Automated Generation of Computational Models from Scientific Publications Using AI

Jean-Baptiste Gourlet[1]*, Elodie Prevot[1], Jeremy Villard[1], **Frederic Cogny**[1]
* Corresponding authors, [1] Nova In Silico, Lyon, France

Using **AI**, **publications** with equations are turned into ready-to-run computational **models**, reducing manual work and improving **reproducibility**

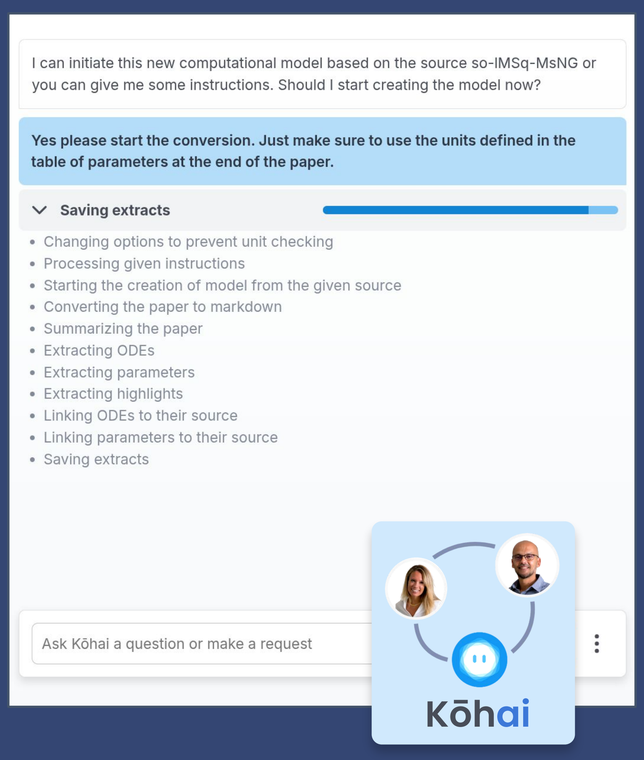

Fig. 1: Kōhai AI chat interface for model extraction.
Modelers interact with Kōhai via a chat interface, providing instructions executed step by step with full visibility. The agent flags inconsistencies and offers guidance, ensuring a transparent, reliable workflow

