

東北大学未踏スケールデータ アナリティクスセンターの取組み



TOHOKU
UNIVERSITY

令和5年（2023年）1月13日（金）

総長特命教授・副理事
未踏スケールデータアナリティクスセンター長
中尾光之



東北大学青葉山新キャンパス

2



薬学研究科

理学研究科

工学研究科

情報科学
研究科

産学連携機構

災害科学
国際研究所

国際集積エレクト
ロニクス研究開発
センター(CIES)



農学研究科

環境科学
研究科

地下鉄
青葉山駅

国際混住寮

次世代放射光施設
NanoTerasu

サイエンスパーク
4万m²

次世代放射光施設はナノを見る巨大な顕微鏡、価値ある膨大なデータを生成

- 新材料やデバイスの開発、生命機能の解明、カーボンニュートラルを目指す環境科学の開拓、さらには医薬品・食品・畜産・農業・漁業等の多様な産業分野で威力を発揮



次世代放射光施設 NanoTerasu



加速器エネルギー 3 GeV
蓄積電流 >400 mA
リング周長 349 m
消費電力 5 MW
年間最大運転時間 6,000 時間 (目標)

国の主体：量子科学技術研究開発機構 QST
パートナー：
光科学イノベーションセンター PhoSIC、
宮城県、仙台市、東北大学、東北経済連合会



課題解決の本丸 東北大学サイエンスパーク

GX・DX を加速するイノベーションエコシステムの創造

カーボンニュートラル時代のGXを牽引

革新的電池、省電力半導体、環境負荷を低減する材料、持続可能な農業など多様な研究開発

計測・計算融合のDXによる課題解決

NanoTerasu と 先端計測装置群が生成する膨大な画像データでナノ世界をデジタル化

東北大学青葉山新キャンパス

仙台駅から9分

仙台駅

地下鉄青葉山駅

次世代放射光施設
NanoTerasu
2024年本格稼働

国際放射光イノベーション・スマート研究センター
など2棟（計8,000㎡）が2023年完成予定

サイエンスパーク：4万㎡

60ペタバイト級
データストア

マルチモーダル
先端計測装置群
クライオ電子顕微鏡
NMR装置ほか

SINETや民間
ネットを介して
全国へサービス

多様な先端計測装置群から生み出される膨大なデータを統合的に分析、あらゆる現象の見える化とシミュレーションを通して広範な科学・産業現場の課題解決に挑戦

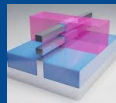


