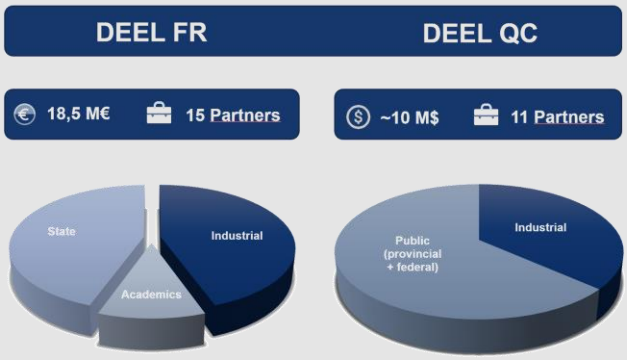
DEEL
DEpendable & Explainable Learning

DEEL MAG #01

TOULOUSE & QUEBEC JOIN FORCES TO DEVELOP AI FOR CRITICAL SYSTEMS

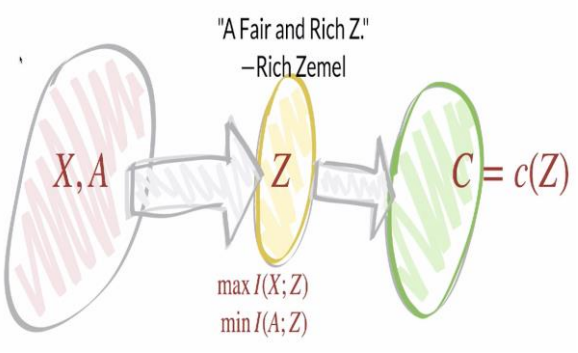
DEEL QUEBEC: 100% FUNDED



On October, the 15th we received the notification by the CRSNG of 50% of the funding of DEEL QC It was already a good news. We know since December the 20th that the Quebec government has also signed the provincial part of the funding. This notification officialises the **French/Quebec collaboration** and makes possible a common communication on DEEL. Next step is contracting the international DEEL partnership.

FOCUS ON FAIRNESS CHALLENGE

Fairness is an important topic in Artificial Intelligence, as pointed out in the book by Cathy O'Neil's "*Weapons of maths destruction*"; Machine Learning (ML) practitioners became aware that when some bias is already present in the observations, possibly resulting from past discrimination of certain categories of the population, then ML algorithms trained on this data reproduce, even reinforce the same discrimination, hence generalizing it to new observations.



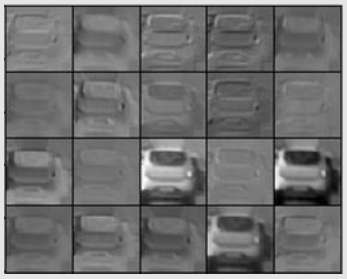
In DEEL, we **first improve the fairness metrics** recently developed by the ML community, by **adding confidence intervals** to better certify the presence of bias. Then we tackle bias issues for industrial use cases where we have sensitive parameters that shall not have too much influence on the outcome of the algorithm (source of image for a land cover prediction, gender and accent for a speech-to-text model, etc.). In a second phase, **we want to focus on industrial use** cases where there is no explicit sensitive parameter. Our goal is to be able to automatically detect bias in training data, measure its impact on the model outputs and finally be able to learn a model robust to distribution change between training and operations or adapt a model to specific operational conditions.

DEEL Publications on fairness

[2019] JM. Loubes, P. Gordaliza, E. Del Barrio, F. Gamboa, Obtaining Fairness with Optimal Transport, ICML
[2020] E. Pauwels, M. Serrurier, JM. Loubes, Fairness with Wasserstein Adversarial Networks, submitted in ECAI 2020
[2019] L. Risser, Q. Vincenot, N. Couellan, JM Loubes, Using Wassertstein-2 regularization to ensure fair decision with neural-network classifiers, To be submitted

FOCUS ON BLINKING DATASET

A blinking dataset has been developed. It is based on real imagery from **Renault prototype of Autonomous Driving** Level 4 vehicle ("hands off, eyes off"). It contains about 5000 images of the rear of the vehicles (100Mo of data) divided into 4 classes: No Blink, Blink Right, Blink Left and Warnings. This dataset is under testing regarding the **Ethik AI** explainability toolkit. It will be soon an input to the fairness evaluation algorithms.



MOBILIT.AI 2020 will take place in Toulouse on May 26-27

As in 2019, the 2020 edition aims at **creating an expert community in the field of dependable AI applied to the transport sector.** It will bring together academic and industrial experts around questions such as: **Certification issues and AI, which guarantees for AI, towards an embedded AI.** Both days will be organized around guest speakers and round tables. New for 2020: **3 master classes in AI** will be held during the forum. Another new feature for 2020 is that special attention will be paid to the integration of students into the forum, giving recruiters the opportunity to get contacts.

The venue of the **MobiliT.AI forum** has been selected, it will be at the auditorium of the Paul Sabatier University. The scientific committee is at work to specify the program and select the speakers. It is composed of **Jean-Michel Loubes** (Professor @ IMT/DEEL/ANITI), **Francois Laviolette** (Professor @ Université Laval/DEEL), **Francis Bach** (Researcher @ INRIA/ENS), **Juliette Mattioli** (Senior Expert @ Thales), **Francois Provencher** (Innovation Officer @ P&WC) and **Christophe Lecante** (President @ TecKnowMetrix).

KEY DATES

9 & 10 jan	Certification mission workshop
23 & 24 jan	Workshop on explainability
28 jan	DEEL COMOPS #3
29-31 jan	ERTS DEEL – **Current Challenges in the Certification of Machine Learning for Safety Critical Systems**, Eric Jenn, Alexandre Albore, Franck Mamalet, Grégory Flandin, Christophe Gabreau, Hervé Delseny, Adrien Gauffriau, Hugues Bonnin, Lucian Alecu, Jérémy Pirard, Baptiste Lefevre, Jean-Marc Gabriel, Cyril Cappi, Laurent Gardès, Sylvaine Picard, Gilles Dulon, Brice Beltran, Jean-Christophe Bianic, Mathieu Damour, Kevin Delmas, Claire Pagetti
6 & 7 feb	**AI Safety Landscape** & **SafeAI** in New York A High-Probability Safety Guarantee for Shifted Neural Network Surrogates, M. Ducoffe, S. Gerchinovitz, J. Sen Gupta

PROJECT MANAGEMENT

Project Contract R2V1 has been issued and signed by 8/11 partners. ANR report will be available end January.

Hiring:	
✚ 14 secondments for the certification mission (20% each)	✚ 2 / ▪ 1 secondments for the core team
✚ 2 CDI signed in December	✚ 4 ongoing PhD, 1 pending institutional decision