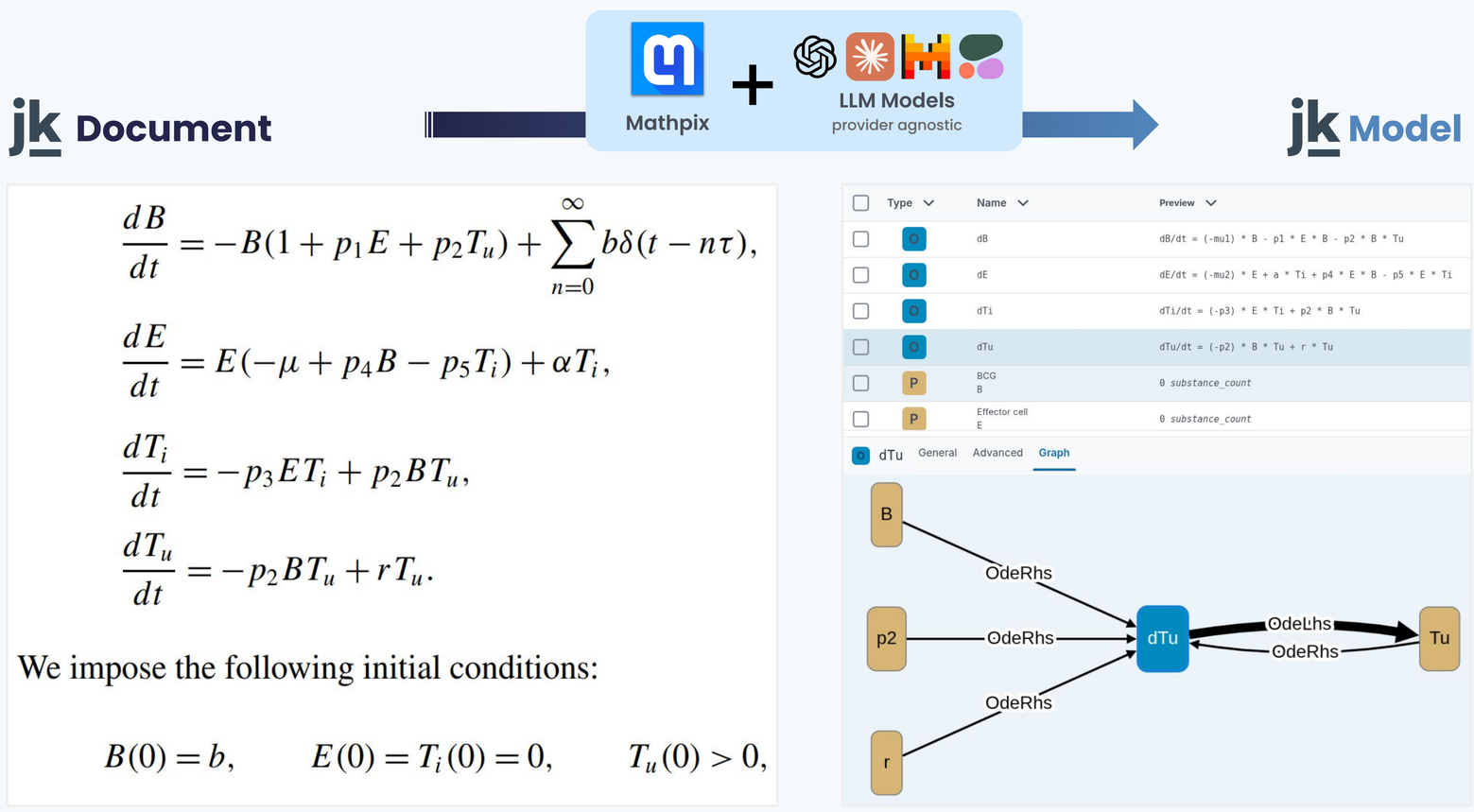

Fig. 2: Extraction sample of model components in Jinkō from Bunimovich-Mendrazitsky et al., 2008 [1].
- Mathpix converts the document into LLM-readable scientific text.
- Model is automatically populated in Jinkō.
- Model is available for simulation or export.

BACKGROUND

Reproducing models from scientific publications is time-consuming and error-prone. ODE-based models are often reported as equations and tables in LaTeX or PDFs, and translating them into executable form can take several days. Advances in AI, particularly Optical Character Recognition (OCR) and Large Language Models (LLMs), create new opportunities to automate this process. By integrating these technologies into the Jinkō platform, we aim to accelerate model reconstruction and make published work readily usable.

METHODS

- **AI Agent Integration**: Utilizing an AI agent within the jinkō modeling and simulation platform to process scientific publications containing model descriptions, including LaTeX-formatted equations and parameter data.
- **OCR and Data Extraction**: Employing the Mathpix[2] API to perform OCR on PDFs, extracting LaTeX equations and parameter tables.
- **Pre-Instruction Input**: Users can provide specific instructions or context before the LLM processes the extracted content, guiding the model to focus on particular aspects of the publication or emphasize certain relationships.
- **Iterative Model Construction**: Leveraging LLMs, specifically GPT-4o-mini[3], to interact with jinkō's API, iteratively building and validating the computational model through function calls and embedded tools. The current focus is on ODE-based models.
- **Unit Analysis and Annotation**: Extracting unit information from source papers, converting it into a standardized SI-compliant format for automated unit checking within the modeling platform, and annotating equations with links to their specific locations in the source document for traceability.

RESULTS

The implementation of this workflow has dramatically reduced the effort needed to translate models from the literature into executable form. Reconstructions that once required several days of focused work can now be completed within a few hours, with unit adjustments remaining the only significant manual step due to inconsistencies in source publications.
Since rolling out the “paper To Model” feature, more than 50% of the models imported in our model library have been generated with the help of Kōhai with a **time savings of roughly 60%** versus building the model by hand. Internal statistics indicate that in some cases the **reduction in effort can reach up to 80%**.
Beyond efficiency, users also report increased confidence in the accuracy and reproducibility of reconstructed models. Features that trace each equation back to its original location in the source paper greatly facilitate verification and validation, while export options to standard formats such as SBML and Julia extend the applicability of the models beyond Jinkō, enabling seamless integration into external workflows and fostering broader reuse within the scientific community.

CONCLUSION

- ➔ By combining OCR and LLMs within the Jinkō platform, we have established a workflow that enables the automated generation of computational models directly from scientific publications. This approach improves reproducibility, reduces the burden of manual implementation, and accelerates the use of existing models in new research contexts.
- ➔ Ongoing development focuses on strengthening the interaction between the AI agent and human modelers, including diagrams and chart extraction, with the goal of further improving precision and reliability in extracting model details from the literature.

REFERENCES

[1] Bunimovich-Mendrazitsky, S., Byrne, H., & Stone, L. (2008). Mathematical Model of Pulsed Immunotherapy for Superficial Bladder Cancer. Bulletin of Mathematical Biology, 70(7). https://doi.org/10.1007/s11538-008-9344-z
[2] Mathpix. Snip: Convert PDFs, images, and handwritten notes to LaTeX, Markdown, and more. Available at: https://mathpix.com (accessed September 2025).
[3]OpenAI. Chat Completions API (GPT-4). Available at: https://platform.openai.com (accessed September 2025).