

# ワレン・レックス (Lexington Whalen)

[in linkedin.com/in/lxaw](https://www.linkedin.com/in/lxaw)

## 概要

- **機械学習研究リーダーシップ**: NVIDIAにてフラッグシップとなる拡散言語モデルの開発を共同リード。1Bから14Bパラメータ規模のモデルに対する新しい学習・推論インフラストラクチャを設計。チームは2名から複数チームへと拡大。成果はCEOジェンスン・ファン氏に直接報告され、学習・推論手法は社内5つ以上の研究チームで採用。従来の主要モデルと比較して速度約10倍、精度約10%の向上を達成。
- **研究インパクト**: 博士課程・修士課程の学生からなる学際的チームをリードし、モデル品質を維持しつつ学習効率を5倍向上させる先駆的な拡散モデル研究を実施。トップレベルのコンピュータビジョン国際会議で論文を発表。研究手法はNVIDIAの社内開発に採用。状態空間モデルおよびTransformerアーキテクチャ双方における学習・推論最適化の専門性を実証。
- **機械学習エンジニアリング**: NVIDIAの社内拡散言語モデルプロジェクト向けに包括的な評価・学習パイプラインを開発。数百万から数十億パラメータ、数十億から数兆トークンに及ぶモデル学習をスケール。高度な並列化戦略を活用し、数千基のGPUを用いた分散学習システムの経験を有する。
- **ヘルスケアAI**: 米国ARPA-HのTHORヘルスケアイニシアチブのテクニカルリードとして、がん検知センサーの劣化を70%以上低減する革新的な小型PyTorchモデルを開発。カーネギーメロン大学、スタンフォード大学、パデュー大学、ジョージア工科大学等のトップ研究機関の研究者と協働。
- **受賞歴**: NSF大学院研究フェロースHIP受賞 (150,000ドル相当)、サウスカロライナ大学トップスカラー賞 (100,000ドル以上相当)、米商務省日本語クリティカル・ランゲージ・スカラー (5,000名以上の応募者から500名のみ選出)。
- **技術開発**: Django/React/SQLを用いた研究用フルスタックWebサイトの開発者。数千件の参加者エントリーを管理するデータベースを運用。総額20,000ドル以上の助成金を獲得し、開発チームをリード。

## 職歴

### ソフトバンク – SB Intuitions

東京

R&Dエンジニア、リーダーシップ&プロフェッショナルトラック (L/Pグレード3)

2026年4月 –

- ソフトバンクのSB Intuitions R&D部門にて、大規模言語モデルの研究開発をリード。技術戦略およびプロジェクト方針に貢献。
- 2026年入社におけるソフトバンクグループ代表に選出。ソフトバンクグループ各社の新入社員を代表。

### NVIDIA

アトランタ, GA & サンタクララ, CA

効率的深層学習インターン

2025年3月 – 2026年3月

- NV Research 1-14Bフラッグシップ拡散言語モデルシリーズのベース・指示・RLバリエーションの学習、ファインチューニング、評価を実施。評価・推論インフラストラクチャを構築し、ビジョン言語モデル、エンベディング、潜在推論に関する5つ以上の社内研究チームに採用。社内ダウンロード数100万以上。
- モデル蒸留、効率的学習レシピ、効率的推論レシピに従事。
- SLURMで管理されたサーバー上の数千基のNVIDIA GPUクラスターを活用した高性能コンピューティングインフラストラクチャ上での拡散モデルに特化。
- 20名以上のシニア研究者およびクロスファンクショナルエンジニアリングチームに対し、毎週の進捗報告および技術的知見を発表。
- Qwen3自己回帰シリーズを精度・速度の両面で上回るEfficientDLMシリーズの拡散言語モデルを開発。社内ダウンロード数100万以上。
- ビジョンチームと協力し、Nemotronシリーズの拡散言語モデルをリリース。2026年3月公開予

定。

## NVIDIA

データフィルタリングチャレンジリード

アトランタ, GA

2024年12月 - 2025年12月

- Google Analyticsを活用してエンゲージメントを追跡し、30以上の大学研究チームおよび企業におけるユーザー体験を最適化するチャレンジWebサイトの開発・管理をリード。
- 参加者向けに初期4億パラメータのベースラインモデルのプロファイリングを支援し、提出物の品質評価のための標準化ベンチマークを作成。
- 100億トークンのファインチューニングデータセットを活用し、モデル提出物の一貫した公正な評価を保証する包括的な評価フレームワークを設計・実装。
- 世界トップ3の機械学習国際会議であるICML（参加者約9,000名、論文採択率25%未満）にてチャレンジをプロモーション。
- Lambda LabsやTuring等のスポンサー企業と連携しGPUリソースを確保。チーム提出物の評価およびコンペティション全体のワークフローを管理。