従来の100倍の明るさの軟X線で世界最先端へ

5

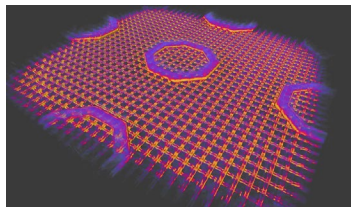
分野を問わずナノ世界を見える化（高精細デジタル化）

電子デバイス



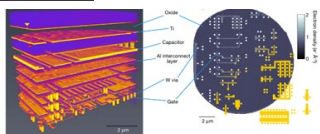
高いコヒーレント性を用いた非破壊の品質管理

デバイス内部のナノの欠陥を見る



デバイスの
元素識別
イメージング

高性能デバイスの実現

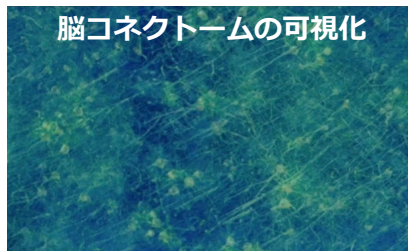


医療

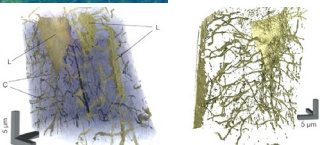


軽元素からなる組織の可視化
医用材料の生体適合性解明

脳コネクトームの可視化



マウス大腿骨の
内部微細構造

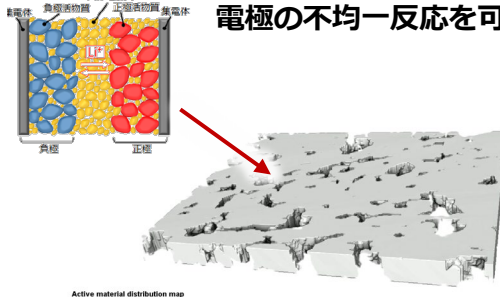


電池



機能に関わる電子状態の変化をリアルタイムで可視化

全固体電池の充電時における
電極の不均一反応を可視化



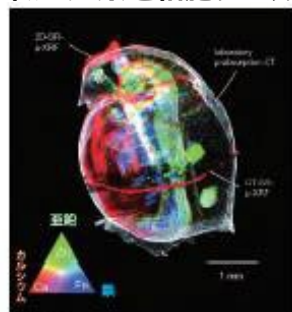
放電 充電 資料：東北大・雨澤浩史

環境



軽元素の分布や状態を直接的
かつリアルタイムで可視化

ミジンコの微量元素を細胞レベルで可視化



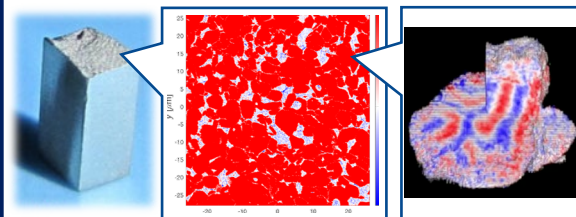
資料提供：B. De Sinter et al. Journal of Analytical Atomic Spectrometry.

磁気デバイス



電子のスピンが見える
（偏光で磁気分布を可視化）

磁区構造を可視化して
強力な磁石をつくる



資料：東北大・中村哲也

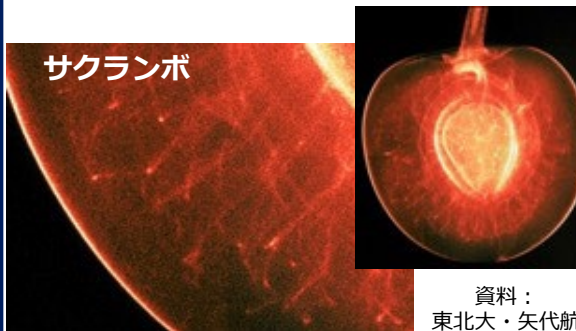
食品・畜産・農業・漁業



軽元素の分布や状態を直接的
かつリアルタイムで可視化

食の安全と高付加価値化の実現

サクランボ



資料：
東北大・矢代航



東北メディカル・メガバンク計画

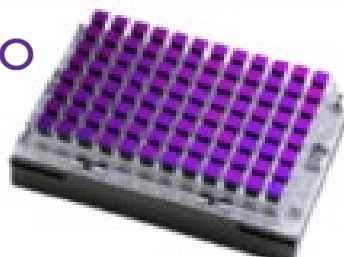
2012年2月 東北メディカル・メガバンク機構 設立
世界初の7万人規模の三世代コホート調査
日本最大規模の15万人地域住民・三世代コホート調査
一般住民（前向き）コホートを基盤として未来型医療・未来型予防の創造へ

