

## 機械学習を用いたシステムの高品質化・実用化を加速する“Engineerable AI”技術の開発

研究開発代表者： 石川 冬樹 国立情報学研究所 アーキテクチャ科学研究系 准教授

共同研究機関： 東京科学大学、北海道大学、九州大学、早稲田大学、  
国立がん研究センター、慶應義塾大学



### 目的：

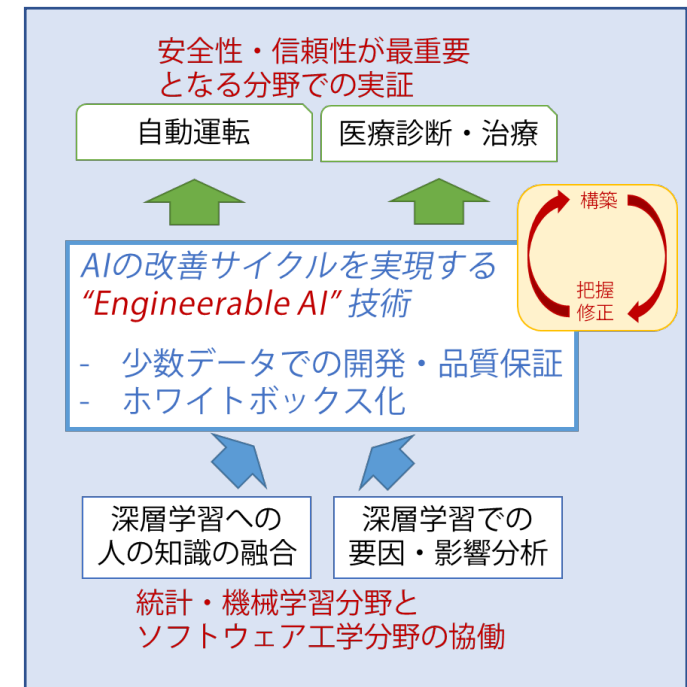
深層学習を用いるAIシステムに対し、構築、分析や検査、修正を継続的に反復し、安全性や信頼性を効果的・効率的に向上・保証するための“Engineerable AI”技術を確立し、自動運転および医療診断・治療においてその有効性を示す。

### 研究概要：

機械学習の利点は、データセット全体からファジーなゴール全体を帰納的・経験的に実現できることである。しかし、大量データから機能全体を大まかに導くというアプローチ上、レアなニーズ・状況に対応したり、多数の状態に対する性能限界を把握・修正したりすることができない。このことは、交通・医療など安全性・信頼性が最重要である領域において、AIの展開を困難にしている。

本研究開発では“Engineerable AI”というビジョンを提唱し、AIシステムの開発・品質保証・運用を安定化・効率化することで、その高品質化・実用化を加速する。具体的には、大量データを用いた学習・反復修正に基づく従来AIに対し、深層学習に人の知識を交え反映しAIを構築する技術や、品質に影響する要因を抽出・分析しての保証や修正を行う技術に取り組む。

本研究開発は統計・機械学習分野とソフトウェア工学分野の研究者が密に連携して行い、自動運転および医療診断・治療においてその有効性を実証する。



## Engineerable AI Techniques for Practical Applications of High-Quality Machine Learning-based Systems

**Project Leader :** Fuyuki Ishikawa

Associate Professor, Information Systems Architecture Science Research Division,  
National Institute of Informatics

**R&D Team :** Institute of Science Tokyo, Hokkaido University, Kyushu University, Waseda University, National Cancer Center, Keio University



### Summary :

Machine learning techniques enable to implement functionality that satisfies a fuzzy goal in an inductive way by training from data. However, this approach makes it impossible to handle rare situations or to understand and fix performance limitations over many situations. This characteristic poses an obstacle in application to safety- and reliability-critical domains.

We envision techniques for “Engineerable AI” that support development, quality assurance, and operation of AI systems, thus promoting practical applications of high-quality AI systems. We develop techniques for construction of AI by incorporating human knowledge into deep learning and for assurance and improvement by extracting and analyzing factors affecting the target quality. We promote this research by collaboration of the machine learning area and the software engineering area and demonstrate the effectiveness in autonomous driving and healthcare.

