

LLM やエージェントの基盤技術を習得！モダン Transformer の核心に迫る
『モダン Transformer モデルの内部構造から、
マルチモーダル/RAG/SLM の実装と最適化まで』を7月2日に発売
発売を記念して第2章を2週間、無料公開

インプレスグループで IT・デザイン関連メディア事業を展開する株式会社インプレス（本社：東京都千代田区、代表取締役社長：高橋隆志）は、最新技術による AI モデルの内部構造と実践技術を網羅した書籍『モダン Transformer モデルの内部構造から、マルチモーダル/RAG/SLM の実装と最適化まで』を2026年7月2日（木）に発売いたします。



■AIの進化を支える Transformer の内部構造をひも解く

ChatGPTや Geminiをはじめとする生成 AI の急速な普及に伴い、その基盤技術である「Transformer」への深い理解は、エンジニアやデータサイエンティストにとって極めて重要になっています。特に現在は、言語を含むマルチモーダルに対応した次世代モデルが続々と登場し、モデルの小規模化（SLM）や高度なエージェント化が加速する大きな転換期を迎えています。しかし、提供されるモデルを表面上利用するだけでは、高度なカスタマイズや現場の課題解決には限界があります。本書は現在、求められているモダン Transformer モデルの本質を、基礎から応用まで体系的に学べるタイムリーな一冊です。

■数学的基盤から現場レベルの実装までを一気通貫

本書の大きな特徴は、Transformer の構造を詳細に解説しつつ、Hugging Face や vLLM といった最新ライブラリを用いた実践的なコード例を豊富に掲載している点です。デコーダー/エンコーダーのみの設計や Mixture-of-Experts (MoE) などのアーキテクチャのバリエーションに加え、テキスト生成戦略（思考の木など）、RAG（検索拡張生成）、ファインチューニング（QLoRA など）、推論の最適化（FlashAttention など）までを網羅。実際に手を動かして動作を確認しながら学べるため、実務につなが

■本書の第2章を2週間限定で無料公開

書籍の発売を記念して、第2章の内容を2週間限定で無料公開いたします。それらの内容は、弊社開発の「インプレス Web ブックビューアー」を使って、登録不要かつ Web ブラウザで紙面を閲覧できるので、場所や時間を選ばずに PC やスマートフォンなどから利用できます。使用するビューアーは SNS で書籍の情報を共有したり、書籍の購入や詳細を確認したりできる機能も備えています。

- ・『モダン Transformer モデルの内部構造から、マルチモーダル/RAG/SLM の実装と最適化まで』

無料公開へのリンクを掲載したページ：<https://book.impress.co.jp/books/1125101146>

【公開期間：2026年7月2日（木）から7月15日（水）まで】

■本書の構成

◆第1部 現代の Transformer モデルの基礎

第1章 Transformer はなぜ必要か

第2章 詳解 Transformer

◆第2部 生成系 Transformer

第3章 モデルファミリーとアーキテクチャのバリエーション

第4章 テキスト生成戦略とプロンプティングテクニック

第5章 選好アライメントと RAG

◆第3部 専門化されたモデル

第6章 マルチモーダルモデル

第7章 効率的な特化型 SLM

第8章 LLM の訓練と評価

第9章 LLM の最適化とスケールリング

第10章 倫理的かつ責任ある LLM

■書誌情報



書名：モダン Transformer モデルの内部構造から、マルチモーダル/RAG/SLM の実装と最適化まで

シリーズ名：impress top gear

著者：Nicole Koenigstein（著）、株式会社クイープ（訳）

発売日：2026年7月2日（木）

ページ数：312 ページ

サイズ：B5 変型判

定価：3,630 円（本体 3,300 円＋税 10%）

電子版価格：3,630 円（本体 3,300 円＋税 10%）※インプレス直販価格

ISBN：978-4-295-02446-0

◇Amazon の書籍情報ページ：

<https://www.amazon.co.jp/dp/4295024465/>

◇インプレスの書籍情報ページ：

<https://book.impress.co.jp/books/1125101146>

■著者プロフィール

Nicole Koenigstein（ニコール ケーニヒシュタイン）

著名なデータサイエンティスト兼定量分析研究者。資産運用会社ワイデン・キャピタルのチーフデータサイエンティスト兼 AI・定量分析部門責任者を務めている。LLM 研究への資金提供に関して欧州委員会

(European Commission) の外部専門家を務めるほか、規制産業における生成 AI の導入について助言を行うため、証券監督者国際機構 (IOSCO) の外部専門家を務める。著書に『Math for Machine Learning』 (Manning Publications) がある。

■翻訳者プロフィール

株式会社クイープ

コンピュータシステムの開発、ローカライズ、コンサルティングを手がけている。最近の訳書に『Python ではじめるクリーンアーキテクチャ SOLID 原則 / ドメイン駆動設計 / テスト駆動開発を実践』がある。そのほか『LLM 本番システム構築ノウハウ 基礎から実装・運用の方法、アプリ構築の実例まで』『Python ライブラリによる因果推論・因果探索 [概念と実践] 因果機械学習の鍵を解く』などを翻訳 (いずれもインプレス発行)。

■impress top gear シリーズについて



impress top gear シリーズは、21 世紀の IT 時代に求められる新たなトレンドを取り上げることに加えて、これまでの技術・知識を時代に合わせた形で提供する技術解説書です。読者が IT のフィールドで「トップギア」へとシフトチェンジできるような技術・知識を書籍として提供していきます。

以上

【株式会社インプレス】 <https://www.impress.co.jp/>

シリーズ累計 8,000 万部突破のパソコン解説書「できる」シリーズ、デザイン・グラフィック関連の書籍ブランド「MdN」、「デジタルカメラマガジン」等の定期雑誌、IT 関連の専門メディアとして国内最大級のアクセスを誇るデジタル総合ニュースサービス「Impress Watch シリーズ」等のコンシューマ向けメディア、「IT Leaders」をはじめとする企業向け IT 関連メディアなどを総合的に展開・運営する事業会社です。IT・デザイン関連出版メディア事業、およびデジタルメディア&サービス事業を幅広く展開しています。

【インプレスグループ】 <https://www.impressholdings.com/>

株式会社インプレスホールディングス (本社：東京都千代田区、代表取締役：塚本由紀) を持株会社とするメディアグループ。「IT・デザイン」「音楽」「山岳・自然」「航空・鉄道」「モバイルサービス」「学術・理工学」を主要テーマに専門性の高いメディア&サービスおよびソリューション事業を展開しています。さらに、コンテンツビジネスのプラットフォーム開発・運営も手がけています。

【本件に関するお問合せ先】

株式会社インプレス 広報担当：丸山

E-mail: pr-info@impress.co.jp URL: <https://www.impress.co.jp/>