50ペタバイト級データ
解析センターを併設



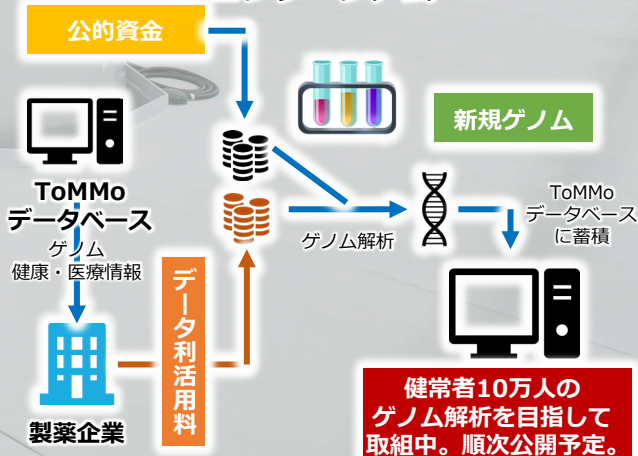
日本人に最適化した簡易ゲノム解析ツール「**ジャポニカレイ®NEO**」を開発

Japanica
ArrayNEO



山本 雅之 機構長

全ゲノム情報と医療・健康情報の統合解析 コンソーシアム





先端コアファシリティ群

- クライオ電子顕微鏡：2台
(最新鋭300kV透過型)
- 無冷媒超伝導マグネット
(利用可能な世界最強25テスラ)
- 低乱熱伝達風洞実験装置
(国内最高水準の性能)



ほか多様な設備を保有、利用企業多数

半導体共創体 ファシリティ群

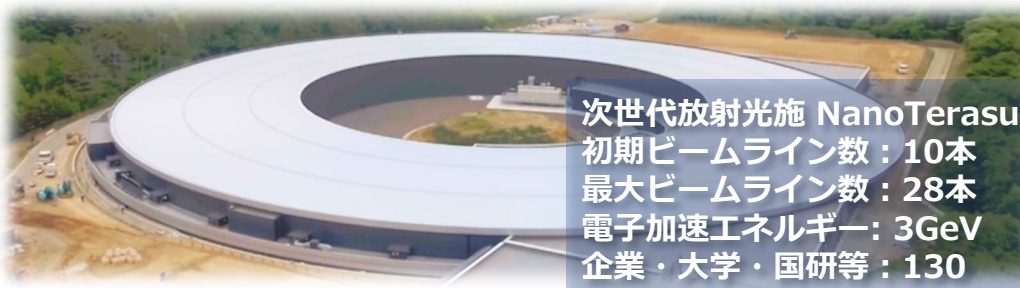
- 国内最大級の学内クリーンルーム群
(総面積8,500m²)
- 利用企業数：約450社
- 国際集積エレクトロニクス研究開発センターCIES、未来科学技術共同研究センターNICHe、マイクロシステム融合研究開発センターμSIC等



理工学



ライフ
サイエンス



次世代放射光施設 NanoTerasu
初期ビームライン数：10本
最大ビームライン数：28本
電子加速エネルギー：3GeV
企業・大学・国研等：130

半導体工学



情報科学



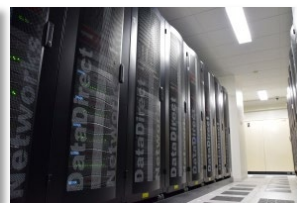
東北メディカル・メガバンク ファシリティ群

- 世界初の7万人規模の三世代コホート調査、日本最大の15万人規模の住民コホート調査
- バイオバンク (日本最大級の約400万本の生体試料)
- 50ペタバイト級データ解析センターを併設



スーパーコンピュータ システム群

- 総合性能：12ペタフロップス
- 合計容量：81ペタバイト (ストレージ)
- 利用企業数：42社





東北大学データシナジー創生機構

CDOを機構長としたデータシナジー創生機構を構築し、機構長の強力なガバナンスの下、3つの組織を一体的に連動させながら教育・研究活動を展開して、分野を越境したデータ・解析手法・人材・設備の一元化・共有化・強化を強力に推進

サイバーサイエンスセンター(CSC)

- ・ 民間企業との協働(共同研究部門)による次世代スーパーコンピュータの研究開発
- ・ スーパーコンピュータを活用した実用システムの社会実装に関する研究開発
- ・ 耐災害性、セキュリティ強度に優れた柔軟な高速ネットワークに関する研究開発
- ・ 未踏スケールデータに対応した超大規模データレイクに関する研究開発

データ駆動科学・AI教育研究センター(CDS)

