Fact Sheet
EMBARGOED TILL 1030HRS, 22 NOVEMBER, 2018

OPEN SOURCE AI LIBRARIES

Artificial Intelligence (AI) has been identified as a frontier technology essential to growing Singapore's Digital Economy. To maintain Singapore's key node status in the global AI ecosystem, IMDA announced in October 2017 that it will spearhead the development of AI areas that can help to grow and create new capabilities.

One key focus area is AI libraries, which are collections of Application Programming Interfaces (APIs), source codes, databases and more. IMDA-created libraries can be released as open-source libraries to spur the growth and innovation of AI-related apps in Singapore. Two open source AI libraries have been created. They are:

1) National Speech Corpus V1.0
2) Intelligent Sensing Toolbox

National Speech Corpus V1.0

First announced in November 2017, the first version of the National Speech Corpus (NSC) was released on 22 November 2018. It contains 2000 hours of locally accented audio and corresponding text transcriptions. There are more than 40 thousand unique words within the text transcriptions which consist of local words such as "Tanjong Pagar", "Ice Kachang", or "Nasi Lemak". The data is made available via the Singapore Open Data License[1].

Automatic speech recognition engines use multiple corpus collections (collectively called corpora) to accurately train themselves to interpret spoken words and transcribe this into text. The NSC thus enables global technology providers to provide speech-related applications (such as voice assistants) for use here. The NSC will be continually updated as-and-when necessary.

Interested parties can request for a copy of the corpus at:

https://form.gov.sg/forms/imda/5bd7f0ec4af98a000f4e826a

---

[1] https://data.gov.sg/open-data-licence

Intelligent Sensing Toolbox

In Big Data analysis, machine learning-based engines can analyse large datasets from multiple sources and identify patterns to generate insights.

The Intelligent Sensing Toolbox (IST) is a suite of open-source AI tools and technologies developed by IMDA. This sense-making AI algorithm will offer businesses a plug-and-play, open-source source code which can be quickly adapted and layered on top of its existing data analytics system to help it make better decisions. The IST is open sourced on the Apache 2.0 License[2] and more functionality could be added progressively, in collaboration with industry partners.

IMDA had successfully completed a proof-of-concept of the IST through its Intelligent Data Centre programme with the National Supercomputing Centre in 2017. The proof-of-concept demonstrated then that the IST could draw new insights in data centres by identifying the probable cause of abnormal behaviours of server nodes in the data centre.

The toolbox's algorithms are available at https://gitlab.com/IMDA-DSL/Intelligent-Sensing-Toolbox

AI as an enabler in the Digital Economy Framework for Action

As part of the AI strategy, IMDA will:

1. Help drive industry adoption and understanding of the benefits from implementing AI;

2. Build a vibrant core of AI developers and companies to foster an ecosystem of users and suppliers as well as build an AI-ready pool of talented researchers and engineers;

3. Address specific AI leadership areas to grow industry capabilities and develop R&D capabilities; and

4. Continue to provide clarity and review relevant regulations for a conducive AI business environment.

---

[2] https://www.apache.org/licenses/LICENSE-2.0

IMDA, in collaboration with AI Singapore, had previously announced initiatives in support of the above.

For media clarifications, please contact:

NEUBRONNER Eugene (Mr)
Manager, Communications and Marketing, IMDA
DID: (65) 6211 1182
Email: Eugene_Neubronner@imda.gov.sg

KOH Christopher (Mr)
Senior Manager, Communications and Marketing, IMDA
DID: (65) 6211 1709
Email: Christopher_Koh@imda.gov.sg