screening. The split is itself a finding: MES research is scattered across journals that do not self-identify as MES.

4.2 · Text extraction

Scientific PDFs are hostile to machines: two columns, equations, captions interleaved with body text, tables as images. We use Meta's Nougat, an open-source OCR model purpose-built for academic documents, which converts a PDF into structured markup (including LaTeX for equations). On a hand-checked sample ($n = 50$) it reaches ~98% accuracy. 2,405 full-text PDFs have been processed so far.

4.3 · Six domain extractors

Clean text is not structured data. A second stage runs six domain-specific extractors:

EXTRACTOR	WHAT IT PULLS
Reactor configuration	chamber type, working volume, membrane / separator
Electrode specifications	materials (17 anode + 10 cathode categories), area, surface modifications
Microbial community	species / consortium, inoculum source, biofilm properties
Operating conditions	temperature, pH, HRT, substrate, external resistance
Statistical measures	confidence intervals, p-values, replicate counts
Figure / table classification	which figures carry polarization data, which tables carry parameters

Each extractor is a **hybrid**: fast regex rules catch values stated in standard forms, and an LLM handles the messy prose — "the anode was a 2×3 cm piece of carbon felt" becomes `electrode_area = 6 cm2` with the right slot and unit. Rules for the easy 70%, AI for the hard 30%.

4.4 · The ontology

Every value is mapped into a **687-slot parameter ontology** (13 categories, 119 subcategories). It is **QUDT-compliant** — every slot has a defined physical quantity and unit, so the platform converts $\text{mW}/\text{cm}^2 \leftrightarrow \text{W}/\text{m}^3 \leftrightarrow \text{A}/\text{m}^2$ safely instead of comparing incompatible numbers. Quality against a human-annotated gold set: **F1 = 94.2%**, **Cohen's $\kappa = 0.91$** , **unit accuracy 99.7%**. Of the 687 slots, only a subset are populated — **the empty slots are not a bug; they are a map of the field's reporting gaps**.

4.5 · Quality control and provenance

- **Provenance on every value** — page, figure/table, model version, timestamp, and the verbatim source snippet are stored alongside the number.
- **Confidence scores** — low-confidence extractions are flagged, not silently promoted.
- **One-click correction** — a "flag incorrect" button routes a human correction back into the dataset.
- **Published integrity notes** — caveats and open statistical issues live in a public `SCIENTIFIC_INTEGRITY.md`. We would rather flag our own limitations than have you find them.

4.6 · Cross-study analytics

- **Correlation discovery** — the engine surfaces **183,615 candidate relationships**. Surfacing is not confirming: of 14 pairs tested rigorously (Student's t , Bonferroni $\alpha = 3.6 \times 10^{-3}$), **2 survived** ($r \approx 0.57$; $n = 36, 55$) and remain exploratory.
- **Reproducibility scoring** — every paper gets a completeness score against the reporting criteria (§6).
- **Contradiction detection** — surfacing studies that report conflicting parameter–performance relationships.

4.7 · Calibrated prediction

Built the way an electrochemist would want it — physics first, statistics second, honesty throughout:

- **Physics-aware, per-system-class models** routed by physics family (anodic oxidation, cathodic reduction, ion transport, selective reduction, sensing). An MEC is not scored with an MFC's formula.
- **Bayesian hierarchical priors** borrow strength across related systems (PyMC, NUTS), log-aware for the heavy-tailed quantities the field actually produces.
- **Calibrated uncertainty** via conformal prediction intervals with a coverage guarantee — a 90% interval contains the truth ~90% of the time on held-out data.
- Every numeric output ships `{value, unit, CI, confidence, source}`; when inputs are missing it reports `data_status` instead of inventing a number.

The point is not to reduce your reactor to one number — it is to **place candidate designs on a trade-off frontier** (energy vs. treatment efficiency) so you can narrow the design space *before* committing months of bench time.

05 · WHAT THE DATA REVEALS

The comparability problem is concentrated, not uniform

- **Average reporting completeness is 23.1%** — the typical paper reports fewer than one in four comparison-relevant parameters. But a single blended number is itself too coarse.
- **Reproducibility is not one number** — it depends on the experiment. A polarization curve, a continuous-flow treatment study, and a CO₂-electrosynthesis run each need *different* parameters, so we score each paper against the minimum-reporting set for its study type, crossed with system configuration. Coverage then ranges from **~12% to ~52%** across study types: **electrochemical characterization is the worst-covered** common study type (reference electrode and measurement method are almost never stated), while materials-screening and long-term-stability studies report more of their essentials.
- **A small set of criteria dominate the gap**: normalization basis, reference electrode, electrode area, and measurement method — each reported in only single-digit-to-low-double-digit percentages. Most strikingly, of the papers reporting a power/current-density value, the **area it is normalized to is almost never stated** (anode / cathode / membrane / projected). The problem isn't "report more parameters" — it's "report the **interpreters** of the number you already report."