- ・ 学部新入生全員に対するリテラシー教育 (AI & Data for All) 、学部専門教育でのスペシャリスト養成 (Specialists with AI & Data)、そして、大学院での高度AIプロフェッショナル人材育成 (AI & Data Top leader) による、全学的な教育体制
- ・ 学修データ利活用による教育のデジタル変革への取り組み

未踏スケールデータアナリティクスセンター(UDAC)

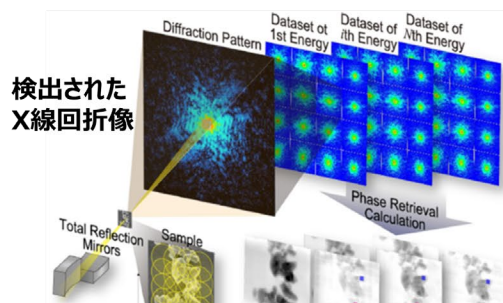
未踏スケールデータ (サイズ、分解能、スピード、多様性において、従来のスケールを遥かに凌駕するデータ) のデータ科学を創生し、未踏領域の価値創造を実現する。その成果に基づいた産学連携、人材育成、起業を促進する。



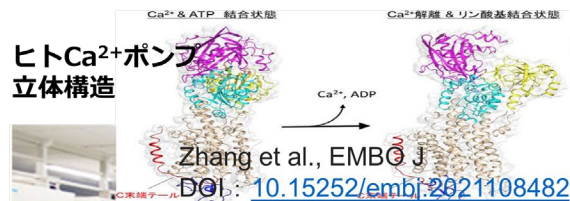
未踏スケールデータアナリティクスセンター

UDAC: Unprecedented-scale Data Analytics Center

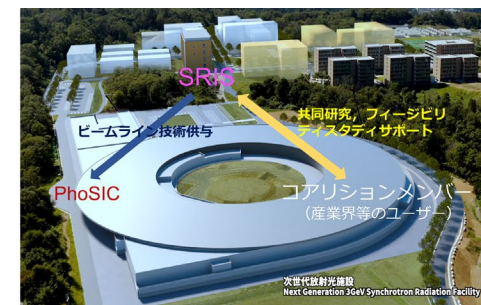
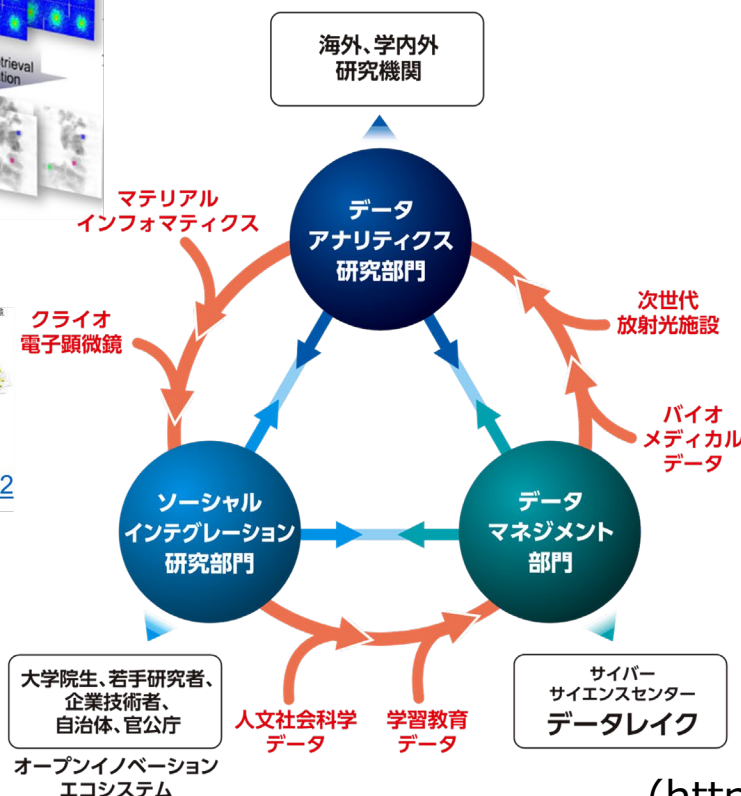
未踏スケールデータ（サイズ、分解能、スピード、多様性において、従来のスケールを遥かに凌駕するデータ）のデータ科学を創生し、未踏領域の価値創造を実現する。その成果に基づいた産学連携、人材育成、起業を促進する。



