Technique for Safe and Trusted AI

Tech ID: 33855 / UC Case 2024-9B0-0

ABSTRACT

Researchers at the University of California Davis have developed a technology that enables the provable editing of DNNs (deep neural networks) to meet specified safety criteria without altering their architecture.

FULL DESCRIPTION

This invention presents systems and methods for editing deep neural networks (DNNs) to ensure they satisfy given safety specifications. Unlike traditional approaches that may require retraining from scratch, this method employs formulas and efficient programming solvers to adjust DNNs, ensuring they adhere to specified input-output criteria without modifying the DNN's structure.

APPLICATIONS

- Enhancement of safety-critical applications such as self-driving cars and healthcare systems.
- Improvement of pattern recognition and problem-solving in computational models.
- Development of more reliable and efficient neural network editing tools.

FEATURES/BENEFITS

- Supports safety specifications using quantified linear formulas, accommodating infinite data sets in high-dimensional spaces.
- Maintains the original architecture of the DNN, avoiding complex structural changes.
- Provides a provable editing approach that ensures DNNs meet specified safety criteria.
- Significantly reduces the time, processor resources, memory, and power typically required for DNN editing.
- Reduces time-consuming and resource-intensive retraining of DNNs for error correction.
- Provides guidance for correcting DNNs identified as inaccurate by verifiers.
- Eases difficulty in ensuring DNNs meet safety-critical application standards.

PATENT STATUS

Patent Pending

CONTACT

Byron N. Roberts
bnroberts@ucdavis.edu
tel: 530-754-8689.

INVENTORS

- Tao, Zhe
- Thakur, Aditya

OTHER INFORMATION

KEYWORDS

artificial intelligence, deep neural networks (DNN), safety-critical application enhancement, pattern recognition, quantified linear formulas, DNN correction

CATEGORIZED AS

- **Engineering**
  - Engineering
  - Other
  - Robotics and Automation
- **Computer**
  - Other
  - Security
  - Software
- **Medical**

- Other
- **Security and Defense**
  - Cyber security
  - Other
- **Transportation**
  - Other

RELATED CASES

2024-9B0-0

University of California, Davis
Technology Transfer Office
1 Shields Avenue, Mrak Hall 4th Floor,
Davis,CA 95616

Tel:
530.754.8649
techtransfer@ucdavis.edu
https://research.ucdavis.edu/technology-transfer/
Fax:
530.754.7620