## ジョージア工科大学

機械学習エンジニア

アトランタ, GA

2024年8月 - 2025年12月

- 博士課程・修士課程の3名のチームをリードし、生成品質を維持しつつ拡散モデルの学習時間を平均最大5倍短縮する革新的手法を設計。手法はCVPR 2025に採択。
- 米国ARPA-HのTHORヘルスケアイニシアチブのアルゴリズムリード。がん検知センサーの劣化を70%以上低減し、センサー寿命を数時間から数日へと10倍以上改善する取り組みを主導。
- 人体のエネルギーで駆動可能なリソース効率の高い機械学習アーキテクチャを設計。
- スタンフォード大学、ノースウェスタン大学、カーネギーメロン大学、MIT等のトップ大学の5つ以上のチームと協働し、ジョージア工科大学を代表。
- 連邦スポンサーおよびヘルスケア関係者への技術的進捗を発表するため全国を巡回し、多様な聴衆に対して複雑な技術概念を効果的に伝達。

## サウスカロライナ大学

ソフトウェアデベロッパー

コロンビア, SC

2021年1月 - 2024年8月

- PHP、Django、React、SQLを使用して3つのフルスタック研究用Webサイトを設計・構築。
- 数千件の参加者エントリーを含むSQLデータベースを設計・管理し、データの整合性、セキュリティコンプライアンス、クエリパフォーマンスの最適化を確保。
- 連邦・州の助成金申請の開発をリードし、20,000ドル以上の資金を獲得。
- 4名以上の開発者チームをリードし、KPIおよびパフォーマンス目標を達成しつつメンタリングと技術指導を実施。
- クラスタリングアルゴリズム、ランダムフォレスト、ニューラルネットワーク等の高度な機械学習手法を実装し、患者データの分析、パターンの特定、予測モデルの開発を実施。
- numpyおよびpandasを活用した革新的な自動言語類似性分析システムを開発し、文書比較時間を約99%短縮（2-3日から10分未満）。米国言語学会（LSA）に論文を発表。

## 主要論文・特許

- Y. Fu, **L. Whalen**, X. Dong, S. Diao, C. Wu, E. Xie, S. Han, J. Kautz, Y. (Celine) Lin, P. Molchanov. "Training Framework for Converting Autoregressive Language Models into Faster Diffusion Language Models." 米国特許、出願予定, 2026. [NVIDIA]
- Y. Fu\*, **L. Whalen\***, Z. Ye, X. Dong, S. Diao, J. Liu, C. Wu, H. Zhang, E. Xie, S. Han, M. Khadkevich, J. Kautz, Y. (Celine) Lin, P. Molchanov. "Efficient-DLM: From Autoregressive to Diffusion Language Models, and Beyond in Speed." *arXiv preprint arXiv:2512.14067*, 2025年12月. [NVIDIA]

- **L. Whalen\***, Z. Du\*, H. You\*, C. Li, S. Li, Y. (Celine) Lin. “Early-Bird Diffusion: Investigating and Leveraging Timestep-Aware Early-Bird Tickets in Diffusion Models for Efficient Training.” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. [ジョージア工科大学]
- Z. Ye, Z. Wang, K. Xia, J. Hong, L. Li, **L. A. Whalen**, C. Wan, H. You, C. Lin, S. Kundu. “LAMB: A Training-Free Method to Enhance the Long-Context Understanding of SSMs via Attention-Guided Token Filtering.” *Association for Computational Linguistics (ACL)*, 2025. [ジョージア工科大学]
- A. Murphy, **L. Whalen**, S. Dubinsky, M. Gavin, J. F. Bailyn, J. Ginn. “On ‘historical unity’ of Russian and Ukrainian: A linguistic perspective on language conflict and change.” *Linguistic Society of America (LSA)*, 2023年4月. [サウスカロライナ大学]

## 学歴

---

ジョージア工科大学 大学院研究員 – 効率的機械学習システム	アトランタ, GA 2024年8月 – 2025年12月
サウスカロライナ大学 コンピュータサイエンス加速修士課程	コロンビア, SC 2024年1月 – 2024年8月
サウスカロライナ大学 コンピュータサイエンス学士	コロンビア, SC 2021年1月 – 2023年12月

## 受賞・表彰

---

- ソフトバンクグループ入社式代表 2026 2026年3月
- NSF GRFP受賞 (150,000ドル相当) 2024年4月
- 東洋大学交換留学生代表 2023年7月
- サウスカロライナ大学 2023年度優秀卒業生 2023年2月
- 米務省日本語クリティカル・ランゲージ・スカラー 2023年1月
- サウスカロライナ大学トップスカラー (100,000ドル以上相当) 2020年8月

## 資格

---

- 日本語能力試験 (N1)

## 語学力

---

- 英語 – ネイティブレベル
- 日本語 – ネイティブレベル (N1認定)
- 中国語 – 中級レベル
- ドイツ語 – 初級