タイコグラフィXFSA法 (Hirose et al., Angew. Chem. Int. Ed. 2018, 57, 1474–1479)



クライオ電子顕微鏡単粒子解析：生体機能分子立体構造の高解像度測定



次世代放射光施設（ナノテラス）：従来の軟X線領域の100倍の輝度・コヒーレンスで、ナノの機能を可視化することにより、**農学、材料科学、理学、工学、環境科学、生命科学、創薬**など様々な分野での画期的な研究成果が期待される。学術、産業界のデータ駆動型研究開発を支援すると共に新しい利用制度“**コアリション・コンセプト**”で、産学融合を加速する。2023年稼働開始。



東北メディカル・メガバンク：世界最大規模の15万人フルゲノム、7万人3世代コホート＋情報科学と医学の**分野融合研究者** => **数少ないビッグデータ研究の成功事例**

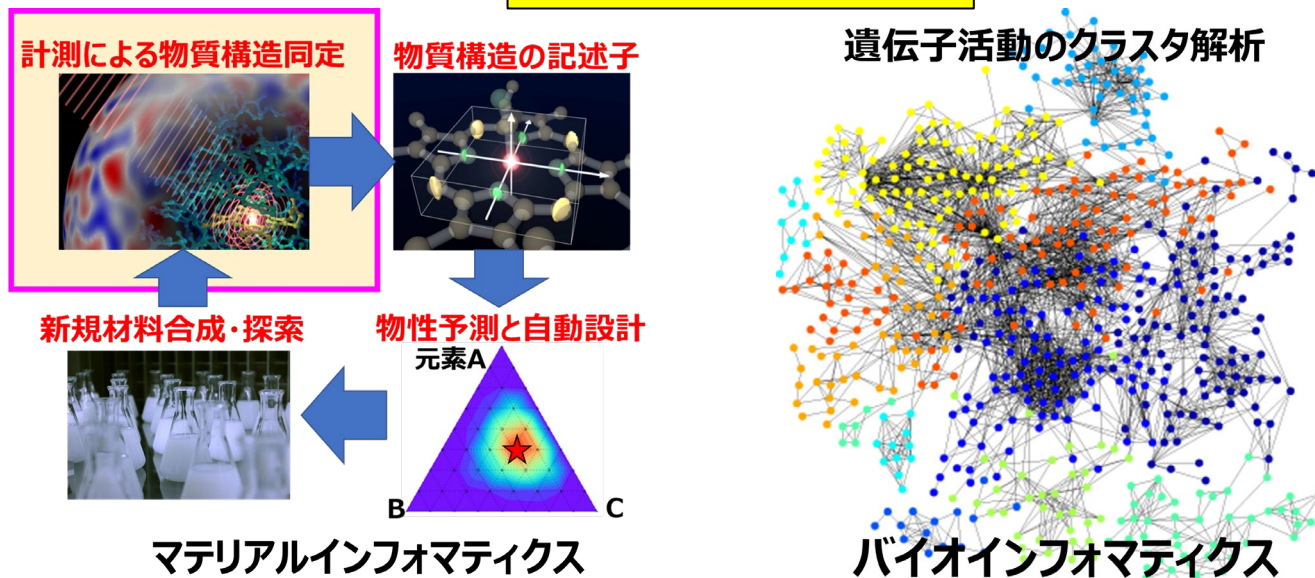
(<https://udac.tohoku.ac.jp/>)



