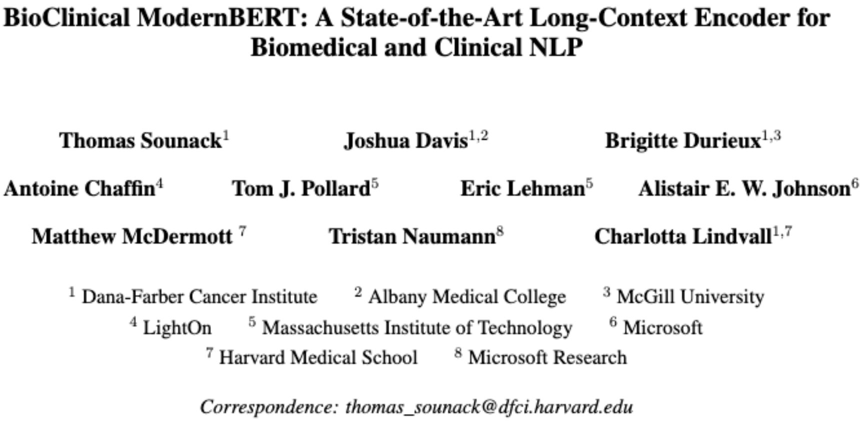
Press Release
Paris, June 13, 2025

Announcing BioClinical ModernBERT: a new state-of-the-art encoder model for Medical NLP

The recent release of ModernBERT by LightOn and AnswerAI aims at providing the best base model that can be then used in different industry verticals.

Today, Thomas Sounack from **the Dana-Farber Cancer Institute** in collaboration with researchers at **LightOn**, **Harvard University**, **MIT**, **McGill University**, **Albany Medical College** and **Microsoft Research**, used this capability and trained a new State-Of-The-Art (SOTA) medical encoder named **BioClinical ModernBERT**.



BioClinical ModernBERT: A State-of-the-Art Long-Context Encoder for Biomedical and Clinical NLP

Thomas Sounack[1] **Joshua Davis**[1,2] **Brigitte Durieux**[1,3]
Antoine Chaffin[4] **Tom J. Pollard**[5] **Eric Lehman**[5] **Alistair E. W. Johnson**[6]
Matthew McDermott [7] **Tristan Naumann**[8] **Charlotta Lindvall**[1,7]

[1] Dana-Farber Cancer Institute [2] Albany Medical College [3] McGill University
[4] LightOn [5] Massachusetts Institute of Technology [6] Microsoft
[7] Harvard Medical School [8] Microsoft Research

Correspondence: thomas_sounack@dfci.harvard.edu

[BioClinical ModernBERT: A State-of-the-Art Long-Context Encoder for Biomedical and Clinical NLP]

Efficient Continued Pre-Training, Streamlined for Medicine

One critical but lesser-known scheduling feature of ModernBERT allows researchers a seamless continued pre-training while eliminating cold restarts. Stable-phase checkpoints and a decay phase contribute to having models that can efficiently converge on specific domains.

Leveraging this scheduling feature, Thomas Sounack continued the pre-training of ModernBERT on an extensive collection of medical texts. The result: BioClinical ModernBERT, a new model that outperforms all existing encoders on medical classification and Named Entity Recognition (NER) tasks, setting a new SOTA benchmark for medical NLP applications.

Optimized for the Realities of Clinical Context

Real-world medical texts can be very long; they span full clinical notes and as well as large reports. BioClinical ModernBERT's ModernBERT backbone provides long-context document support, with hybrid attention and unpadding mechanisms for rapid processing, crucial for healthcare and clinical workflows.

A Recipe for Continued Pre-Training

Beyond the model itself, this experience refines continued pre-training for domain adaptation. This approach is reproducible: BioClinical ModernBERT demonstrates robust transfer to new domains, opening the door for anyone seeking to tailor ModernBERT for their own specialized data.

Try It Out

Interested in leveraging continued pre-training or ModernBERT's long-context expertise in a different domain? Explore the BioClinical ModernBERT collection and see how it can advance specialized NLP tasks in your domain.

About LightOn

Founded in 2016 in Paris and the first European generative AI company listed on Euronext Growth, LightOn is a pioneering player in the field of sovereign GenAI. Its Paradigm platform enables organizations to deploy large-scale AI while ensuring the confidentiality of their data. LightOn's technology ensures essential strategic independence by offering tailored solutions. This technological mastery is accompanied by the ability to process large volumes of data for industrial uses, with applications in various sectors such as finance, industry, health, defense, and public services.

LightOn is listed on Euronext Growth® Paris (ISIN: FR0013230950, ticker: ALTAI-FR). The company qualifies for PEA and PEA PME investment plans and is recognized as an "Innovative Company" by Bpifrance. To learn more: https://www.lighton.ai

Contacts

LIGHTON invest@lighton.ai	**SEITOSEI●ACTIFIN** **Investor Relations** Benjamin LEHARI lighton@seitosei-actifin.com
KALAMARI **Media Relations** Camille Bernisson - +33 7 64 44 14 49 Maroua Derdega - +33 7 63 77 73 20 lighton@kalamari.agency	**SEITOSEI●ACTIFIN** **Financial Media Relations** Jennifer JULLIA - +33 6 47 97 54 87 jennifer.jullia@seitosei-actifin.com