- The 5-parameter checklist is associated with a >40% reduction in power-density IQR ($487 \rightarrow 278 \text{ mW/m}^2$) across a 63-paper retrospective subset ($p < 0.05$). This is **associational, not causal** — the clean test is a prospective intervention study (§6).

Absolute percentages from the current sample are extraction-limited lower bounds; the relative pattern across study types is the robust signal, and the full-corpus run firms up the absolutes. Method and the cross-tab are in `experiment-type-coverage.md`; live coverage is on `/tools/reproducibility`.

06 · THE STANDARD

The 5-parameter minimum-reporting checklist

A floor every MES paper can clear and every reviewer can check in under a minute:

1. **Electrode spacing**
2. **Electrode surface area** (with the normalization basis: which face, which projected area)
3. **External resistance**
4. **Measurement method** (polarization-curve maximum vs. steady-state)
5. **Unit normalization basis** (areal vs. volumetric; which area or volume)

Every MES paper already contains these in the lab notebook — they just do not all reach the methods section. The checklist is a live tool (`messa.io/tools/reproducibility`): paste a DOI or drop a PDF, get a score and the list of missing parameters, ready to paste into a review. The decisive experiment is a **prospective intervention study** — a journal special issue whose authors meet the checklist, compared against a control issue. That is a study an editor in the MES community can run, and we will be the scoring partner.

07 · FOR YOU

How you can use AI like this in your own research

You do not need to build a platform to get most of this leverage. Five patterns, each usable in your group this year:

1. **Triage the literature at scale.** AI-assisted semantic search + extraction pulls the 30 papers that share your reactor type and substrate, with parameters already in a table. Months $\rightarrow$ an afternoon.
2. **Structure your own backlog.** Old papers, theses, and notebooks are unstructured data. Rules for the standard 70%, an LLM for the messy 30%, a human check on low-confidence values — a decade of your work becomes queryable.
3. **Self-score before you submit.** Run your manuscript through a reproducibility check. It costs a minute and removes the most common reviewer complaint.
4. **Predict before you build.** Use calibrated priors to bound a proposed design's performance envelope *before* machining the reactor. Insist on a prediction that comes with a calibrated interval and a stated source.
5. **Make your data FAIR and shared.** The more the community standardizes reporting, the more every downstream AI tool — yours included — improves.

The meta-lesson: **demand provenance and calibrated uncertainty from any AI you use**. A tool that cannot show you where a number came from, or admit when it does not know, has no place in a methods section.

08 · THE PLATFORM

What's live, free, and open

CAPABILITY	STATUS	WHERE
Semantic literature search across 10,829 papers	Live	<code>messai.io/research</code>
Per-paper extracted parameters + provenance + flag-incorrect	Live	paper detail pages
Reproducibility scoring (5-param + comprehensive)	Live	<code>/tools/reproducibility</code>
Open REST API (100+ endpoints)	Live	<code>messai.io / docs</code>
Cross-study compare	In progress	<code>/research/compare</code>
Calibrated performance prediction (per-class)	In progress	<code>/lab</code> , <code>/api/ml/predict</code>
Multi-objective (Pareto) optimization	In progress / demo	<code>/lab</code>

Everything is free for the MES community and open-source at github.com/Messai-io. We label what is live versus in progress deliberately — overclaiming is the fastest way to lose a rigorous audience's trust.

Four ways to plug in

1. **Verify your own paper** — paste a DOI at `messai.io` , check our extraction, flag anything wrong (~30 seconds).
2. **Claim an ontology category** — own the slots your group knows best (anode materials, biofilm properties...) via a GitHub issue.
3. **Pilot the 5-parameter checklist** in your reviewing/editing — we provide the copy-paste reviewer comment and the scoring tool.
4. **Run the intervention study** — a pre-registered special issue, with us as scoring partner.

09 · INTEGRITY

Limitations

In the spirit of the published `SCIENTIFIC_INTEGRITY.md` :

- The corpus split (verified-MES vs. tangential) relies on automated screening and is imperfect.
- Many ontology slots are unpopulated; coverage is a work in progress, not a finished map.
- The two surviving correlations are exploratory (small n, uncontrolled confounds), not confirmatory.

- The checklist↔IQR result is retrospective and observational; causality requires the prospective study.
- Prediction is honest about its inputs: when it lacks them, it reports `data_status` rather than guessing.

We would rather you trust a smaller set of well-qualified claims than a larger set of fragile ones.

10 · REFERENCES & LINKS

References

Platform: messai.io · Code / data: github.com/Messai-io · Contact: sam@messai.io

- Logan, B.E., et al. (2006). Microbial fuel cells: methodology and technology. *Environ. Sci. Technol.* 40, 5181–92.
- Schröder, U. (2007). Anodic electron transfer mechanisms in microbial fuel cells. *Phys. Chem. Chem. Phys.* 9, 2619–29.
- Wilkinson, M.D., et al. (2016). The FAIR Guiding Principles. *Sci. Data* 3, 160018.
- Logan, B.E., et al. (2019). Electroactive microorganisms in bioelectrochemical systems. *Nat. Rev. Microbiol.* 17, 307–319.