未踏スケールデータのデータアナリティクスを極め、学術・産業界と共有します。

- 先端的なAI やデータ科学をさらに発展させることによって未踏スケールデータのハンドリング・分析手法の研究開発
- 分野を横断した研究領域融合による新たな分析技術の開発・共有、そして、新たな社会的価値の創造
- 研究・教育データの高度な加工・分析する技術に関する研究開発および研究教育拠点に対するコンサルティング
- データ駆動型の新しい教育改善手法の開発とその実践的な検証、およびそのサービス化に関する研究

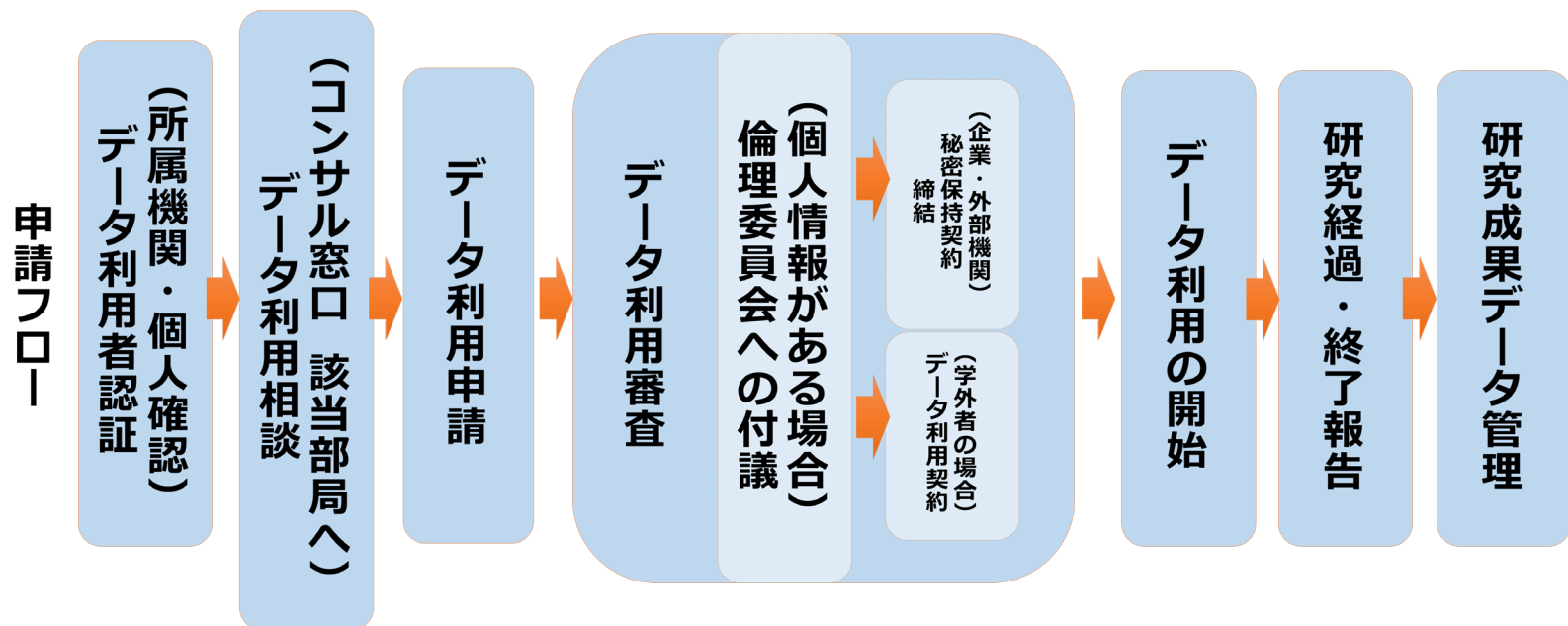
志賀元紀教授の研究





未踏スケール研究データの効率的でセキュアな利用を促進するマネジメントを行います。

- 研究デジタルデータ保存サービスとアーカイブ
- データ関連の法令遵守、規程制定、契約業務
- 研究データ管理に関係する機能・業務を集約化
- サイバーサイエンスセンターと連携し超大規模ストレージ(データレイク)、スパコン、高速ネットワーク等を基盤としたデータプラットフォーム構築支援、並びにそれに基づく未踏スケールデータの収集・格納・管理・共有利用



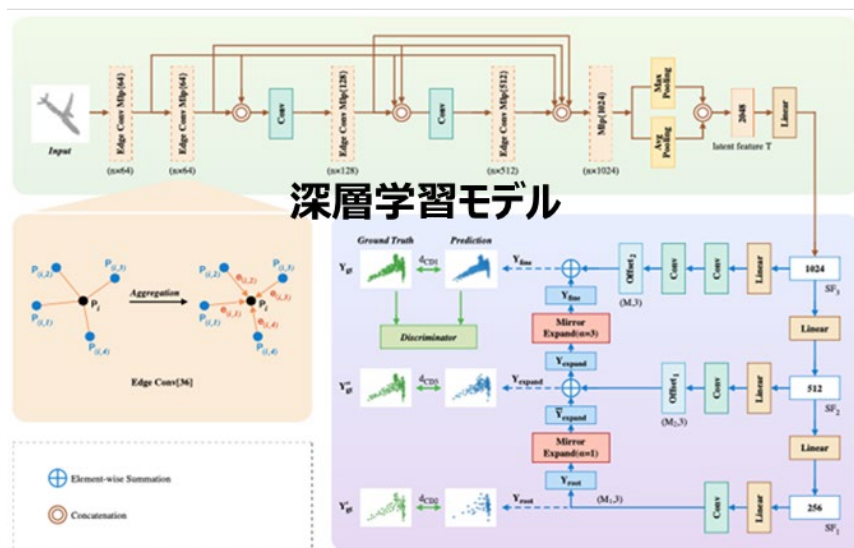


未踏スケールデータに挑む学生・研究者・企業技術者を受入れ、オープンエコシステムによる価値創生や起業を支援します。

- 学内の学部・研究科・研究所・センター等との連携、および海外研究教育機関、企業・政府・自治体との共同研究による社会実装の推進
- データ駆動科学・AI 教育研究センターとの協働による社会人特別プログラムの企画
- オープンエコシステムによる国内外からの大学院生、若手研究者、企業技術者の受入れ、価値創生への取り組みや起業の支援

山田和範教授の研究

AIアルゴリズム開発



ドメインデータサイエンティスト養成

1. データ科学と機械学習
2. Pythonの基本的な使用方法
3. Pythonの発展的な使用方法
4. 教師あり学習法
5. 教師なし学習法
6. TensorFlowの基本的な利用方法
7. 多層パーセプトロン
8. 再帰型ニューラルネットワーク
9. アテンションネットワーク
10. Hugging Faceの利用方法
11. 敵対的生成ネットワーク
12. 強化学習法



Senior Domain Specialist with AIMD

データ科学国際共同大学院

- ・情報科学研究科が6研究科と連携して提供する，留学生と日本人学生のための双方向・分野横断型学習環境
- ・実世界のビッグデータ課題に取り組むプロジェクト型学習（PBL）の実施
- ・海外研究機関のメンターとの最低6ヶ月間の共同研究機会の提供

情報科学研究科および連携する大学院群

Leading Researcher with Advanced AIMD

- ・未踏スケールデータ科学を学び、研究を行う機会を提供
- ・アナリスト、ビジネスエキスパートとのユニットによる課題解決を行いデータ科学の利活用の仕方を理解する

未踏スケールデータアナリティクスセンター (UDAC)

Domain Specialists with AIMD

- ・AIMDについて学ぶ機会を文理問わず学生の専門性に応じて提供する
- ・専門知識とAIMDスキルのダブルメジャー修得を強く推奨する

学部、学科専門教育

AIMD for All

- ・新入生全員（2,500名）にAIMDリテラシー教育を実施
- ・AIMDの先進的なトピックを学ぶと共に，現在進行中の研究とその最前線を知る機会を提供し，学生のモチベーションを高めるためのユニークなコース

全学教育

